Statistical detection for digital image forensics

Tong Qiao

Thèse
de doctorat
de l'UTT

Tong QIAO

Statistical Detection for Digital Image Forensics

Spécialité :
Optimisation et Sûreté des Systèmes

2016TROY0006

Année 2016

utt
université de technologie
Troyes

THESE

pour l'obtention du grade de

DOCTEUR de l'UNIVERSITE DE TECHNOLOGIE DE TROYES

Spécialité : OPTIMISATION ET SURETE DES SYSTEMES

présentée et soutenue par

Tong QIAO

le 25 avril 2016

Statistical Detection for Digital Image Forensics

JURY

M. F. MORAIN-NICOLIER	PROFESSEUR DES UNIVERSITES	Président
M. P. BAS	DIRECTEUR DE RECHERCHE CNRS	Rapporteur
M. R. COGRANNE	MAITRE DE CONFERENCES	Examinateur
M. W. PUECH	PROFESSEUR DES UNIVERSITES	Rapporteur
M. F. RETRAINT	ENSEIGNANT CHERCHEUR	Directeur de thèse
M. T. H. THAI	RESEARCHER	Examinateur

To my Mother, my Father, and my Grandmother,

for their unlimited support, encouragement, and love.

Acknowledgments

This work has been carried out within the Laboratory of Systems Modeling and Dependability (LM2S) at University of Technology of Troyes (UTT). It is funded by China Scholarship Council (CSC) and the region Champagne-Ardenne, IDENT project.

This work has been accomplished under the supervision of M. Florent RETRAINT. I would like to express my deepest appreciation to him for his highly professional guidance and incessant support. I highly value the friendly yet professional environment created during my three and a half year doctoral period. The high confidence he has given to me is definitely the greatest reward of my endeavors.

I also would like to express my deepest appreciation to my PhD co-advisor, M. Rémi COGRANNE who has guided me with high efficiency and availability. I still remember when I came to UTT three and a half year ago, Rémi indeed gave me great help at the first year of my doctoral period. It is my honor and pleasure to work with him. His personal and professional help during my doctoral period is definitely priceless.

I give my endless appreciation to the members of our team, especially Dr. Thanh Hai THAI for his selfless sharing his research fruits and patient discussion with me. I also would like to give my sincerely thanks to Ms. Monia CHEIKH ROUHOU, M. Moustapha EL AOUN and Ms. Thi Ngoc Canh DOAN. It was the most valuable time to work with them in my life.

I also would like to express my special thanks to M. Patrick BAS and M. William PUECH for accepting to review my PhD thesis. I would also like to thank M. Frédéric MORAIN-NICOLIER and M. Thanh Hai THAI for agreeing to examine this thesis. Valuable remarks provided by the respectful experts in this field like them would improve the thesis' quality.

Most of all, I would like to thank my family for supporting me, encouraging me, and loving me all my life.

Résumé

Le XXIème siècle étant le siècle du passage au tout numérique, les médias digitaux jouent un rlôe de plus en plus important. les logiciels sophistiqués de retouche d'images se sont démocratisés et permettent de diffuser facilement des images falsifiées. Ceci pose un problème sociétal puisqu'il s'agit de savoir si ce que l'on voit a été manipulé. Cette thèse s'inscrit dans le cadre de la criminalistique des images. Trois problèmes sont abordés: l'identification de l'origine d'une image, la détection d'informations cachées dans une image et la détection d'un exemple falsification : le rééchantillonnage. Ces travaux s'inscrivent dans le cadre de la théorie de la décision statistique et proposent la construction de détecteurs permettant de respecter une contrainte sur la probabilité de fausse alarme. Afin d'atteindre une performance de détection élevée, il est proposé d'exploiter les propriétés des images naturelles en modélisant les principales étapes de la chaîne d'acquisition d'un appareil photographique. La méthodologie, tout au long de ce manuscrit, consiste à étudier le détecteur optimal donné par le test du rapport de vraisemblance dans le contexte idéal où tous les paramètres du modèle sont connus. Lorsque des paramètres du modèle sont inconnus, ces derniers sont estimés afin de construire le test du rapport de vraisemblance généralisé dont les performances statistiques sont analytiquement établies. De nombreuses expérimentations sur des images simulées et réelles permettent de souligner la pertinence de l'approche proposée.

Abstract

The remarkable evolution of information technologies and digital imaging technology in the past decades allow digital images to be ubiquitous. The tampering of these images has become an unavoidable reality, especially in the field of cybercrime. The credibility and trustworthiness of digital images have been eroded, resulting in important consequences in terms of political, economic, and social issues. To restore the trust to digital images, the field of digital forensics was born. Three important problems are addressed in this thesis: image origin identification, detection of hidden information in a digital image and an example of tampering image detection: the resampling. The goal is to develop a statistical decision approach as reliable as possible that allows to guarantee a prescribed false alarm probability. To this end, the approach involves designing a statistical test within the framework of hypothesis testing theory based on a parametric model that characterizes physical and statistical properties of natural images. This model is developed by studying the image processing pipeline of a digital camera. As part of this work, the difficulty of the presence of unknown parameters is addressed using statistical estimation, making the application of statistical tests straightforward in practice. Numerical experiments on simulated and real images have highlighted the relevance of the proposed approach.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

Acronym	**W**hat (it) **S**tands **F**or
AWGN	Additive White Gaussian Noise.
AC	Alternating Current.
bpp	bits per pixel.
cdf	cumulative distribution function.
CCD	Charge-Coupled Device.
CCN	Circular Correlation Norm.
CFA	Color Filter Array.
CLT	Central Limit Theorem.
CMOS	Complementary Metal-Oxide Semiconductor.
CG	computer generated.
DC	Direct Current.
DCT	Discrete Cosine Transform.
DFT	Discrete Fourier Transform.
DSC	Digital Still Camera.
EXIF	Exchangeable Image File.
EM	Expectation Maximization.
FPN	Fixed Pattern Noise.
FAR	False Alarm Rate.
GLR	Generalized Likelihood Ratio.
GLRT	Generalized Likelihood Ratio Test.
HSV	Hue Saturation Value
i.i.d.	independent and identically distributed.
JPEG	Join Photographic Expert Group.
LSB	Least Significant Bit.
LR	Likelihood Ratio.
LRT	Likelihood Ratio Test.
LS	Least Squares.
LAUMP	Locally Asymptotically Uniformly Most Powerful.
MAE	Mean Absolute Error.
mAE	Median Absolute Error.
ML	Maximum Likelihood.
PCE	Peak to Correlation Energy.
pdf	probability density function.
pmf	probability mass function.
PIM	photographic image.
PRNU	Photo-Response Non-Uniformity.
ROC	Receiver Operating Characteristic.
RGB	Red Green Blue.
RLE	Run-Length Encoding.

SPA	Sample Pair Analysis.
SPN	Sensor Pattern Noise.
SVM	Support Vector Machine.
SVD	Singular Value Decomposition.
TIFF	Tagged Image File Format.
WF	Weighting Factor.

Glossary of Notations

Notation	Definition
α_0	False alarm rate.
β_δ	Power of the test δ.
χ^2_K	Chi square distribution with K degree of freedom.
γ	Gamma factor.
δ	Statistical test.
η	Noise.
ξ	Noise.
Γ	Gamma distribution.
$\mathcal{B}$	Binomial distribution.
μ	Expectation.
σ	Standard deviation.
τ	Decision threshold.
ϕ	Probability density function of a standard Gaussian random variable.
Δ	Quantization step in the DCT domain.
Λ	Likelihood Ratio.
Φ	Cumulative distribution function of a standard Gaussian random variable.
θ	Parameter space.
e	estimate error.
$\mathbb{E}$	Mathematical expectation.
$\mathbb{R}$	Set of real numbers.
$\mathbb{V}\mathrm{ar}$	Variance.
$\mathbb{Z}$	Set of integer numbers.
$\Gamma(k, \theta)$	Gamma distribution with shape parameter k and scale parameter θ.
$\mathcal{H}_0$, $\mathcal{H}_1$	Null hypothesis and alternative hypothesis.
$\mathcal{K}_{\alpha_0}$	Class of tests whose false alarm probability is upper-bounded by α_0.
$\mathcal{N}(\mu, \sigma^2)$	Gaussian distribution with mean μ and variance σ^2.
$\mathcal{P}(\lambda)$	Poisson distribution with mean λ and variance λ.
$\mathcal{S}$	Source of digital images.
$\mathcal{Z}$	Set of possible pixel values.
$\mathcal{Z}^N$	Image space.
$\mathbf{Z}$	Natural image.
(a, b)	Parameters of Poissonian-Gaussian noise model.
(a_k, b_k)	Parameters of enhanced Poissonian-Gaussian noise model.
$(\tilde{a}, \tilde{b})$	Parameters of generalized signal-dependent noise model.
$(\tilde{a_k}, \tilde{b_k})$	Parameters of improved signal-dependent noise model.

(c, d)	Parameters characterizing the relation between $\tilde{a}$ and $\tilde{b}$, which are the parameters of improved signal-dependent noise model.

Chapter 1

General Introduction

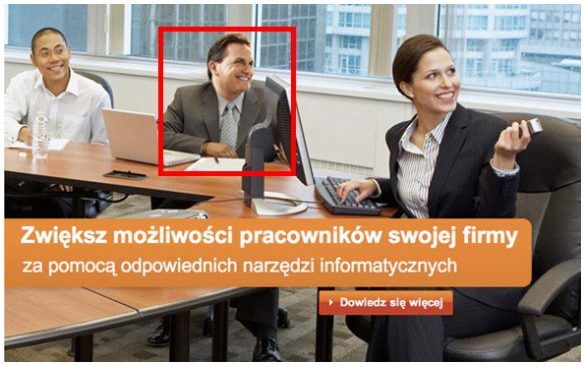
1.1 Research Background and Problem Description

During the past two decades, digital industry revolution changes the way human-being are living in this world. With the digital still camera coming out, we can escape from the film age and strike into the digital world. Hence, mass of digital images are flooding into our daily life. Generally speaking, an image has three main functions: conveying information, artistry and sharing memory. Compared with text information, an image is more direct and instant. The value of a good photograph is priceless. A photo can also help us record the beautiful moment, but does not vanish as time passing by.

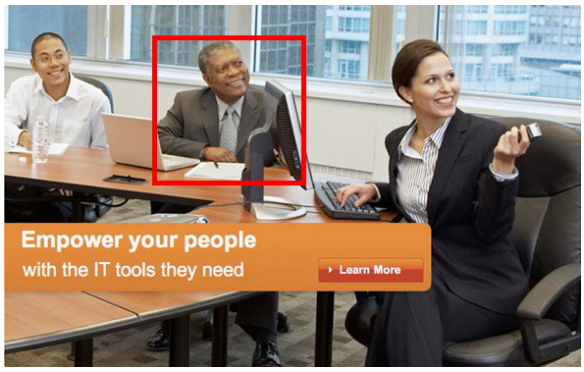
We always believe that we are living in the world where what we are seeing are definitely true. Does not an image possibly cheat our eyes? We are gradually suspecting authenticity of a given image. Since the photo-editing software can easily manipulate the digital image, it is unavoidable that some image forgers maliciously falsify the digital image to distort the truth. Some irresponsible journalists publish the manipulated photo on the newspaper in order to fool the public. Supposed that the falsified image is presented in the court, the jury probably makes wrong judgment. Moreover, with the development of smart mobile phone, hundreds of image processing application softwares spring up recently, hands-on operation becomes more and more feasible and easier. Therefore, it is very urgent to restore our confidence to the digital image. In this background, digital image forensics emerges.

Prior to our discussion about image forensics, we first illustrate two simple examples of falsified images. On August 2009, *Microsoft U.S.* published an advertisement picture on its official website (see Figure 1.1a), but the original picture was altered when the Polish subsidiary of *Microsoft* used it replacing a black person sitting in the middle by a white person (see Figure 1.1b). This change unavoidably arose argument about racial discrimination. Technically, if the original image can not be obtained, it is hardly possible to authenticate the faked image by visual observation. In general, we define this manipulation as image *splicing*.

Then let us illustrate the second example of the altered image. Different from the first example of falsification (see Figure 1.1), we can not visually distinguish two images at all even if both original and manipulated images are displayed simultaneously (see Figure 1.2). By using the redundancies existing in the digital image, the secret information is embedded into the original image (see Figure 1.2a) while degree of distortion from the new altered image (see Figure 1.2b) can not attract human's suspicion. By means of the image as the carrier, a secret message is transmitted



(a) The altered photo posted on the Polish subsidiary of Microsoft.



(b) The original photo appeared on Microsoft U.S. website.

Figure 1.1: Example of falsification.

between sender and receiver without worrying about the surveillance from a third party. In general, we define this information hiding technique as *steganography*. In this scenario, even though the original image is obtained, we can still not tell the difference between the faked image (see Figure 1.2a) and original image (see Figure 1.2b) by visual observation. In recent years, a large number of steganographic softwares appear on the Internet. If the steganography is used maliciously by the terrorists for transmitting secret information, it is possible that national defense security will be threatened.

To reestablish the trust to digital images, *digital image forensics* has emerged. Academically, it is proposed to generalize our research into three steps: raising a question, analyzing the question and solving the underlying problem. In the first step, we propose the questions as follows:

1. Where does the digital image come from?
2. Is the content of the digital image genuine?

In the second step, let us analyze the two proposed problems. We can classify the first problem into the field of *image origin identification*. In this scenario, we

(a) The altered photo embedded with secret information.

(b) The original photo acquired by a digital still camera.

Figure 1.2: Illustration of falsification.

mainly investigate a given digital image captured by what camera brand, model or even individual device. In fact, when we do research on image origin identification, it is assumed that we can not obtain any prior information about the digital camera, which definitely largely increases the difficulties of identification. Thus, our task is to extract intrinsic fingerprints existing in the digital image which characterize the unique digital still camera. Besides, we also need to authenticate if the inspected image is generated by the realistic digital still camera, not by the computer. The second problem can be classified into the field of *image content integrity*. In this scenario, by inspecting a given image, we question the authenticity of this digital image. The forensic research of image content integrity addresses two main problems: *image forgery authentication* and *steganographic image detection*. In virtue of the development of digital technology and software industry, more and more manipulation techniques appear which also increases the difficulties for image forensic researches. Figure 1.1 illustrates a splicing technique which can be investigated by the image forgery authentication. Detection of steganographic images studies whether an inspected digital image contains a secret information, as illustrated by Figure 1.2.

Academically, we denote the study of data hidden detection as *steganalysis*.

In the third step, let us specifically solve our proposed two problems. In this thesis, about image origin identification, we first mainly study how to distinguish computer generated (CG) images from photographic images (PIM). Then, this thesis investigates source camera identification based on natural images, respectively in RAW and JPEG format. Finally, image content integrity had been studied through the detection of Least Significant Bit (LSB) replacement steganography and image resampling detection.

1.2 Outline of the Thesis

The goal of this thesis is to investigate the challenges of above mentioned digital image forensic problems. More specifically, this thesis mainly studies the problem of image origin identification and the problem of image content integrity. Thus, the overall structure of the thesis is the following:

- Chapter 1 mainly introduces our research background and generally describes the problems studied in the thesis. Besides, it also presents in more details the outline of this thesis. Finally, all the publications from the author's doctoral work are listed.

- Chapter 2 provides an overview of current-art of digital image forensics. It is proposed to introduce the image processing pipeline which is the fundamental for studying image forensics. Then, we introduce in details a classification of image falsifications and describe each of them. Similarly, the digital forensics is also classified into two categories which are described in detail: *active forensics* and *passive forensics*. Finally, we provide an overview of current-art in steganography and steganalysis.

- Chapter 3 studies the discrimination between PIM and CG images. The proposed method exploits traces of CFA interpolation, present in PIM images, together with the use of hypothesis testing theory. By using the Likelihood Ratio Test (LRT), the method proposed to distinguish PIM from CG images warrants a prescribed False Alarm Rate (FAR) and achieves the maximal detection power. Experimental results show the efficiency of the proposed approach and the high robustness with respect to counter-forensic techniques.

- Chapter 4 explores the statistical detection of JSteg steganography. The approach is based on a statistical model of DCT coefficients challenging the usual assumption that among a subband all the coefficients are independent and identically distributed (i. i. d.). The hidden information detection problem is cast in the framework of hypothesis testing theory. In an ideal context where all model parameters are perfectly known, the LRT is presented and its performance is theoretically established. For a practical use where the distribution parameters are unknown, by exploiting an ad-hoc selection channel approach

in DCT coefficients, a detector based on the estimation of those parameters is designed. The loss of power of the proposed detector, compared with the optimal LRT, is small. Numerical results show the relevance of the proposed approach.

- Chapter 5 investigates the problem of identifying a source imaging device, possibly from the the same model, from natural images in RAW format. It exploits the proposed enhanced Poissonian-Gaussian model from which the noise parameters are considered as unique fingerprints and hence used to identify source camera device. Again, a similar approach is presented, that is the statistical detection problem is cast within the framework of hypothesis testing theory. Then, the LRT is presented and its performance is theoretically established. Finally, a practical test is proposed to deal with the unknown nuisance parameters: a practical test based on the estimation of those parameters is designed. Numerical results that support the relevance of our proposed approach are presented to conclude this chapter.

- Chapter 6 also addresses the problem of individual camera device identification for natural images but when the images are compressed using the JPEG standard. This approach is based on the generalized signal-dependent noise model but improves it by studying the relationship between noise model parameters. It is shown that, because of the non-linear response of pixels, the noise model parameters exhibit the linear relation. This relation is used to identify source camera device. Then, the same approach is proposed, and the problem is cast within the framework of hypothesis testing theory. The LRT is presented and its performance is theoretically established. Then, when the nuisance parameters are unknown, two practical tests are proposed based on the estimation of those parameters. Numerical results on simulated data and real natural images are presented to support the relevance of the proposed approach.

- Chapter 7 first explores the periodic characteristic of one-dimensional (1-D) resampled signal, which can be exposed by using linear parametric model. After dealing with the nuisance parameters, together with Bayes' rule, it is proposed to use the conditional probability of residual noises as the unique fingerprint of the resampled signal. Similarly, we study the resampling artifact of a two-dimensional (2-D) image. It is also proposed to establish the practical LRT for dealing with the problem of detecting a batch of resampled images. Numerical experiments show the relevance of our proposed algorithm with low complexity and high efficiency.

- Chapter 8 concludes this thesis and present the perspectives of future works opened in the field of digital image forensics.

1.3 Publications and Authors' Contribution

Most of the material presented in this thesis appears in the following publications that represent original work, of which the author has been the main contributor.

Journal articles

1. **T. QIAO**, F. Retraint, R. Cogranne, and C. Zitzmann "Steganalysis of JSteg algorithm using hypothesis testing theory", *EURASIP Journal on Information Security*, vol. 2015, no. 1, pp. 1-16, Mar. 2015.

Journal articles under review

1. **T. QIAO**, F. Retraint, R. Cogranne, and T. H. Thai, "Individual Camera Device Identification Based on RAW Images", submitted to *IEEE Transaction on Information Forensics and Security*, 2016.

2. **T. QIAO**, F. Retraint, R. Cogranne, and T. H. Thai, "Individual Camera Device Identification from JPEG Images", submitted to *EURASIP Journal on Signal Processing: Image Communication*, 2016.

Conference papers

1. **T. QIAO**, F. Retraint, R. Cogranne, and T. H. Thai, "Source camera device identification based on RAW images", in *IEEE International Conference on Image Processing, ICIP '15*, pp. 3812-3816, Sep. 2015.

2. **T. QIAO**, C. Zitzmann, F. Retraint, and R. Cogranne, "Statistical detection of Jsteg steganography using hypothesis testing theory", in *IEEE International Conference on Image Processing, ICIP '14*, pp. 5617-5521, Oct. 2014.

3. **T. QIAO**, C. Zitzmann, R. Cogranne, and F. Retraint, "Detection of Jsteg algorithm using hypothesis testing theory and a statistical model with nuisance parameters", in *2nd ACM workshop on Information hiding and multimedia security, IH&MMSec '14*, pp. 3-13, Jun. 2014.

4. **T. QIAO**, F. Retraint, and R. Cogranne, "Image authentication by statistical analysis", in *21st European Signal Processing Conference, EUSIPCO '13*, pp. 1-5, Sep. 2013.

Chapter 2

Overview on Digital Image Forensics

Contents

2.1 Introduction

As we discussed in Chapter 1, digital image forensics concerns two main problems: image origin identification and image content integrity. To address these problems, forensic researchers have developed a variety of image authentication algorithms to protect against attacks from image forgers. In general, it is proposed to classify all the forensic methodologies into two categories: *active forensics* and *passive forensics*. It should be noted that, generally speaking, steganalysis can be included within the field of digital image forensics; however due to its specificities, steganalysis is usually studied independently, therefore we can hardly categorize it into active or passive forensics. Active forensics involves forensic techniques which authenticate a digital image by using the prior-embedded relevant information such as a digital watermark or signature. Active forensics indeed restores the human credibility of digital image due to its powerful detection efficiency. However, since the embedding mechanism has to be available, active forensics has its limitation of a widely-adopted utilization. Besides, active forensics faces the other questions/problems like "what happens when several people merge their media ?" or "how to embed a robust information that can be retrieved regardless of the modification of the media an active attacker can perform ?" Therefore, passive forensics without embedding any prior information is the focus of the digital image forensic investigation.

In this chapter, we mainly provide an overview on the field of digital image forensics that has emerged over the past decades. The chapter is organized as follows. First of all, Section 2.2 illustrates the general procedure of image acquisition which lies at the foundation of the work presented in this thesis. In other words, the research fruits of this thesis are inspired by the pipeline of image acquisition. Section 2.3 presents an overview of the variety of usual image forgeries which should be addressed by researches from the image forensic community. Then, Section 2.4 briefly presents the development of digital image forensic techniques over the past decades. Overview of the current art in steganography and steganalysis is presented in Section 2.5. Finally, Section 2.6 concludes this chapter.

2.2 Image Processing Pipeline in Digital Cameras

When studying digital image forensics, the very first step is to understand the process of image acquisition. To this end, it is proposed to generally illustrate the image acquisition procedure of a digital camera (see Figure 2.1). When a scene is captured by a digital camera, photons radiating from the scene pass through the optical system (e.g. lens) and converge onto the focal plane. Then, Color Filter Array (CFA) filters the incident light spectrum. Because the cost for producing full-color sensors, for all three primary color components (red, green, and blue) is prohibitive, CFA filtering together with post-processing technique is a tradeoff between image quality and product cost. Subsequently, the image sensor converts it to electrical energy which is then converted to digital signal by an analog-to-digital (A/D) converter inside the camera. During the storing of electrical charges and

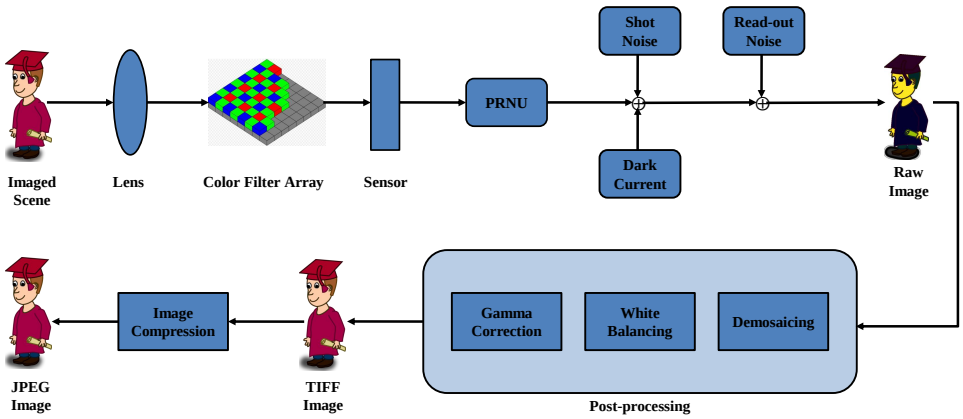


Figure 2.1: Illustration of image acquisition pipeline in a digital camera.

reading of voltage, a RAW image is unavoidably corrupted with random additive noises (shot noise, read-out noise, dark current) and multiplicative noise (Photo-Response Non-Uniformity noise, PRNU). Then, after several post-processing stages (such as demosaicing, white balancing, gamma correction), a full-color and high-quality image is obtained. This type of image is referred to as a Tagged Image File Format (TIFF) image in this thesis. The TIFF format is not very practical for transmission or storage. Thus, a lossy compression algorithm is very often used to reduce the image size. It is no doubt that JPEG compression can remarkably strike the balance between image quality and size.

Let us use a three-dimensional matrix with size $N_\mathrm{r} \times N_\mathrm{c} \times 3$ to represent a full-color image, where N_r denotes the number of rows and N_c the number of columns. It is proposed to utilize $c \in \{\mathrm{R}, \mathrm{G}, \mathrm{B}\}$ describing a color channel in which R, G and B respectively represent a red, green and blue color. In general, a digital still image is saved with l bits and each pixel value is described as a natural integer. Then, all the possible values of pixels can be denoted by a set $Z = \{0, 1, \ldots, K\}$ with $K = 2^l - 1$. Thus, an arbitrary image is definitely in the finite image space Z^N with $N = N_\mathrm{r} \times N_\mathrm{c} \times 3$. For clarity, let us denote $\mathbf{Z}$ as a digital image in RAW format and $\mathbf{X}$ an image in TIFF or JPEG format. $\mathbf{Z}^c$ denotes each color channel of the image $\mathbf{Z}$, where $z^c(m, n)$ with $1 \leq m \leq N_\mathrm{r}$, $1 \leq n \leq N_\mathrm{c}$ denotes a pixel value of the color channel c at the location (m, n). In the following subsections, let us briefly discuss the image acquisition in a digital still camera.

2.2.1 RAW Image Formation

A digital image consists of an optical system (e.g. lens), an image sensor and electronic system, which are very similar to the human visual system such as the eyes, retina and brain. Here, we mainly discuss the image sensor system consisting of a two-dimensional arrays of photodiodes, which is in charge of collecting photons from the scene after passing through the lens. In general, the image sensor can be

Figure 2.2: Illustration of samples from CFA.

classified into two categories: Charge-Coupled Device (CCD) and Complementary Metal-Oxide Semiconductor (COMS). It should be noted that the image sensor can convert light energy (or collected photons) to electrical energy while the output signal of the image sensor is still analog. Afterwards, an A/D converter enables the analog signal to become the digital signal, which is recorded as the digital image in RAW format. In general, an image in RAW data can be recorded with 12, 14 or 16 bits due to the different A/D converter equipped within the digital still camera. A RAW image preserves all the original information collected from the captured natural scene while does not go through the operation of post-processing. This advantage provides the photographer more flexibility for further adjustments by using the image-editing software.

Since the image sensor can not distinguish the light wavelength (or different color channels), a CFA pattern is overlaid on the image sensor. Figure 2.2 illustrates some samples of CFA. Among all the possible CFA patterns, the Bayer pattern is the most widely-adopted. Since the human eyes are more sensitive to green than red or blue light, the Bayer pattern contains 50% green while only 25% red and 25% blue. In practice, based on this human visual-sensitivity designed pattern, the acquired digital image can achieve the relevant level of image quality. However, there are few digital cameras adopting a full-color filter which records all the information from three color channels (e.g. *Sigma SD9* or *Polaroid x530*). Due to the high production cost, this technique has not been widely-adopted. Thus, we do not discuss this case in this thesis.

It is proposed to define $\mathbf{Z}$ as an image in RAW format. Due to CFA filtering, the single-channel image $\mathbf{Z}$ is represented by a two-dimensional matrix with size $N_{\mathrm{r}} \times N_{\mathrm{c}}$. It should be noted that each color channel c only records one pixel value of the RAW image $\mathbf{Z}$. Therefore, the pixel value can be represented by:

$$z^c(m,n) = \begin{cases} z(m,n) & \text{if} \quad \mathbf{P}_{\mathrm{CFA}}(m,n) = c \\ 0 & \text{otherwise,} \end{cases} \tag{2.1}$$

where $\mathbf{P}_{\mathrm{CFA}}$ denotes the CFA pattern.

However, during the procedure of the RAW image formation, it is unavoidable that several noise sources are introduced. The first category of random noise consists of Poisson-distributed noise (or shot noise) and dark current. Due to the photo-counting process, the Poisson-distributed noise is generated. The thermal energy in the absence of light leads to dark current. All the remaining electronic

noises including read-out noise can be categorized into the second random noise source. Typically, the distribution of read-out noise can be modelled by the Gaussian model with zero mean. In fact, there also exists another kind of noise source, referred to as a multiplicative noise source, which does not belong to the category of random noise. Due to imperfections during sensor manufacturing process and non-uniformity of photo-electronic conversion caused by inhomogeneity of silicon wafers, it is unavoidable that Sensor Pattern Noise (SPN) is generated, which includes two main components: Fixed Pattern Noise (FPN) and PRNU.

2.2.2 Post-processing

It should be noted that due to the property of the CFA pattern, only one pixel value from one color channel (R, G or B) is recorded at the pixel location (m, n), which leads to the incomplete information for a digital image. Therefore, it requires CFA interpolation or demosaicing [2] to fill up the missing color values, which allows to estimate all the zero values in the sub-images. Technical speaking, interpolation technique utilizes the neighourhood pixel values to estimate a missing pixel value. Let us briefly categorize the CFA demosaicing algorithms into: *non-adaptive* and *adaptive*. Non-adaptive algorithm, such as *Nearest-neighbor*, *Bilinear* and *Bicubic* algorithm, arbitrarily uses the same interpolation algorithm dealing with all the pixels. For instance, let us describe the *Bilinear* algorithm as:

$$\mathbf{Z}_{\mathrm{D}}^{c} = \mathbf{F}_{\mathrm{D}}^{c} * \mathbf{Z}^{c}, \tag{2.2}$$

where the symbol $*$ denotes the two-dimensional convolution, and $\mathbf{Z}_{\mathrm{D}}^{c}$ denotes the demosaiced image from the color channel c, and $\mathbf{F}_{\mathrm{D}}^{c}$ represents the linear filter of the color channel c. In general, the interpolation algorithm is independently applied in three different color channels. Thus, the linear filter $\mathbf{F}_{\mathrm{D}}^{c}$ is designed for R, G and B as:

$$\mathbf{F}_{\mathrm{D}}^{\mathrm{G}} = \frac{1}{4}\begin{pmatrix} 0 & 1 & 0 \\ 1 & 4 & 1 \\ 0 & 1 & 0 \end{pmatrix}, \qquad \mathbf{F}_{\mathrm{D}}^{\mathrm{R}} = \mathbf{F}_{\mathrm{D}}^{\mathrm{B}} = \frac{1}{4}\begin{pmatrix} 1 & 2 & 1 \\ 2 & 4 & 2 \\ 1 & 2 & 1 \end{pmatrix}. \tag{2.3}$$

In fact, the non-adaptive algorithm can deal with the missing pixel values in smooth regions of the image while does not perform well in textured regions or edges. Thus, it is proposed to use the adaptive algorithm which allows to estimate the more accurate missing pixel values with minimizing the demosaicing errors, based on the different regions of the image. Since missing pixels are estimated by using the neighbourhood information, CFA demosaicing leads to the linear correlation among adjacent pixels, which can be amplified in the spatial domain during the following post-processing operations.

After CFA demosaicing, it requires furthermore to improve the image visual quality which allows the demosaiced image $\mathbf{Z}_{\mathrm{D}}^{c}$ to pass through another post-processing technique: white balancing [2]. Because of the color temperature diversity from the light sources, which leads to the shift of the reflection spectrum of the object

from the true color, the captured object without white balancing in the image possibly appears different colors when different light sources cast onto the object. For instance, in the case that a light source with low color temperature illuminates a white object, the captured object in the image appears reddish. In the case that a light source with high color temperature illuminates that same white object, the captured object in the image become bluish. Although the digital still camera can not directly distinguish the color changes in different color temperature, the human visual system is smart enough to make a distinction under different cases which is called *color constancy*. Therefore, by compensating the illumination imbalance, white balancing is a very important step which allows to render the captured object appearing white as what the human visualizes. Then, let us briefly introduce a white balancing algorithm named *Gray World*, which assumes that the average value of each color channel c will average into a common gray value:

$$\overline{z}_{\mathrm{D}}^{\mathrm{R}} = \overline{z}_{\mathrm{D}}^{\mathrm{G}} = \overline{z}_{\mathrm{D}}^{\mathrm{B}}, \tag{2.4}$$

where $\overline{z}_{\mathrm{D}}^{c}$, $c \in \{\mathrm{R}, \mathrm{G}, \mathrm{B}\}$ denotes the average intensity of the demosaiced image $\mathbf{Z}_{\mathrm{D}}^{c}$, which is formulated as:

$$\overline{z}_{\mathrm{D}}^{c} = \frac{1}{N_{\mathrm{r}} \cdot N_{\mathrm{c}}} \sum_{m=1}^{N_{\mathrm{r}}} \sum_{n=1}^{N_{\mathrm{c}}} z_{\mathrm{D}}^{c}(m,n). \tag{2.5}$$

Then, let us define g_{W}^{c} as the gain factor from the color channel c for white balancing. Since the human visual system is more sensitive to the green channel, this algorithm assumes that the gain factor of the green channel $g_{\mathrm{W}}^{\mathrm{G}}$ is a constant. The other gain factors are formulated by:

$$g_{\mathrm{W}}^{\mathrm{R}} = \frac{\overline{z}_{\mathrm{D}}^{\mathrm{G}}}{\overline{z}_{\mathrm{D}}^{\mathrm{R}}}, \quad \text{and} \quad g_{\mathrm{W}}^{\mathrm{B}} = \frac{\overline{z}_{\mathrm{D}}^{\mathrm{G}}}{\overline{z}_{\mathrm{D}}^{\mathrm{B}}}. \tag{2.6}$$

Finally, the white-balanced image $\mathbf{Z}_{\mathrm{W}}^{c}$ is written by:

$$\mathbf{Z}_{\mathrm{W}}^{c} = g_{\mathrm{W}}^{c} \cdot \mathbf{Z}_{\mathrm{D}}^{c}. \tag{2.7}$$

In fact, due to the estimate or selection of the appropriate gain factors for three color channels, it is not very easy for white balancing. The digital still cameras usually select the gain factors depending on the prior knowledge of light sources, referred to as some typical scene lights such as *daylight*, *indoor* and *outdoor* stored in the processor of the camera.

After CFA demosaicing and white balancing, the pixel intensity only appears the linear characteristic. Therefore, it requires the gamma correction [2] to compensate the non-linear property for better visually displaying. To adjust the brightness with former CTR (Cathode Ray Tube) display device and render the luminance into a perceptually uniform domain, it is proposed to use the gamma correction technique during the post-processing acquisition, which is a pixel-wise non-linear operation defined as:

$$z_{\mathrm{G}}^{c}(m,n) = \left(z_{\mathrm{W}}^{c}(m,n)\right)^{\frac{1}{\gamma}}, \tag{2.8}$$

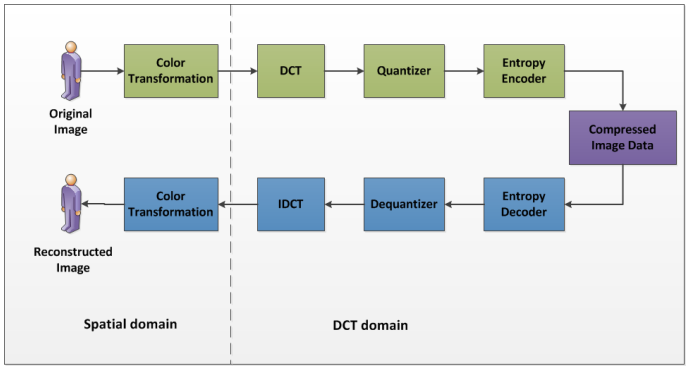


Figure 2.3: Illustration of the JPEG compression pipeline.

where the gamma factor γ is typically set as 2.2. Till this step, a full-color image in TIFF format has been created, which is denoted as $\mathbf{X}_{\text{TIFF}}$. In the following subsection, let us mainly discuss how to convert a TIFF image $\mathbf{X}_{\text{TIFF}}$ to a JPEG image $\mathbf{X}_{\text{JPEG}}$.

2.2.3 Image Compression

In general, the TIFF format is not very practical for transmission or storage. Therefore, a lossy compression algorithm is usually proposed to reduce the image data size. Among many lossy compression algorithms, JPEG compression can remarkably strike the balance between image visual quality and size. Typically, the JPEG compression scheme includes three primary settings: color space, sub-sampling technique, and quantization table. The fundamental steps of the JPEG compression are illustrated in Figure 2.3. Since the JPEG compression scheme works in three different color spaces, referred to as YCbCr instead of RGB color space, three color components from R, G and B color channel are first transformed to the corresponding ones from Y, Cb and Cr color channel. In this case, the channel Y denotes the luminance of the pixel value, and the chrominance is respectively represented by the channels Cb and Cr. It should be noted that the transformation of each channel works independently. Besides, by improving the efficiency of image compression, the channels Cb and Cr are respectively down-sampled by a factor of 2 horizontally and vertically. Let us formulate the specific transformation by:

$$\begin{pmatrix} \text{Y} \\ \text{Cb} \\ \text{Cr} \end{pmatrix} = \begin{pmatrix} 0.299 & 0.587 & 0.114 \\ -0.169 & -0.331 & 0.5 \\ 0.5 & -0.419 & 0.081 \end{pmatrix} \begin{pmatrix} \text{R} \\ \text{G} \\ \text{B} \end{pmatrix} + \begin{pmatrix} 0 \\ 128 \\ 128 \end{pmatrix}. \tag{2.9}$$

Typically, the JPEG compression algorithm mainly contains two steps: Discrete Cosine Transform (DCT) and quantization. In each 8×8 block of each color channel, the DCT operation converts the pixel value in the spatial domain to the correspond-

ing coefficients in DCT domain by using:

$$I(u,v) = \frac{1}{4} T_u T_v \sum_{m=0}^{7} \sum_{n=0}^{7} x_{\text{TIFF}}(m,n) \cdot \cos\left(\frac{(2m+1)u\pi}{16}\right) \cos\left(\frac{(2n+1)v\pi}{16}\right), \tag{2.10}$$

where for simplicity and clarity, x_{TIFF} from the YCbCr color space denotes a pixel value in a 8×8 block, $0 \leq m, n \leq 7$, $I(u,v)$ represents the coefficient in DCT domain, $0 \leq u, v \leq 7$, and T_u denotes the normalized weight

$$T_u = \begin{cases} \frac{1}{\sqrt{2}} & \text{for } u = 0 \\ 1 & \text{for } u > 0 \end{cases}. \tag{2.11}$$

It should be noted that DCT operation works independently in three color spaces Y, Cb and Cr. The Direct Current (DC) coefficient representing the mean value of pixels in a 8×8 block from the spatial domain locates at the position $(0,0)$ in DCT domain. The Alternating Current (AC) coefficient refers to as the remaining 63 coefficients in the 8×8 block. In general, the main energy of the image from the low-frequency components concentrates in the upper left corner of 8×8 grid representing the background of the image in the spatial domain. On the contrary, the DCT coefficients from the high-frequency components which is not visually important represent the details of the image. Besides, it should be noted that the DCT operation is the lossless process.

Afterwards, the DCT coefficients go through the lossy compression operation, referred to as the quantization. By dividing each unquantized DCT coefficient using the quantization step and rounding it, the quantization operation is formulated by:

$$D(u,v) = \text{round}\left(\frac{I(u,v)}{\Delta(u,v)}\right), \tag{2.12}$$

where $D(u,v)$ denotes the quantized DCT coefficient and $\Delta(u,v)$ the corresponding quantization step which constitutes the 8×8 quantization table.

After quantization, the Run-length Encoding (RLE) algorithm deals with the quantized DCT coefficients in order of the zig-zag sequence. Then, the entropy coding algorithm (e.g. Huffman coding) is employed. Finally, the compressed image data is obtained.

The process of the JPEG decompression is briefly illustrated in Figure 2.3, which works in the reverse order: entropy decoding, dequantization and Inverse DCT (IDCT). First of all, the entropy decoder extracts the quantized DCT coefficients $D(u,v)$, which is multiplied by the corresponding quantization step $\Delta(u,v)$

$$I_d(u,v) = \Delta(u,v) \cdot D(u,v), \tag{2.13}$$

where $I_d(u,v)$ denotes the dequantized DCT coefficient. Then, let us reconstruct the image in the spatial domain by transforming the dequantized DCT coefficients

using IDCT operation

$$x_{\mathrm{IDCT}}(m,n) = \sum_{u_{\mathrm{h}}=0}^{7}\sum_{u_{\mathrm{v}}=0}^{7}\frac{1}{4}T_uT_vI_d(u,v) \cdot \cos\left(\frac{(2m+1)u\pi}{16}\right)\cos\left(\frac{(2n+1)v\pi}{16}\right). \tag{2.14}$$

Subsequently, the components in the YCbCr color space are respectively transformed to the corresponding pixel values in the RGB color space. Typically, the pixel values are possibly not integers or beyond the finite dynamic range (e.g. $[0, 255]$). Therefore, the rounding and truncation operation is absolutely necessary

$$x_{\mathrm{JPEG}}(m,n) = \mathrm{trunc}\Big(\mathrm{round}\Big[x_{\mathrm{IDCT}}(m,n)\Big]\Big), \tag{2.15}$$

where the JPEG pixel $x_{\mathrm{JPEG}}(m,n)$ from the image $\mathbf{X}_{\mathrm{JPEG}}$ is different form the TIFF pixel $x_{\mathrm{TIFF}}(m,n)$ from the image $\mathbf{X}_{\mathrm{TIFF}}$ because of the quantization, rounding and truncation errors.

A full-resolution image carries all the intrinsic characteristics generated by an image acquisition process. Any manipulation on a natural image unavoidably introduces the change of those intrinsic characteristics. Most image forensic tools make use of the characteristics left by image acquisition pipeline to perform digital image forensics. In this thesis, it is proposed to investigate mainly the following processes of an image acquisition: CFA filtering, demosaicing, image compression, random noise (e.g. shot noise and read-out noise), the post-camera[1] operation (e.g. resampling, steganography). It should be noted that the post-camera manipulation of an image is operated after acquiring a digital still image while the post-processing of the image acquisition is executed, using within the camera, before obtaining a natural image that can be interpreted visually.

In fact, the study of digital image forensics has very close linkage with image acquisition and post-camera operations. Describing the whole image acquisition pipeline, as done in this section, is a first natural step. The following step is to define what image forgery is. Since the ultimate target of digital image forensics is to restore the credibility of a digital image, it is crucial to understand the different type of image forgery techniques. Hence, in the following section, it is proposed to solve the following questions:

1. What is a falsified image?
2. How many categories are there among different image tempering techniques?

2.3 Image Falsification Classification

In this section, we mainly discuss what image tempering is and how many tempering techniques exist. In the context of digital images, tempering means that one intentionally manipulates a digital image and, hence, modifies the semantic meaning of

[1]The term *post-camera* is referred to as the operation after acquiring a digital image.

(a) The three clocks in the building look very similar to each other, especially with the same size.

(b) The model of three traffic lights on the road has no differences. Two of them have been scaled.

Figure 2.4: Example of copy-move tempering.

the image for malicious purposes. In fact, it is very common that one can use the photo-editing software to modify unsatisfying images taken by a digital still camera. If the goal of such modification is not to distort the truth, creating a scene that never has been, and possibly transmitting this image publicly, then, the modified image can not be defined as tempered one. On the opposite, in this thesis we adopt the definition of image forgery as the modification of an image with the purpose to alter the fact that the scene capture and to distribute this image publicly.

Due to the advent of the high-performance computers or smart cellphones, one can manipulate a digital image with relative ease. Additionally, ubiquitous photo-editing tools result in a variety of image falsification techniques. In digital image forensic community, it is proposed to define the manipulation technique which maliciously modifies the original realistic image as forgery. In fact, it is difficult to classify all the image tempering techniques. With the development of information technology, there is no doubt that there will spring up more and more image editing techniques. Therefore, in this thesis, it is proposed to generally illustrate several prevalent image tempering techniques.

2.3.1 Copy-move

Copy-move is the most widely-adopted manipulation technique. This technique denotes that one object in an image is first copied, then pasted within the same image but at a different location. In general, copy-move forgery intends to achieve two goals: duplicating the object and hiding the truth. By using the surrounding elements, a forger may cover a figure or an object within the photo that it is aimed at removing. To confuse the public and transmit wrong information, a forger can also make one or more duplicated copies from the original realistic object within the image. Two typical examples are illustrated in Figure 2.4. Two clocks which are the copies of the original clock appear at different positions of the same building

(a) A spliced photo appears on the front page of the *Los Angeles Times* in 2003.

(b) The first original photo.

(c) The second original photo.

Figure 2.5: Example of splicing tempering.

in Figure 2.4a. In fact, we have no clue that which the original one is or which the duplicates are. A more complicated copy-move technique is displayed in Figure 2.4b. Two duplicated traffic lights are scaled, a little different from the original one. Obviously, the forger considers the perspective effect. Therefore, it is necessary that forensic researchers design copy-move detectors to authenticate a copy-move tempered image, and even locate duplicated objects within a given image.

2.3.2 Splicing

Splicing is another common manipulation technique that duplicates one or more objects from a first image and copy them into a second image. Since the new forged image is made up from disparate elements from two or more original real images, this tempering technique can be also defined as composite forgery. Different from copy-move tempering, the spliced object is from the different image. A famous example of image splicing is illustrated in Figure 2.5. The solider in Figure 2.5a originates from Figure 2.5b while the civilian holding a child in Figure 2.5a is spliced from Figure 2.5c. In fact, for increasing the credibility of the composite image, the forger usually utilizes some background from two real original images, see Figure 2.5b and 2.5c. Besides, The forged photo illustrated in Figure 1.1 is also a very typical example of a spliced image. Thus, the problem of detecting splicing forgery receives relevant

(a) An original image.

(b) An image forged by enlargement of 120% involving resampling.

Figure 2.6: Example of resampling tempering.

focus in the digital image forensic field.

2.3.3 Resampling

To create convincing forged images, a forger needs to use some geometric operations which require a resampling of pixels. Resampling is a post-camera processing technique. In general, it is proposed not to classify resampling among the image tempering technique. However, resampling of pixels is the basis of many forgeries because it is involved in almost all geometric modification (rotation, enlargement, etc. ...). While the evidence of resampling operation within an image does not necessarily imply that the image has been tempered it is a very important clue. Thus, resampling detectors play a very important role as an auxiliary forensic tool. Additionally, resampling falsification often happens together with other forgery techniques, for instance, copy-move or splicing forgery. For instance, in order to enhance the visual realism, splicing and copy-move forgeries often require scaling and/or rotation operations which requires pixels resampling, see Figure 2.4b.

A typical example is illustrated in Figure 2.6. The original image shown in Figure 2.6a is first resized, through an enlargement of 120%, and then cropped to obtain an image of the same size as the original; the result is shown in Figure 2.6b. In Chapter 7, we mainly discuss the problem of detecting a resampled image.

2.3.4 Brightness modification

In the process of image acquisition, gamma correction is a typical technique that modifies the brightness of an image to enhance the display on a screen (rendering realistic colors). From a forensics point-of-view, however, by altering the brightness or illumination within an image one can often emphasize on a portion of the image in order to modify the semantic purpose. Another usual example of brightness modification is related to copy-move or splicing forgery after which, to enhance visual realism, the brightness of the altered region must be adjusted to remain

(a) An original image published on the *Newsweek*.

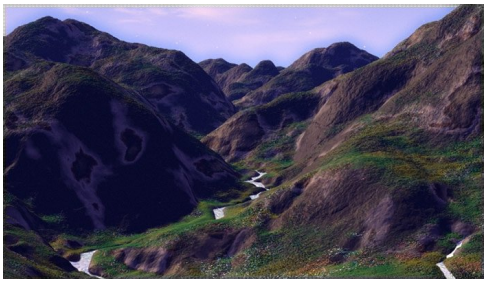
(b) A faked image, with a modification of the brightness, published on the *Times*.

Figure 2.7: Example of brightness modification tempering.

consistent with the original image. Like resampling, brightness modification can hardly be considered as a forgery alone, but it is a usual operation that is used along with other types of forgeries. Thus, brightness tempering is also an auxiliary forensic tool that provides important clues. An example of image tampering through brightness modification is presented in Figure 2.7. The original image published on the *Newsweek* is shown in Figure 2.7a and a tempered version of this image, with a much smaller brightness, has been published on the *Times* and shown in Figure 2.7b). This brightness operation obviously changes the display effect which probably transmits wrong information in the public. Therefore, the investigation of image brightness modification is crucial.

2.3.5 Median Filtering

Median filtering is a widely-adopted denoising operator because it is simple and has the ability to preserve rather well the edges. Different from most linear post-camera techniques such as resampling, the median filter is a non-linear operator which, hence, cannot be detected by most of the state-of-the-art forensic detectors that are often designed on the underlying assumption that modification is linear. Thus, the detection of median filtering receives more and more attention in the digital image forensic community.



(a) A computer generated image (CG).

(b) A photographic image (PIM) captured by a digital still camera.

Figure 2.8: Example of comparison between CG and PIM.

2.3.6 Double Compression

Image lossy compression (e.g. JPEG compression) is a general operation in the imaging pipeline, see Figure 2.1. In general, JPEG compression is the most popular image format due to its good tradeoff between efficiency, or image resulting size, and image quality. Most of the consumers' digital camera only output images already compressed using the JPEG standard and do not offer the ability to export uncompressed image. Then, when such a JPEG image is modified by using the post-camera operation, the forged image will likely be saved once again using the JPEG format. Thus, the resulting tempered image has, all in all, been compressed twice; once in the operation of image acquisition and twice in the post-camera processing. Therefore, the detection of double compression interests many forensic researchers as even though it is not really a type of forgery it provides an important clue like resampling and brightness modification.

2.3.7 Computer Generated

With the help of modelling and rendering software, an image forger can easily generate a computer graphics looking very like a real photographic image. By transmitting computer generated (CG) images, the forger can scratch a scene which actually does not exist in the real world. As shown in Figure 2.8a, the scene generated by a computer plausibly exits on earth and the image is rather realistic. The computer generated image usually simulates a three dimensional model that resembles to a realistic environment using 3D modelling and rendering software. Since the procedure of generating CG does not go through the pipeline of image acquisition, the problem of authenticating CG images generally focuses on how to effectively extract intrinsic features that reveals the use of the acquisition pipeline from a digital image.

To restore the trust to digital images, forensic researchers have the responsibility for designing robust detectors to fight against falsification of digital images. In the following section it is proposed to review the main digital image authentication methods that have been proposed over the past decades.

2.4 Overview on Digital Image Authentication Methods

To prevent digital image falsification, forensic analysts have designed efficient detectors to help the public restore the trustworthiness to digital images. In general, image authentication methods can be classified into two categories: active forensics and passive forensics. Active forensics refers to the techniques which authenticate a digital image with the use of an embedded digital watermark or a cryptographic signature before; this process requires to process the digital image before its transmission/sharing in order to embed the watermark. A large effort has been done in the research community in this direction and the problem encountered by active forensics is typically to ensure high robustness of the watermark against a wide range of attacks, such as geometric and collusion attacks to cite few [3].

Passive forensics refers to the techniques which inspect a digital image without any prior embedded information. The goal of passive forensics is, generally speaking, very similar to those of active forensics, however because the available information is different, the methodologies also differ dramatically. The problem of passive forensics is mostly related to the finding of "bullet scratches", also referred to as "intrinsic fingerprints", in order to get some information about the image and its acquisition. In addition, Image hashing technique, which extracts a short sequence representing the content of the whole image, can also help forensic researchers detect the altered image. For instance, if the content of the image is modified, the corresponding hash (a short sequence) will be changed as well. It also should be noted that different from hash functions in cryptography, referring to as MD5 or SHA-1, changing very sensitively resulted from the input data, image hashing requires its extracted sequence is only sensitive to malicious image tempering while performing robustly to innocent operations.

Besides, in recent years, forensic community has proposed a new definition to reflect better the current-art in the field of information security: the "*counter-forensics*". Generally, forensic investigators assume that the digital image forger has no knowledge about digital image forensics. Most of the proposed authentication algorithms are valid under that assumption.

If, however, a forger is aware of the current-art in the image forensics, one may try to modify how the medium is modified in order to escape from forensic detection or identification. This very concept of cleverly adapting the forgeries in order to fool the forensics detector is the field of "*counter-forensics*". The existence of counter-forensics opens a cat and mouse game in the sense that forensic investigators always aim at designing more robust authentication algorithms to deal with the possible use of counter-forensic attacks. On their side, counter-forensic researches always try to find more efficient methods to prevent detection of forgeries in order not to being detectable by possible forensic examination.

In the following subsections, we first briefly introduce active forensics and discuss its main advantages and disadvantages. Then, the detailed state-of-the-art on passive forensic methods is proposed. Finally, counter-forensics is briefly discussed to expose the current limits of image forensics.

2.4.1 Active Forensics

Active forensic methods rely on digital watermarking techniques to authenticate the inspected digital image. Digital watermarking is referred to as the procedure of embedding a watermark or a signature into a multimedia carrier (e.g. image, video or audio) [4]. If the receiver of the multimedia carrier can not successfully extract an untempered version of digital watermark, active forensic analysts can declare that the multimedia carrier has been manipulated [5]. In general, it is proposed to satisfy three requirements when designing a watermarking scheme [4]:

- Robustness: the capacity to extract the embedded data (digital watermark or signature) after application of post-camera processing;
- Imperceptibility: the watermarked content is perceptually indistinguishable from the original multimedia carrier;
- Capacity: the amount of the embedded data.

For a practical application of active forensics, a complete digital watermarking scheme needs satisfy those three requirements. However, depending on the targeted application of the digital watermark one may focus on one of those requirement. It is thus usually proposed to classify digital watermarking systems into the following categories [4, 6]:

- Copyright protection: in scenarios in which unauthorized parties claim multimedia content ownership, robust watermarking is proposed to protect the copyright of the medium, even in the case in which the medium has been subject to important modification (such as removing, forging) from unauthorized parties. In this case, large capacity is not required, but imperceptibility and robustness (with respect to geometric modifications) must reach acceptable levels [7].
- Copy prevention control: when multimedia content is distributed at a large scale, the content owner may need efficient methods to avoid the illegal duplication and/or re-transmission by unauthorized parties. Therefore, a robust digital watermark, or fingerprint, is embedded into the multimedia carrier in order to identify each user that gets access to the medium. The ultimate goal may prevent the copy of the medium or, more often, to identify sources of leakages. In this scenario, imperceptibility, capacity and robustness (with respect to collusion attacks) are mainly concerned.
- Tempering detection: when a digital image is maliciously forged, its very integrity is attacked. In this scenario, active forensic researchers can accurately detect whether or not the inspected image is tempered using a fragile/semi-fragile watermark, and even locate the portion of tempering [8, 9].

In the digital forensic community, since digital watermarking techniques can be applied in the field of copyright protection, copy prevention control, and image tempering detection, watermarking techniques would indeed help reestablish the trustworthiness to digital images. It, however, has some unavoidable disadvantages. A digital watermark or signature needs to be embedded into a multimedia carrier before its transmission, which limits its utilization. It should be noted that active forensic algorithms are valid only in the scenario in which a digital watermark or signature is pre-embedded into the multimedia carrier. Thus, forensic researchers have been seeking for another type of forensic technique, passive forensics, to address the general case when no watermark is embedded within a given medium.

All the methods proposed in this thesis belong to the passive forensics. We will thus, from now, mainly focus on passive forensic methods and, for simplification, digital forensics is sometimes used to refer to passive forensics. Beside, we also mainly discuss the methodologies of natural image authentication, source camera identification, and image resampling detection which are more specifically the fields in which are the core contributions of the present thesis.

2.4.2 Passive Forensics

As we have discussed in Section 1.1, digital image forensics addresses two main problems: image origin identification and image content integrity.

2.4.2.1 Image Origin Identification

In general, when a photographer captures a digital image with a camera, the image is stored with a file header that provides meta-data about the image. Similar to a black box in an airplane, the header files (e.g. Exchangeable Image File, EXIF and Joint Photographic Experts Group, JPEG headers) contain all recording and compression history such as camera model, exposure, date and time of acquisition. Thus, forensic analysts can readily obtain all the information on the inspected image by extracting its header. Similar to a digital watermark, file header are extrinsic fingerprints that can used to solve the problem of identifying source camera. In this scenario, the problem of passive forensics is simply converted to active forensics. The original header, however, can be easily removed or replaced if the image is processed by the post-camera operation[2]. Besides, a mass of photos shared on the Internet, especially on the social web-site such as Facebook or Twitter, do not have their header files. Therefore, in the passive forensic research, we do not use file header as reliable fingerprints for identifying source camera. In the image acquisition pipeline, as illustrated in Figure 2.1, each processing step probably leaves some traces in the image. The passive algorithms of image origin identification make full use of the traces as intrinsic fingerprints to extract information about the history of a given

[2]For instance, the *exiftool* software and the *libexif* C library can easily be used to remove and replace header information stored with the EXIF standard.

image (acquisition device, compression, etc. ...). Any proposed passive approach for image origin identification has to solve the following questions [10]:

- Which intrinsic/in-camera[3] fingerprints can be used for identifying image origin?
- How to extract accurately these fingerprints from an inspected image?
- Under which framework can the proposed detector distinguish the fingerprints from images acquired by different sources[4] ?

Although the image acquisition of different camera sources operates very similarly, distinguishable in-camera fingerprints can still be extracted from some steps of the image processing pipeline. Nevertheless, a fingerprint should satisfy the following four basic requirements:

- Generality: the proposed fingerprint should exist in every image.
- Uniqueness: the proposed fingerprint should be distinguishable for different camera brands/ models/individual devices; that is two sources must share the same fingerprint while any two different sources must have different fingerprints.
- Invariant: the proposed fingerprint should be independent of the image content, but must greatly dependent on the source camera.
- Robustness: the proposed fingerprint should survive the linear operation such as demosaicing and the non-linear operation such as lossy compression or gamma correction.

Generally speaking, it is necessary for forensic analysts to extract the in-camera fingerprints from each given image in order to achieve the target of passive forensics. Based on different applications of identifying source camera brand, model or individual device, the uniqueness of the fingerprint is also very important. For camera brand/model identification, it should be noted that the in-camera fingerprints are shared among cameras of the same brand/model but distinguishable for different camera brands/models. While for identifying the different individual camera instances, the fingerprint should be unique for each device. The latter is a challenging problem since the property of the fingerprints extracted from different devices of same camera model are probably very similar and, hence, largely increases the difficulty of identification.

[3]The term *in-camera* is referred to as all the whole image acquisition pipeline (see Figure 2.1), which does occurs before the image is output from the camera; it is the opposite in that sense to the post-camera operations.

[4]The term *source* means an individual camera device, a camera model, or a camera brand. Other sources such as scanners are not addressed in this thesis.

About the problem of image origin identification, many forensic investigators have been proposed, which can be split generally into two categories: *learning-based* methods, based on supervised classification, and *statistical model-based* methods. *Learning-based* approaches, using supervised classification methods for a vast majority [11–13], has been widely-adopted by most of authentication detectors due to its relevant detection results and easy operation [14–19]. *Statistical model-based* methods, which often exploit *hypothesis testing* [20] in this thesis, is also utilized by some forensic methodologies [21–25].

A supervised classifier generally works in three steps: features extraction, training step and testing step. During the features extraction step, all the images are processed to extract some features which are expected to be sensitive to the camera model (or the source). The features extraction step, hence, essentially consists in a mapping from the set of image pixels $\mathcal{Z}$ to a features space, denoted $\mathcal{F}$; for a features set of M_f, the feature space is usually a subset of $\mathbb{R}^{M_f}$. Before applying the training step, one requires to apply the features extraction step on all the images from a training set, that is for which the source is known. Let us denote $\mathcal{F}^M$ the features extracted from all the images from the training set; $\mathcal{F}^M$ can be classified into disjoint sets, each gathering features from the same source. M' different sources are defined as $\{\mathcal{S}_1, \mathcal{S}_2, \ldots, \mathcal{S}_{M'}\}$ which are elements of classification. Technically, each element $\mathcal{S}_m$, in which $m \in \{1, \ldots, M'\}$, is a subset of $\mathcal{F}^M$. For clarity, it is assumed (without loss of generality) that each element $\mathcal{S}_m$ is made from K images for which the feature vectors are denoted $\mathbf{v}_{m,k}$ with $m \in \{1, \ldots, M'\}, k \in \{1, \ldots, K\}$. In the training step, by using supervised machine learning algorithms, a typical classifier is established based on the set of the feature vectors $\{\mathbf{v}_{m,k}\}$. The supervised classification scheme essentially learns a mapping between the feature space $\mathcal{F}$ and the sources $\{\mathcal{S}_1, \mathcal{S}_2, \ldots, \mathcal{S}_{M'}\}$ which ideally classifies an input feature vector into its corresponding source. Thus, in the testing step, the feature vectors are extracted is the same way as in prior to the training step. Then they are put into the trained classifier to achieve the supervised classification. The most popular supervised classification scheme is Support Vector Machine (SVM) [12, 13]. The details of its implementation can be referred to as [26]. Supervised classification scheme, however, unavoidably involves two main drawbacks which somewhat limit its application. During its training step, the scheme first requires a wide rage of different types of samples (e.g. images) from various sources. In practice, it is probably hardly possible to gather a wide enough dataset to capture the huge diversity of image categories for all imaging device. Besides, establishing the statistical performance of the trained supervised classifier remains an open problem in the machine learning community [27]. In the following, let us briefly present the other approach for camera identification: statistical model-based methods based on hypothesis testing.

Let us denote $\mathbf{Z}$ an inspected image and $\mathcal{S}_0$ and $\mathcal{S}_1$ the two sources which it is aimed at identifying. In this scenario, forensic analysts make a choice between the

two following hypotheses:

$$\begin{cases} \mathcal{H}_0 : \mathbf{Z} & \text{is taken by the camera source } \mathcal{S}_0 \\ \mathcal{H}_1 : \mathbf{Z} & \text{is taken by the camera source } \mathcal{S}_1 \text{ different from camera source } \mathcal{S}_0. \end{cases}$$

It should be noted that two different camera sources $\mathcal{S}_0$ and $\mathcal{S}_1$ might refer to as brands/models/individual devices. In this thesis, it is proposed to establish two types of practical detectors for source camera identification. Assuming that the in-camera fingerprints from $\mathcal{S}_0$ and $\mathcal{S}_1$ are both prior-known, then the statistical tests we propose address the problem of identifying whether the given image $\mathbf{Z}$ is acquired by camera source $\mathcal{S}_0$ or by camera source $\mathcal{S}_1$. In another scenario, in which only the fingerprints from $\mathcal{S}_0$ are known (without knowing the other possible source $\mathcal{S}_1$), the proposed statistical test investigates whether the inspected image $\mathbf{Z}$ has been captured by camera source $\mathcal{S}_0$. When the source identification problem is cast within the framework of hypothesis testing theory, it may be possible to design a statistical test which is optimal with respect to a given criterion and it may be possible to establish theoretically the statistical performance of this test; which is of crucial interest for warranting a prescribed false alarm probability. Moreover, this methodology may also allow us to establish some bounds on the statistical performance on can achieve. To the best of our knowledge, in the passive forensic community, few investigations have explored the hypothesis testing theory. We will extend our discussion in Chapter 5 and 6.

Generally speaking, the problem of source camera identification is addressed only for photographic images, that have been acquired by a real camera. To extend the problem of image origin identification, it is proposed to identify whether the given picture is a realistic photographic (PIM) image captured by a digital device or a computer generated (CG) one by using a graphic software (see Figure 2.8). In the passive forensic community, the problem of discriminating PIM from CG images, also referred to as natural image authentication, attracts attention and has been widely studied. Similar to the problem of source camera identification, the problem of natural image authentication can also be categorized into: *learning-based* methods, based on supervised classification, and *statistical model-based* methods.

To the best of our knowledge, the classifiers based on the supervised-learning scheme, such as SVM, are by far the most popular from forensic investigators for source image identification. Based on the features used for natural image authentication [28, 29], supervised learning-based approaches can be classified into two categories: statistical features and physical features. Let us first review some statistical feature-based methodologies. Based on the first order and high order wavelet statistics, the authors of [30] proposed to extract 216 dimensional features and designed a SVM classifier for natural image authentication. The geometry-based algorithm enabled the authors of [31] to reveal different features from PIM and CG images. Based on 144-dimensional feature vectors extracted from characteristic functions of wavelet histograms, the authors of [32] designed a SVM classifier with lower dimension. It was proposed to extract statistical features from HSV (Hue, Saturation,

Value) color space in [33], in which 234 dimensional features are obtained and used with a SVM classifier. This approach has been improved by using a four-scale contourlet transform in HSV color space [34]. Based on the statistical difference of uniform gray-scale invariant local binary patterns, the authors of [35] proposed to extract four groups of 59 dimensional features from each image, and exploit a SVM classifier for natural image authentication. By considering the visual features, together with image features, it was proposed to design a SVM classifier to distinguish between PIM and CG images in [36, 37]. Physical features is referred to as the features which are extracted from the process of image acquisition. Thus we will mainly discuss the related methods in the following paragraphs (*Sensor Imperfections* and *CFA Pattern and Demosaicing*).

Let us discuss the statistical model-based scheme, exploiting hypothesis testing, for solving the problem of identifying a natural image. Given an image $\mathbf{Z}$ whose source has to be identified, forensic analysts decide between two following hypothesis:

$$\begin{cases} \mathcal{H}_0 : \mathbf{Z} & \text{is acquired by a real digital camera device} \\ \mathcal{H}_1 : \mathbf{Z} & \text{is generated by a computer using a rendering software.} \end{cases}$$

To distinguish between PIM and CG images using hypothesis testing theory one requires to explicitly associate each hypothesis with a statistical distribution (of pixels' value for instance). This is not straightforward, however, this framework allows us to establish analytically the statistical performance of the optimal detectors and to warrant a prescribed false alarm probability. In Chapter 3, we will specifically discuss the problem of natural image authentication using hypothesis testing.

It is very important to keep in mind that, though very different, the two approaches based on supervised-learning or on a statistical model exploit different in-camera fingerprints left behind in every stage of imaging acquisition. In general, based on the sequential order of digital image processing pipeline (see Figure 2.1), it is proposed to classify state-of-the-art algorithms into four categories: *Lens Aberration*, *Sensor Imperfections*, *CFA pattern and demosaicing*, and *Image Compression*.

Lens Aberration Due to the design and manufacturing procedure, lens aberrations (e.g. spherical aberration, chromatic aberration, or radial distortion.) impact an output digital image. Among existing types of lens aberrations, lens radial distortion might have the most severe impact on the output images. When the transverse magnification of an image, referred to as the ratio of the image distance to the object distance, does not remain constant but instead becomes a function of the off-axis image distance, the radial distortion causes straight lines in the object space rendered as curved lines on the image sensor. The two different types of lens radial distortion are illustrated in Figure 2.9. Because the lens radial distortion depends on the optical system it can serves as an intrinsic feature for source camera identification [16]. As describes in [16], the lens radial distortion can be modelled as follows:

$$r = r' + k_1(r')^3 + k_2(r')^5 \tag{2.16}$$

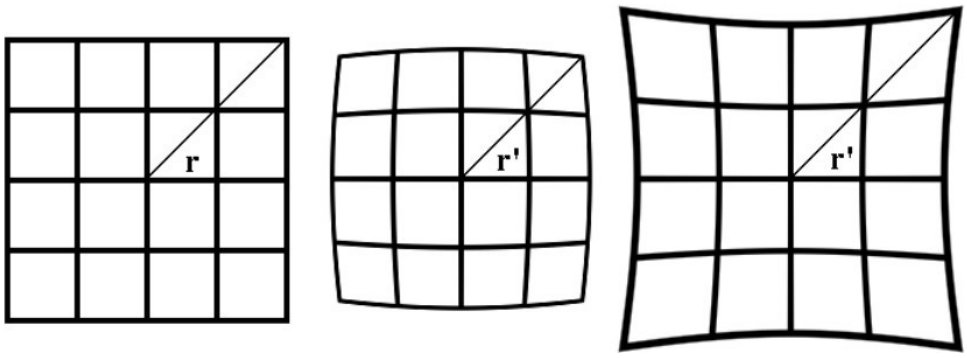


Figure 2.9: Illustration of lens radial distortion. Left: undistorted grid; Middle: grid with barrel distortion; Right: grid with pincushion distortion.

where r and r' are the undistorted radius and distorted radius respectively, and the distortion parameters (k_1, k_2) can be estimated using the straight line method [38, 39]. Based on the intrinsic fingerprints (k_1, k_2) together with other features proposed in [15], the authors of [16] designed a SVM classifier and obtained a relevant detection power. The classifier, however, only investigates the problem of source camera brand identification, and does not give the results of identifying source camera model or individual device on the large scale image database. Besides, the lens aberration-based detector has to face the challenge from interchangeable-lens cameras, that is camera for which the lens can be changed. In this scenario, the proposed detector of [15] is probably unable to identify the source camera.

Sensor Imperfections Due to imperfections during sensor manufacturing process and non-uniformity of photo-electronic conversion caused by inhomogeneity of silicon wafer [40, 41], Sensor Pattern Noise (SPN) extracted from a given image has been one of the first intrinsic fingerprint used for source camera identification [42]. It should be noted that the SPN includes two main components: Fixed Pattern Noise (FPN) and Photo-Response Non-Uniformity noise (PRNU). The FPN represented by dark current, has been used for camera identification in [43]. However, because the FPN can be suppressed by subtracting a dark frame from the output image it cannot be considered as a reliable fingerprint and, hence, has not been used in later works anymore. Compared with the FPN, the PRNU is much more robust. Therefore, SPN is usually referred to as PRNU.

Because the PRNU results from different sensitivity of pixels' silicon wafer, it is a multiplicative noise. However, dealing with multiplicative noise is not straightforward; thus, it has been first proposed to model the PRNU as an Additive White Gaussian Noise (AWGN) formulated as [42]:

$$\mathbf{I}' = \mathbf{I} - F(\mathbf{I}) \tag{2.17}$$

where $\mathbf{I}'$ denotes the unique PRNU, $\mathbf{I}$ is referred to as the original noisy image, and $F(\cdot)$ a function of wavelet denoising filter [44]. To extract the intrinsic PRNU

fingerprints, it is proposed to average K candidate PRNU from the same camera model. Then, the camera reference pattern is formulated as:

$$\mathbf{P} = \frac{\sum_{k=1}^{K} \mathbf{I}'_k}{K} \tag{2.18}$$

where the intrinsic fingerprint $\mathbf{P}$ denotes the camera reference pattern, the index $k \in \{1, ..., K\}$ represents the image number and $K \geq 50$ the total number of images. Moreover, let us define $\mathbf{I}'_{ins}$ as the PRNU of the inspected image $\mathbf{I}_{ins}$ and $\mathbf{P}_0$ as the camera reference pattern of the known camera model $\mathcal{S}_0$. It was proposed in [42] to use correlation c between $\mathbf{I}'_{ins}$ and $\mathbf{P}_0$ as the criteria of identifying source camera model, which can be formulated as:

$$c = \frac{(\mathbf{I}'_{ins} - \overline{\mathbf{I}'}_{ins}) \cdot (\mathbf{P}_0 - \overline{\mathbf{P}}_0)}{\|\mathbf{I}'_{ins} - \overline{\mathbf{I}'}_{ins}\| \|\mathbf{P}_0 - \overline{\mathbf{P}}_0\|} \tag{2.19}$$

where $\overline{\mathbf{I}'}_{ins}$ and $\overline{\mathbf{P}}_0$ respectively denote the mean value of all the coefficients in the matrix $\mathbf{I}'_{ins}$ and $\mathbf{P}_0$, and $\|\cdot\|$ is the symbol of norm. The higher the correlation c is, the most likely the inspected image $\mathbf{I}_{ins}$ comes from $\mathcal{S}_0$. This simple classifier designed in [42] can efficiently identify whether an inspected image $\mathbf{I}_{ins}$ is acquired by the camera model $\mathcal{S}_0$.

This first work is very interesting and efficient but, however, it suffers from two drawbacks. First, because the filter $F(\cdot)$, used to estimate pixels expectation, is not perfect, the extracted PRNU is not independent of the image content, which can be severely contaminated by the details from scenes. The author of [45] gave the assumption that the stronger a signal component is in PRNU, the less trustworthy the component should be. Therefore, based on the intensity of the signal component in PRNU, it is proposed to use 6 models to refine the PRNU. Based on the brightness and smoothness of the image, it is proposed to use the PRNU blocks with weighted factors to generate the new composite senor pattern noise as intrinsic camera fingerprints in [46]. Second, the PRNU acts as a multiplicative noise for which the addition approximation (2.17) might not be always accurate. Inspired by the pioneer work [42], the authors of [47, 48] proposed to estimate the PRNU multiplicative factor instead of estimating the PRNU directly. First, it is proposed to model more accurately a natural image $\mathbf{I}$ captured by a camera as:

$$\mathbf{I} = \mu_{\mathbf{I}} + \mu_{\mathbf{I}}\mathbf{K} + \mathbf{\Xi} \tag{2.20}$$

where $\mu_{\mathbf{I}}$ denotes a noisy-free version of $\mathbf{I}$, $\mathbf{K}$ the multiplicative PRNU factor, and $\mathbf{\Xi}$ represents other types of additive noises which are of no interest for the problem of source camera identification using the PRNU. To estimate the PRNU factor $\mathbf{K}_0$ as the intrinsic fingerprint of source camera $\mathcal{S}_0$, the authors of [47, 48] proposed to utilize the Maximum Likelihood Estimation (MLE) algorithm as follows:

$$\mathbf{K}_0 = \frac{\sum_{k=1}^{K} \mathbf{I}'_k \mathbf{I}_k}{\sum_{k=1}^{K} (\mathbf{I}_k)^2} \tag{2.21}$$

where noise residual $\mathbf{I}'_k$ of the image $\mathbf{I}_k$ can be obtained by (2.17) with $k \in \{1, ..., K\}$. It should be noted that the extracted PRNU may be contaminated by CFA interpolation and JPEG compression. It has thus been proposed in [47, 48], to replace the simple wavelet filter (2.17) by a Wiener filter applied in the wavelet domain. Finally, the work [47, 48] also improves the detection method by using the Peak to Correlation Energy ratio (PCE) as the criteria instead of the cross correlation proposed in [42] for identifying source camera model/device. Indeed, the PCE ia a more stable statistic and independent from the image size. Motivated by the work [48], based on the enhanced PRNU, the authors of [49] proposed to use Circular Correlation Norm (CCN) as the test statistic for identifying source camera model.

It should be noted that in [47, 48] it is also proposed to use the PRNU to detect forgeries. The approach is very similar to the one applied for source camera identification, but instead detects if some small blocks of a given image have the PRNU that does not match the reference PRNU $\mathbf{K}_0$. Those blocks with a different PRNU are indeed likely to reveal traces of forgeries.

Similarly, it has also been proposed to use the correlation of the PRNU for distinguishing between PIM and CG images [50]. The authors of [28] extracted 8 dimensions of multi-fractal spectrum features of PRNU and designed a SVM classifier to solve the problem of natural image identification. Based on the traces of CFA interpolation on the local correlation of PRNU, a SVM identifier was proposed to discriminate between PIM and CG images in [29].

CFA Pattern and Demosaicing The Color Filter Array (CFA) is placed next to the pixels' silicon wafer, that records the light intensity, and is used to filter the incident light spectrum from the optical system. In fact, it is technically possible to produce a full-color sensor, that is for which each pixel directly records all three primary color components: red, green, and blue. This, however, has a prohibitive cost and is thus only used in very high-end cameras. CFA filtering together with demosaicing (or CFA interpolation) technique, hence, constitutes a good tradeoff between image quality and product cost. Assuming that different camera brands/models employ different CFA patterns and demosaicing algorithms, a pioneer work on source camera identification, using CFA design and demosaicing algorithm, has been done in [14, 15].

Inspired by that research, the authors of [51] proposed a novel methodology to estimate the CFA pattern and demosaicing algorithm, which are employed as the features to design an efficient classifier for camera model/brand identification. Let us briefly describe the algorithm proposed in [51]. Based on the observation that a RGB (Red, Green and Blue) type of CFA with a fixed periodicity of 2×2 is employed in most of digital still cameras, the authors first construct a search space containing 36 possible CFA patterns. It should be noted that among different zones of an image, the camera manufacturer generally uses different demosaicing algorithms. Thus, based on the gradient feature in the local neighborhood pixels, three types of regions are classified as follows: the regions with a significant hori-

zontal gradient; the regions with a significant vertical gradient; the regions with the remaining smooth part of the image. Then, singular value decomposition (SVD) is proposed to deal with the contamination from non-linear transformation (e.g. gamma correction) during the image acquisition. Finally, the demosaiscing parameters are estimated and used to match with 36 candidate CFA patterns corresponding to different cameras. Afterwards, the authors proposed to use a SVM classifier based on the estimated parameters of demosaicing algorithm from different camera models to achieve the goal of source camera model/brand identification.

By considering the intra-channel and the cross-channel correlation of demosaicing algorithms, based on the partial second-order derivative correlation model, the authors of [52] proposed a novel classifier to identify source camera models. Besides, it can also detect the acquired TIFF format image edited by using what photo-editing software.

Apart from identifying source camera, the physical features extracted from CFA pattern can be employed to distinguish PIM and CG images. To detect the presence of CFA interpolation coupled with chromatic aberration, it was proposed to discriminate two types of images in [53]. The authors of [54] proposed to improve this approach by using the periodic pattern, in Fourier domain, due to the demosaicing algorithm. However this method is not robust to post-operation processes that largely reduce the periodic pattern due to CFA demosaicing and it is easy to attack the detector [54] with post-camera operations as shows in [55].

Image Compression In the image acquisition pipeline, image compression is the significant final step. Due to its overall good tradeoff between file size and image quality, JPEG compression is by far the most popular option for image comrpession[5]. JPEG compression has two requirements [56]:

- The compressed file size is not allowed to exceed a fixed size.
- The visual quality of JPEG image has to be maximized for a given compression factor.

Since different camera manufacturers or models probably use different parameters in the JPEG compression standard, especially different quantization tables, it was first proposed to extract forensic features from JPEG compression in [57]. The authors of [57] proposed two types of features: bit per pixel and the percentage of non-zero integers in each DCT subband. Bit per pixel is referred to as the number of bit for storing one pixel of an image. It is very easy to obtain this first forensic feature. Let us also briefly describe the second type of features. In the operation of the quantization, the DCT coefficients in the high frequency domain are usually quantized with a larger quantization step size, compared with their counterparts in the low or middle frequency domain. Therefore, in the highest frequencies a more important part of zero-valued DCT coefficients are generated. It is proposed

[5]JPEG popularity may also be explained by the fact that JPEG compression standard is an open, patent-free standard with, thus, no royalty associated with its use.

to calculate the percentage of non-zero integers in each DCT subband, which is formulated as:

$$\begin{aligned} \mathcal{P}_i &= \frac{n_i}{N} \\ \text{with} \quad N &= n_1 + n_2 +, \ldots, n_i, \ldots, +n_{64} \end{aligned} \tag{2.22}$$

where $i \in \{1, 2, \ldots, 64\}$ represents the DCT subband index , and n_i denotes the total number of non-zero integers in i-th subband. Based on the 65 physical features exacted from a JPEG image, a SVM classifier is established to identify source camera brand/model.

Other methods for camera identification based on JPEG compression also have been proposed. For instance, it has been proposed in [58] to model the statistical difference of absolute values of quantized DCT coefficients using a Markov chain model based on the transition probability matrix extracted from four different directional (horizontal, vertical, main diagonal, and minor diagonal). Based on the probability matrices from the Y and Cb components from JPEG images, it is proposed to design a SVM classifier to solve the problem of identifying source camera brand/model.

2.4.2.2 Image Content Integrity

The forensic research of image content integrity fundamentally addresses two problems: image forgery authentication and steganographic image detection. Image forgery authentication aims to detect any manipulation on the image content. In fact, the main goal of image tempering detection can be formulated as follows: "how to extract the post-camera fingerprints or in-camera fingerprints". Post-camera fingerprints are extracted from the image processing after obtaining an image, which is the extrinsic features of an image. While in-camera fingerprints are extracted from the operation of image acquisition, which is the intrinsic features of an image. Generally, any manipulation unavoidably leave traces, which more or less reveal the inconsistency of the image, compared with its un-tempered original version. Therefore, by tracking the inconsistency of post-camera or in-camera fingerprints from the inspected image, forensic researchers have generally proposed the following six types of tempering detection methodologies: *statistical features-based*, *camera fingerprints-based*, *linear manipulation-based*, *non-linear manipulation-based*, *compression history-based*, and *physical inconsistency-based*. The problem of steganographic image detection usually is discussed in the field of steganalysis, which will be mainly analyzed in Section 2.5.

Statistical Features-based Apart from distinguishing between PIM and CG images, the statistical features extracted from the image under investigation can also be used for image tempering detection. This type of detectors mainly authenticates the image with tempering techniques containing copy-move (see Figure 2.4), splicing (see Figure 2.5), resampling (see Figure 2.6), and brightness modification (see Figure 2.7). Several approaches have been proposed to extract statistical-based features

set, for instance using the higher-order wavelet statistics in [59] and using binary similarity measures in [60]. Those features are, very usually, used with a supervised learning algorithm to design a detector. By reducing the dimensionality of the statistical features, the technique of copy-move tempering can be tackled with the algorithms proposed in [61, 62]. Besides, some statistical feature-based algorithms can be extended to detect the image with hidden information [63, 64].

Camera Fingerprints-based This type of forensic detectors extracts in-camera fingerprints which are related, and sensitive, to different operations from image acquisition. Such fingerprints can mainly be used for detecting the splicing technique. When the portion of the innocent image is replaced by another part from another image, acquired with a different camera and/or different acquisition settings, the inconsistency of the intrinsic fingerprints from those different regions reveals the tampering (see Figure 2.5). Several approaches have been proposed based on different elements from the image acquisition pipeline such as lens aberration [65, 66], sensor pattern noise [47], CFA pattern and demosaicing [67–69], and gamma correction [70, 71].

Linear Manipulation-based The linear manipulation-based detectors aim at detecting, in a given image, artifacts from resampling (see Figure 2.6) which are mainly related to the operations of resizing, rotation, and other geometric manipulations. The prior works for exploiting the artifacts of resampled images can generally be classified into two following fundamental categories:

1. Methods in the first category focus on identifying the linear correlation of resampled images in the spatial domain. By using Expectation Maximization (EM) algorithm [72], it first proposed a linear predictor to expose the correlation existing among neighboring pixels of the resampled images in [68]. Although EM algorithm was capable of estimating the linear interpolated parameters, it has a high computational complexity. Driven by this pioneer work, a slightly different method has been proposed in [73], using the second-derivatives and a parametric model of resampling correlation instead of the EM algorithm, with a much lower computation cost. Instead of estimating the interpolation parameters, the fixed parametric predictor indeed improves the efficiency of detection, yet the detection accuracy can not be guaranteed. The problem of how to find a good tradeoff between the computation efficiency and estimation accuracy remains open. Instead of using a single predictor, the authors of [74] proposed to use a global predictor to detect a resized image. It is however limited to the detection of resizing, excluding other geometric manipulation such as rotation or shearing. The methods in this category can efficiently detect linear manipulation but, however, cannot estimate the interpolation period.

2. Methods in the second category investigate the second or high-order derivative of resampled images in the frequency domain. By exploiting the second-derivative algorithm, the periodicity of an interpolated image has been exposed in [75]. The method proposed in [75] can estimate the interpolation period but, however, is limited to the detection of resizing manipulation. By analyzing the variance of the n-th order derivative, the authors of [76] improved this approach to expose the trace of image resampling. However, the linear transformation matrix can not be approximated. Inspired by the method [75], it has been proposed to analyze the relationship between the rotation angle and the frequency in the spectrum of the image's edge map [77]. However, this method is also limited to the detection of rotation which explains it cannot be widely used. By extracting relevant features set and using supervised learning method such as SVM, it has been proposed to detect whether the inspected image is resampled (see [78, 79]). Although those methods could achieve high detection accuracy, supervised statistical learning is rather time-consuming. In addition, several problems such as the robustness to training and testing set mismatch remain open. In order to approximate the linear transformation of an interpolated images, authors of [80] proposed to analyze the second-derivative of images in both the horizontal and vertical directions. Although these methods in the second category could estimate the linear transformation matrix and interpolation period, their statistical performance for the tampering detection problem has not been deeply studied.

Non-linear Manipulation-based A rather classical operation in image processing is the median filtering; it is indeed computationally very simple and provides an overall good tradeoff between denoising and edge preserving. However, medial filtering is typically a non-linear modification which, hence, cannot be detected with linear manipulation-based detectors. Non-linear manipulation-based detectors actually almost only study whether the image under investigation is tempered by a median filter (see subsection 2.3.5). Median filtering tends to create streaks of adjacent pixels with the same value. Thus by considering the difference between adjacent pixels' intensity, authors of [81–83] proposed efficient forensic detectors for the median filtering.

Another type of non-linear manipulation-based detectors investigates the brightness modification (see Figure 2.7). This is a tempering technique used for modifying the illumination conditions of an image. When falsifying a given image, typically using splicing or copy-move technique, one may apply brightness modification to alter illumination condition from the tempered portion such that it matches the background of the image [5]. After dealing with brightness modification, impulsive peaks and gaps clearly appear in the pixel value histogram of an image. By analyzing such artifacts, the authors of [84, 85] proposed a detection technique for images subject to brightness modification.

Compression History-based JPEG standard is, by far, the most popular image lossy compression algorithm due to its overall good tradeoff between file size and image quality. When forging an image by splicing technique, it is thus very likely that the image, or the spliced portion, will be stored twice in JPEG format with different compression parameters. Thus, the general idea of the compression history-based detectors is to reveal the traces of manipulation using JPEG compression artifacts. Those methods are usually based on one of the two obvious JPEG compression artifacts: comb-like distribution of DCT coefficients and blocking artifact. Due to the quantization of the DCT coefficients in JPEG compression, some integers have a more higher probability to appear and even more when the image is compressed twice with different quantization steps. Due to the block-by-block based DCT transformation in JPEG compression process, the block boundaries in the spatial domain are visible.

For clarity, let us consider two tempering scenarios. In the first scenario, a forger maliciously falsifies an innocent JPEG image and re-saves it with an uncompressed image format such as TIFF. By analyzing the JPEG compression artifacts, the detectors proposed by [86–88] is efficient to detect that such an image, though stored with as an uncompressed file, has previously been compressed using the JPEG format. In the second scenario, a forger again maliciously falsifies an image but instead re-saves the tempered image using once again JPEG format. By investigating the inconsistency between the quantization tables and the quantization artifacts from the inspected image, it has been proposed to design an efficient detector to detect such double compression of JPEG images in [89–91]. In addition, in the case that the quantization table used in the single and double JPEG compressed image remains same, by using a random perturbation strategy proposed in [92], the JPEG double-compression history with the same quality factor is revealed. Besides, using the assumption that a portion of a given image (typically the tempered area) has a different compression history than the remaining part of the image, the compression history-based detectors can also authenticate whether the JPEG image is tempered by splicing technique [93–96]. Moreover, the methodology of detecting JPEG compression history can be extended to the research of steganalysis [90, 91].

Physical Inconsistency-based Different from the detectors based on the intrinsic or extrinsic fingerprints, the physical inconsistency-based detectors do not require to extract fingerprints, but make use of properties of lighting conditions in an image under investigation. It is assumed that, since the direction of the light source from different objects in an image can be estimated, the inconsistency in the lighting direction is used to identify the tempered region (e.g. splicing) of an image [97]. Therefore, based on two-dimensional or three-dimensional surface normals and illumination under a single or multiple lighting source, the authors of [98–100] proposed a series of detectors for image tempering authentication.

2.4.3 Counter-forensics

Counter-forensics is referred to as post-camera processing techniques which are designed as countermeasures, for forgers, against digital image forensics [101]. The intention of counter-forensics is to reveal the weakness of some state-of-the-art forensic detectors. In the forensic community, forensic researchers usually assume the very ideal scenario where a forger does not take into account the image forensic techniques. Therefore, the counter-forensics is proposed to challenge that assumption. It is proposed to classify the counter-forensic techniques into two categories: *hiding artifacts* and *forging fingerprints* which are both used for fooling image forensic detectors.

2.4.3.1 Hiding Artifacts

The principle of counter-forensics by covering artifacts is to prevent forgery detection by adding another processing step during the process of image manipulation, to compensate the main artifacts of the forged image. Of course, because each forgery introduces specific artifacts, there is not a general methodology and each forgery has a specific technique to hide specific artifacts. To give a few examples, the resampling artifacts [68] can be concealed using a non-linear filter (e.g. median filter) and adding random white noise [102, 103]. By estimating the specific position of gaps and peaks [84, 85] in the histograms, the forger can easily hide the artifact of the brightness modification by adding random noise [104]. JPEG compression has two artifacts: comb-like histogram in DCT domain and blocking traces [86–88]. Similar to the technique for hiding brightness modification, the the gaps of DCT-values histogram can be compensated using noise dithering technique based on the estimated distribution parameters [105]. The blockiness artifacts can also be compensated [105] by smoothing blocks' boundaries with a median filter.

2.4.3.2 Forging Fingerprints

A more advanced counter-forensic technique not only conceals the tampering artifacts, but also add the new faked fingerprints which can fool the forensic investigators. By synthetically creating the CFA pattern in a digital image, the counter-forensic algorithm [106] effectively forges the fingerprints which are used for analyzing the CFA structure of an image under investigation. When a forger conceals the original sensor pattern noise (e.g. PRNU) and superimposed a different fake pattern noise, the PRNU-based detectors such as [42, 47, 48] are probably not able anymore to identify the source camera model [103].

2.5 Overview on Steganography and Steganalysis

In order to make this thesis rather comprehensive, let us know briefly introduce the last type of very different forgery that has been studied in this work: information

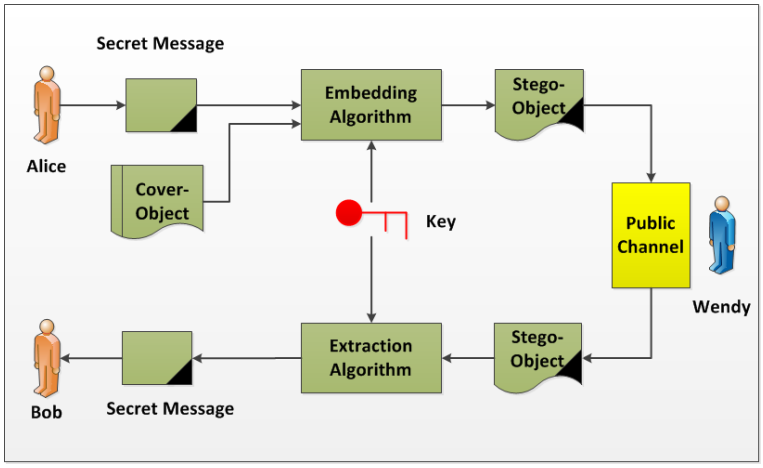


Figure 2.10: Typical steganographic mechanism.

hiding. Steganography is the art and science of hiding secret messages in the *cover-object*, which is used for embedding secret information. The cover-object within which secret information has been hidden is referred to as the *stego-object*. On the opposite, steganalysis aims at detecting the hidden secret information embedded in cover-objects, that is identifying cover-objects from stego-objects. In the digital forensic community, steganography in fact is a type of tempering technique, in which the image content integrity is damaged. In general, steganographic techniques, however, can not be arbitrarily defined as maliciously image forgery. Because the aim of steganography is to create a somewhat hidden communication channel between a sender and a recipient while the secret messages embedded in the cover-object can not be detected by a third party. To counter against steganographic techniques, a steganalyzer as a third party investigates if the object (e.g. image) under investigation is embedded within secret information, which is referred to as a digital forensic tool used for investigating the image content integrity.

The typical steganographic mechanism modelled as the prisoner problem [107] is illustrated in Figure 2.10. Alice and Bob, two prisoners, intend to conspire together a plan for escaping jail. The only communication channel that they can use, however, is unfortunately under surveillance by Wendy who is a warden of the prison. The fundamental problem of steganographic mechanism is that the stego-object, in which the secret message (e.g. the plan for escaping jail) is hidden, should be exchanged freely between Alice and Bob while Wendy can not observe any abnormal object in the public channel. Another key step is how to find a good cover-object as a carrier. In practice, digital media (e.g. image, audio, or video) are very often used as carriers for transmitting secret information. For a practical context, we focus on the digital image steganography and steganalysis, that uses digital images as cover-objects. After obtaining the cover-object, Alice needs to design a secure steganographic methodology involving embedding algorithm and extraction algorithm with a unique key, shared only with Bob and, of course, not known by

Wendy.

In the steganographic mechanism (see Figure 2.10), it is assumed that the public channel, the cover-object, and the secret message are already known, or selected, hence Alice needs to devote all her effort on the design of an optimal steganographic embedding method under the following three requirements:

- Imperceptibility: the stego-object is perceptually indistinguishable from the cover-object.
- Capacity: the amount of hidden secret messages must be "sufficiently" large.
- Robustness: the capacity to preserve the hidden data in case of being removed from the stego-object and counter against some steganographic attacks.

These three requirements are very similar to the criteria of the watermarking schemes (see subsection 2.4.1). Imperceptibility can guarantee that the embedding mechanism can not generate a distortion that is sufficiently large to create visual distortion of the image quality; imperceptibility thus largely decreases the risk of the suspicion from the steganalyst, Wendy. Capacity provides to the steganographer, Alice, enough space for embedding secret messages, for instance the plan of prison break. Robustness is referred to as the ability of extracting the secret message when the stego-object is attacked by a steganalyst, Wendy, or when the public channel is under distortion.

A typical steganographic mechanism mainly involves two stages: embedding stage and extraction stage. When Alice intends to transmit a secret message $\mathbf{I}$, she embeds it into a cover image $\mathbf{C}$ via an embedding algorithm which uses a secure key. The secret message $\mathbf{I}$ is denoted as a binary sequence of L bits, where $\mathbf{I} = (i_1, i_2, \ldots, i_L)^T$ with $i_l \in \{0, 1\}, 1 \leq l \leq L$. The stego image $\mathbf{S}$ is then sent to Bob through an insecure public channel. At the receiver's side, Bob can retrieve the secret message $\mathbf{I}$ by using an extraction algorithm with the secure key which is known in advance. It should be noted that when extracting the secret message $\mathbf{I}$, Bob does not require the original cover image $\mathbf{C}$. During the transmission of the secret message $\mathbf{I}$, if Wendy successfully detects that a stego image $\mathbf{S}$ is exchanged between Alice and Bob, in the public channel, steganography fails. In this scenario, different from cryptanalysis, steganalysis does not care about whether Wendy can retrieve the specific secret message.

In a digital world, there exists many steganographic tools available on the Internet. Because such tools are readily available and easily usable, more and more secret messages embedded in cover media such as images, videos and audio are transmitted in our daily life. Supposed that the steganographic technique is utilized by the criminal with the aim to take part in malicious activities, or by the traitor intending to threaten the homeland security, the consequence becomes very horrible. Thus in this operational context, it is necessary to design the optimal and the most reliable steganalysis methodology to counter against steganography. In general, due to its simplicity most of steganographic schemes insert the secret message into the

Least Significant Bit (LSB) plane of cover media, involving two types of steganography: LSB replacement and LSB matching. The former algorithm aims at replacing LSB plane in the spatial domain or frequency domain of the cover media by 0 or 1 from the secret message $\mathbf{I}$. The latter algorithm, also known as ± 1 embedding (see [6,108,109]), randomly increases or decreases pixel or DCT coefficient value, only when necessary, such that after modification, its value matches the secret message bit i_l it is aimed at embedding.

Since LSB replacement is easier to implement it remains more popular and, hence, as of December 2011, *WetStone* declared that about 70 percent of the available steganographic softwares are based on the LSB replacement algorithm [110,111]. Therefore the research on LSB replacement steganalysis remains an active topic.

As a steganalyst, Wendy only intends to detect the presence of the secret message hidden in the stego image. Moreover, she can also estimate the message length or brute-force the secure key and even retrieve the content of the secret message. The estimation of the message length is defined as *quantitative steganalysis* [112–116]. Brute-forcing the secure key and retrieving the message content are defined as *forensic steganalysis* [117, 118]. In this thesis, we mainly discuss LSB replacement steganalysis which can be implemented in the spatial domain or frequency domain. In general, the current methods for LSB steganalysis can be classified into three categories: *weighted-stego detectors*, *statistical detectors*, and *universal detectors*.

2.5.1 Steganography of LSB Replacement

2.5.1.1 LSB Replacement Description

Let us first define a cover image as $\mathbf{C} = \{c_n\}, n \in \{1, \ldots, N\}$, in which N denotes the total number of pixels. The LSB replacement steganography in the spatial domain rely on the general concept of hiding the secret (binary) message $\mathbf{I}$ of L bits, where $\mathbf{I} = (i_1, i_2, \ldots, i_L)^T$ with $i_l \in \{0, 1\}, 1 \leq l \leq L$, by replacing the original LSB of the cover image. Let us formally define the LSB pixel value of c_n as:

$$\mathrm{LSB}(c_n) = \mod(c_n, 2) \tag{2.23}$$

where $\mathrm{mod}(a, 2)$ denotes the remainder of the Euclidean division of a by n. Thus the value of $\mathrm{LSB}(c_n)$ is thus into the set $\{0, 1\}$. After replacing the LSB of the cover image pixel, $\mathrm{LSB}(c_n)$, by the bit i_l from the secret message, the value of the stego-image pixel s_n is given by:

$$s_n = c_n - \mod(c_n, 2) + i_l. \tag{2.24}$$

From Equation (2.24), one can immediately establish the impact of LSB embedding on pixels' value in the following two cases:

$$s_n = \begin{cases} c_n & \mathrm{mod}(c_n, 2) = i_l \\ \overline{c_n} & \mathrm{mod}(c_n, 2) \neq i_l \end{cases} \tag{2.25}$$

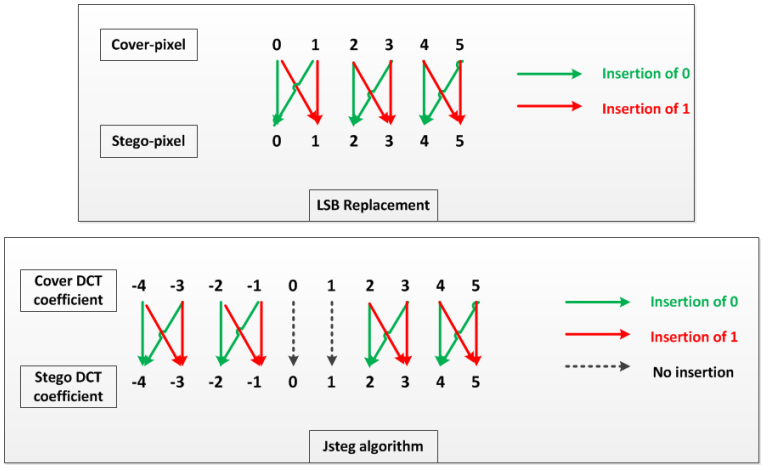


Figure 2.11: Illustration of LSB replacement (top) and JSteg (bottom).

where $\overline{c_n} = c_n + (-1)^{c_n}$ represents the value of n-th the pixel with flipped LSB. Therefore, one can immediately find the fundamental principal of LSB replacement steganography: if the LSB of pixel n equals 0, that is $\text{LSB}(c_n) = 0$, it can either remain the same or be increased to 1; on the contrary, if the LSB of pixel n equals 1, that is $\text{LSB}(c_n) = 1$, it can either remain the same or be decreased to 0. In other words, odd values are never increased while even values are never decreased. Besides, the absolute value of the difference between a stego pixel s_n and a cover pixel c_n cannot be larger than 1, $|s_n - c_n| \leq 1$. It is proposed to illustrate the implementation of LSB replacement technique in Figure 2.11. Steganography of LSB replacement is only operated in the LSB of a cover image, hence the visual distortion of a stego image is largely tolerable, which satisfies the proposed criteria: imperceptibility. For considering the security of steganographic algorithms, it is reasonable that before embedding the secret message $\mathbf{I}$, it is proposed to generate a pseudo-random sequence from a secure key for selecting the candidate cover pixel c_n used for embedding. It should be noted that the secure key is only known, or shared, by the sender and the receiver. Finally, we note that a steganographer can potentially embed a maximal number of bits that corresponds to the number of pixels N of the stego image. In practice, a steganographer very often embeds a much smaller amount of secret message's bits (steganography criteria: capacity) which defines the embedding payload, or embedding rate R, defined in bits per pixel (bpp) as:

$$R = \frac{L}{N} \tag{2.26}$$

where, again, L is the number of bits from the secret message and N denotes the number of pixels from the cover image.

2.5.1.2 JSteg Steganography Description

JSteg steganography is an embedding algorithm which also relies on the LSB replacement mechanism but operates directly on the DCT coefficients from JPEG images, instead of pixels' value from the spatial domain. JSteg steganographer embeds the secret message into the LSB of quantized DCT coefficients subband by subband. The coefficients from the first subband, often referred to as DC coefficients, represent the mean of coefficients over the block of 8×8 pixels. The modification of those coefficients may be obvious and creates artifacts that can be detected easily, hence, they are usually not used for data hiding. Similarly, the JSteg algorithm does not use the coefficients from the other subbands, referred to as AC coefficients, if they equal 0 or 1. In fact, it is known that using the coefficients equal to 0 or 1 modifies significantly statistical properties of AC coefficients; this creates a flaw that can be detected. For JPEG images, the definition of the embedding payload, or embedding rate R, is slightly modified to take into account that all the DCT coefficients are not usable. In such as case, the usual definition of the embedding payload is formulated, in bits per non-zero AC coefficients as:

$$R = \frac{L}{\sum_{k=2}^{64} n_k} \tag{2.27}$$

where n_k denotes the number of usable coefficients in the k-th DCT subband, $2 \leq k \leq 64$.

2.5.2 Steganalysis of LSB Replacement

Let us denote $\mathbf{Z}$ a digital image under investigation. A steganalyst can decide between the following two hypotheses:

$$\begin{cases} \mathcal{H}_0 : \mathbf{Z} = \mathbf{C} & \text{is not embedded in LSB by a secret message} \\ \mathcal{H}_1 : \mathbf{Z} = \mathbf{S} & \text{is embedded in LSB by a secret message.} \end{cases} \tag{2.28}$$

Many steganalysis methodologies have been proposed to solve the problem (2.28). As already states before, the current art for LSB replacement steganalysis methods can generally be classified into three categories[6]: weighted stego-image detectors, statistical detectors, and universal detectors. In the following paragraphs, those three types of detectors are presented in more detail.

2.5.2.1 Weighted Stego-image Detectors

In 2004, the weighted stego-image (WS) method [120] and the test proposed in [121] for LSB replacement steganalysis open the way to optimal detectors. By minimizing

[6]Note that a forth category of steganalysis method, usually referred to as the "Structural detectors", has been proposed but it not studied any longer since almost a decade now [119] and, hence, is not presented in this brief state-of-the-art.

the distance between the weighted stego-image and the cover-image, WS detectors can estimate the length of the secret message. For the clarity of the presentation, let us define a weighted stego-image $\mathbf{Z}$ with scale parameter α as:

$$\forall n \in \mathcal{N}, \quad z_n^{(\alpha)} = (1-\alpha)z_n + \alpha\overline{z_n}, \quad \text{with} \quad \overline{z_n} = z_n + (-1)^{z_n} \tag{2.29}$$

where $z_n^{(\alpha)}$ denotes the weighted stego-image. Roughly speaking, $z_n^{(\alpha)}$ represents a weighted stego-image in the sense that from Equation (2.29) it is a weighted sum of cover pixels' values and stego pixels' values. The idea of the WS algorithm is to estimate the payload using the weighted stego-image (2.29); this is carried out by a Least Square algorithm (LS) which corresponds to minimizing the Euclidian distance between the $z_n^{(\alpha)}$ and the cover-image $\mathbf{C}$:

$$\widehat{R} = 2\arg\min_{\alpha} \sum_{n=1}^{N} \left(z_n^{(\alpha)} - c_n\right)^2 = 2\sum_{n=1}^{N} \frac{1}{N}(z_n - \overline{z_n})(z_n - c_n), \tag{2.30}$$

where N denotes the number of pixels in the inspected image $\mathbf{Z}$. Since the value of the cover image pixel c_n are unknown, the WS algorithm proposes to use a local linear filter $\mathcal{F}$ to find an approximation, or estimation, of c_n as $\widehat{c}_n = \mathcal{F}(z_n)$. Unfortunately, the local estimator $\mathcal{F}$ is not independent of the content of an image $\mathbf{Z}$ under investigation. In fact, the local estimator $\mathcal{F}$ performs more accurately in flat areas and less accurately in textured areas. Thus, it has reasonably been proposed in the original WS method [120], to use a weighting vector $\mathbf{w} = (w_1, \cdots, w_N)$ with $\sum_{n=1}^{N} w_n = 1$ in the estimator $\widehat{R}$. The goal of this vector $\mathbf{w}$ is to put more weight on pixels in flat areas, the most reliable, and less on pixels in textured areas, for which the cover pixel's value is not accurate. Therefore, the modified estimator $\widehat{R}$ can be written:

$$\widehat{R} = 2\sum_{n=1}^{N} w_n(z_n - \overline{z_n})(z_n - \mathcal{F}(z_n)), \tag{2.31}$$

where the authors of [120] proposed to use the a linear filter $\mathcal{F}$ by simply averaging the four-directional neighboring pixels around z_n. More accurate pixel estimators have been proposed in [122]. Apart from the option of local estimator $\mathcal{F}$, a weight vector $\mathbf{w}$ is also significant to improve the performance of WS detectors. the enhanced WS algorithm proposed in [122] improved the detection rate by modifying the pixel predictors, adjusting weighting factors and introducing the idea of bias correction.

It should be noted that the WS detectors can be categorized as a quantitative steganalysis. This detector indeed directly estimates the payload R, while does not directly solve the testing problem (2.28). In the estimation of R, even if the real embedding rate $R = 0$, the estimator $\widehat{R}$ will probably arrive at a positive or negative value. In general, in the field of steganalysis, the performance of an estimator is measured by the Mean Absolute Error (MAE):

$$\frac{1}{K}\sum_{k=1}^{K} |\widehat{R_k} - R| \tag{2.32}$$

where K denotes the number of the inspected images, or by the Median Absolute Error (mAE):

$$\text{med}|\widehat{R_k} - R| \tag{2.33}$$

where med$|\cdot|$ denotes the median value in the absolute vector $|\widehat{R_k} - R|$.

Nevertheless, the drawback of the WS method [120, 122] is that it can only be applied in the spatial domain. Due to the prevalence of images compressed in JPEG format, dealing with this kind of images becomes mandatory. Inspired by the prior studies [120, 122], the WS steganalyzer for JPEG covers was proposed in [123]. In a similar fashion as the WS estimator defined in (2.31), its modified version for the JPEG images is given by:

$$\widehat{R} = \sum_{n=1}^{N} w_n(x_n - \overline{x_n})(x_n - \mathcal{F}_{DCT}(x_n)), \tag{2.34}$$

where x_n denotes the n-th usable DCT coefficient, N the total number of DCT coefficients and the coefficient predictor $\mathcal{F}_{DCT}$ estimates the original DCT coefficient by averaging the four neighboring coefficients, from the same subband, in the four adjacent blocks. It should be noted that the neighboring coefficients come from the same subband excluding the DC coefficients or border coefficient blocks. Two improvements have been proposed based on empirical evidences. First, for the JPEG image WS algorithm, the coefficient predictor $\mathcal{F}_{DCT}$ is ignored for better estimation accuracy. Second, an offset factor $W^\star$ is added in the estimation. The WS payload estimator $\widehat{R}$ proposed for JPEG images is thus given by:

$$\widehat{R} = \sum_{n=1}^{N} w_n(x_n - \overline{x_n})x_n + W^\star. \tag{2.35}$$

The interested reader can refer to [123].

2.5.2.2 Statistical Detectors

When it is assumed that the embedding payload R is known, the approach of statistical detectors is to exploit a distribution model for cover images and address the testing problem (2.28). In this scenario, it is possible to design "optimal detectors" [124], by using hypothesis testing theory. Let us assume that the pixels from the inspected image $\mathbf{Z} = \{z_n\}$, $n \in \{1, \cdots, N\}$ are independent and they all follow the probability distribution, denoted as $\mathcal{P}_{\boldsymbol{\theta}}$, parameterized by the parameter $\boldsymbol{\theta}$. By considering that the pixels are quantized, the distribution $\mathcal{P}_{\boldsymbol{\theta}}$ is represented by its probability mass function (pmf) denoted $P_{\boldsymbol{\theta}} = \{p_{\boldsymbol{\theta}}[z_n]\}$. Supposed that a cover image $\mathbf{C}$ is used as a carrier to hide a secret message with embedding payload R. The pixels from the resulting stego-image, denoted $\mathbf{S}$, follow the statistical distribution $\mathcal{Q}^R_{\boldsymbol{\theta}}$ which is completely characterized by its pmf $Q^R_{\boldsymbol{\theta}} = \{q^R_{\boldsymbol{\theta}}[z_n]\}$ given, from Equations (2.23)-(2.25), by:

$$q^R_{\boldsymbol{\theta}}[z_n] = (1 - \frac{R}{2})p_{\boldsymbol{\theta}}[z_n] + \frac{R}{2}p_{\boldsymbol{\theta}}[\overline{z_n}], \tag{2.36}$$

where, again, $\overline{z_n} = z_n + (-1)^{z_n}$ represents the integer z_n with flipped LSB. Since the sequence of secret message bits, which are independent and identically distributed (i. i. d.), follow a Binomial distribution $\mathcal{B}(1, 1/2)$, the probability that a stego pixel s_n equals the cover c_n with flipped LSB equals $R/2$ while, on the contrary, the probability that a stego pixel s_n equals the cover pixel c_n equals $1 - R/2$.

When investigating an image $\mathbf{Z}$, a statistical steganalyst always makes a decision between the following two hypotheses: $\mathcal{H}_0$: "*the pixels z_n follow the distribution $\mathcal{P}_{\boldsymbol{\theta}}$*" and $\mathcal{H}_1$: "*the pixels z_n follow the distribution $\mathcal{Q}^R_{\boldsymbol{\theta}}$*" which can be written formally as:

$$\begin{cases} \mathcal{H}_0 : \{\mathbf{Z} \sim \mathcal{P}_{\boldsymbol{\theta}}\}, \\ \mathcal{H}_1 : \{\mathbf{Z} \sim \mathcal{Q}^R_{\boldsymbol{\theta}}\}. \end{cases} \tag{2.37}$$

Assuming that all the pixels are i. i. d. , it has been proposed in [125] to design two detectors based on the the probability distribution $\mathcal{P}_{\boldsymbol{\theta}}$. When $\mathcal{P}_{\boldsymbol{\theta}}$ is unknown, based on Hoeffding's test [126], it is proposed to develop an asymptotically optimal detector; When $\mathcal{P}_{\boldsymbol{\theta}}$ is prior-known, based on Neyman-Pearson Lemma [127, theorem 3.2.1] an optimal detector is established. Since the cover model proposed in [125], assuming that all the pixels follow the same distribution, cannot describe accurately the pixels' distribution, the proposed detectors cannot achieve a high detection power. In fact, the problem of finding the most accurate statistical model for cover images remains open [128] and, thus, so is the problem of using such model to design statistical test with the highest detection accuracy.

Relying on a simplistic local polynomial model for pixels' expectation, some statistical tests, such as those proposed in [129–133], achieved a high detection performance compared with WS detectors. Among these proposed steganalysis, based on the hypothesis testing theory, the Likelihood Ratio Test (LRT), which has some optimal properties, is given by:

$$\delta^{\mathrm{lr}}(\mathbf{Z}) = \begin{cases} \mathcal{H}_0 \ \text{ if } \Lambda^{\mathrm{lr}}(\mathbf{Z}) = \dfrac{\mathcal{Q}^R_{\boldsymbol{\theta}}[\mathbf{Z}]}{\mathcal{P}_{\boldsymbol{\theta}}[\mathbf{Z}]} < \tau^{\mathrm{lr}} \\ \mathcal{H}_1 \ \text{ if } \Lambda^{\mathrm{lr}}(\mathbf{Z}) = \dfrac{\mathcal{Q}^R_{\boldsymbol{\theta}}[\mathbf{Z}]}{\mathcal{P}_{\boldsymbol{\theta}}[\mathbf{Z}]} \geq \tau^{\mathrm{lr}}, \end{cases} \tag{2.38}$$

where Λ^{lr} denotes the Likelihood Ratio (LR) and τ^{lr} a decision threshold set to guarantee a prescribed false alarm rate (FAR). By challenging the assumption that all the pixels are i. i. d. , the detectors proposed in [129–133] assume that each pixel follows a Gaussian distribution, $z_n \sim \mathcal{N}(\mu_n, \sigma_n^2)$; using this statistical model to solve the testing problem (2.38), the LR Λ^{lr} can be written as:

$$\Lambda^{\mathrm{lr}}(\mathbf{Z}) \propto \sum_n \frac{1}{\sigma_n^2}(z_n - \overline{z}_n)(z_n - \mu_n), \tag{2.39}$$

where the model parameter $\boldsymbol{\theta} = (\mu_n, \sigma_n^2)$ characterizes each pixel. The $\Lambda^{\mathrm{lr}}(\mathbf{Z})$ is very similar to the WS detector (2.31) which justifies its good detection performance. Similarly, a LSB matching detector has been obtained using the same methodology

in [134]; for the LSB matching embedding mechanism, the LR given by:

$$\Lambda^{\text{lr}}(\mathbf{Z}) \propto \sum_n \frac{1}{\sigma_n^4}((z_n - \mu_n)^2 - \frac{1}{12}). \tag{2.40}$$

Although the most powerful detector can maximize the detection power while guaranteeing a prescribed false alarm probability, it is not a practical steganalyzer since it assumes a perfect knowledge of the payload R and the model parameter $\boldsymbol{\theta}$, while they are both unknown in practice. Therefore, it has also been proposed to design practical detectors. When the embedding rate R is unknown, one can try to design a test which is locally optimal around a given embedding payload R, namely a Locally Asymptotically Uniformly Most Powerful (LAUMP) test, as proposed in [130, 132]. Besides, one can employ the Maximum Likelihood Estimation (MLE) to obtain the estimated model parameter $\widehat{\boldsymbol{\theta}}$, which is then used in the design of a practical Generalized Likelihood Ratio Test (GLRT).

One can also extend the statistical detector in the spatial domain (2.38) to JSteg steganalysis. The detector proposed in [135] is based on the same interesting methodology; however it is based on the assumption that the DCT coefficients are i. i. d. within a subband; that means all the DCT coefficients from a subband follow a Laplacian distribution with a zero expectation and a constant scale parameter. The authors of [23, 136], however, proposed a much more accurate steganalysis of JPEG images by exploiting a more accurate cover model for DCT coefficients.

2.5.2.3 Universal Detectors

Different from the previous two types of detectors, universal steganalyzers are more practical steganalyzers for the two following main reasons. First of all, current art of universal detectors based on high dimensional features and supervised learning methods indeed achieves detection accuracy that is largely better than other types of steganalyzers. Second, those detectors are highly flexible and can potentially be trained to detect a wide range of different steganographic schemes.

In general, universal steganalysis can be classified into two types: supervised and unsupervised. Similar to the passive forensic detectors in the supervised framework (see Section 2.4.2), the labeled images in the training stage are used for training a steganalysis classifier. Authors of [64, 115, 137–143] proposed to investigate a series of supervised universal steganalyzers based on several supervised learning methods, among which the SVM classifier has been the most widely-adopted. If the training data is not perfectly representative of the cover model, the accuracy of steganalysis can not be guaranteed, which may result in the mismatch problem [124]. Unlike supervised universal steganalysis, unsupervised universal detectors do not meet the problem of mismatch since it can not establish a cover model until the classification stage [124]. To the best of our knowledge, no literatures have been proposed to solve the problem of unsupervised universal steganalysis.

The performance of universal steganalysis largely relies on the features extracted from an image under investigation. These features should be sensitive to changes

caused by steganographic algorithms, but robust to some post-camera non-steganographic processing techniques. Therefore, the problem of designing suitable features for universal steganalysis remains open [124].

2.6 Conclusion

This chapter first introduces some popular image forgery techniques. Then it is proposed to mainly address the research of digital image forensics involving two general problems: image origin identification and image content integrity. Forensic investigators employ two schemes, which are active and passive forensic techniques, to solve those two general problems. Since active forensics needs the pre-embedded watermark/signature in practice, we focus our discussion on passive forensics. Based on the pipeline of image acquisition, we discuss some state-of-the-art forensic detectors which extract in-camera/post-camera fingerprints as the evidence for digital image forensics. To detect LSB replacement steganography, we mainly investigate three types of special steganalysts.

To the best of our knowledge, very limited number of forensic detectors investigate digital image forensics based on the framework of hypothesis testing theory. To fill this gap, in this thesis, we will propose to design forensic detectors based on hypothesis testing theory and to move those into the practical operational context. In general, the main contributions of this thesis are classified as follows:

- After CFA filtering, each pixel records an intensity for only one of the three color channels (red or green or blue); demosaicing is the post-process by which it is possible to restore the full-color image. Due to this process, a linear relationship is created among neighboring pixels which can be used as an intrinsic fingerprint of a natural image. It is proposed to exploit such a fingerprint in order to distinguish a photographic image (PIM) from a computer generated image (CG). In Chapter 3, we specifically investigate image forensics with the goal to classifying between PIM and CG.

- In the process of image compression, using JPEG compression standard, the Discrete Cosine Transform (DCT) is a key element. JSteg steganography embeds secret information into JPEG format image by using the redundancies of DCT coefficients from JPEG images. By studying the distribution of DCT coefficients, we analyze the characteristics of the image before and after JSteg embedding. In Chapter 4, we mainly investigate steganalysis of JSteg algorithm.

- Due to imperfections during sensor manufacturing process, it is unavoidably that a variety of noises corrupt an image during its acquisition. Those noises are usually used as intrinsic fingerprints for identifying source camera origin. It is proposed to study the distribution of the noise (e.g. shot noise and readout noise) to accomplish the goal of source camera identification. In Chapter 5, we particularly explore the problem of source camera device identification

based on RAW data. Moreover, in Chapter 6, it is proposed to identify source camera device from JPEG images.

- Apart from the basic image acquisition process illustrated in Figure 2.1, image resampling is a typical post-camera technique which leaves very similar characteristics as demosaicing. By exploiting the linear relationship among neighboring pixels from a natural image, it is proposed to develop an algorithm to detect image resampling. In Chapter 7, we mainly study the image resampling detection.

Chapter 3

Natural Image Authentication by Statistical Modelling Analysis

Contents

3.1 Introduction and Contributions

To distinguish a photographic image (PIM) from a computer generated (CG) one is a research subfield of digital image forensics on which this chapter focuses. Driven by the pioneer work [30], most of the latest methods proposed to exploit high order statistics and/or physical features to distinguish PIM from CG using supervised learning method (such as Support Vector Machine, SVM). Although those methods can achieve high detection accuracy, supervised statistical learning is time-consuming. In addition, several problems such as the robustness to training and testing set mismatch and the establishment of detection performance remain open.

Note that PIM and CG images fundamentally differ as PIM images are obtained from a complex imaging process, see Figure C.1, while CG images are generated by a graphics software, not by a digital still camera. In [14], it is proposed to exploit the features of Color Filter Array (CFA) to detect CG images. Estimating the CFA pattern is also an effective approach to authenticate PIM images under some restrictive conditions, see [52]. Figure C.1 illustrates the image acquisition pipeline in a digital camera. Photons radiating from an object go though the optical system. Then, the CFA filters the light spectrum so that each pixel records only one color channel (red, green, or blue) ; the two missing color channels are padded by color interpolation. Finally, a natural image is generated after several image post-processes such as white balancing and gamma correction. In [67], the statistical features resulting from the image acquisition pipeline are used for differentiating PIM from CG images.

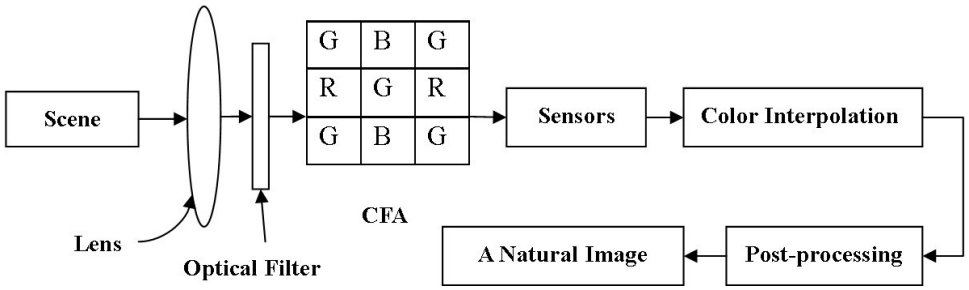


Figure 3.1: Illustration of the image acquisition pipeline.

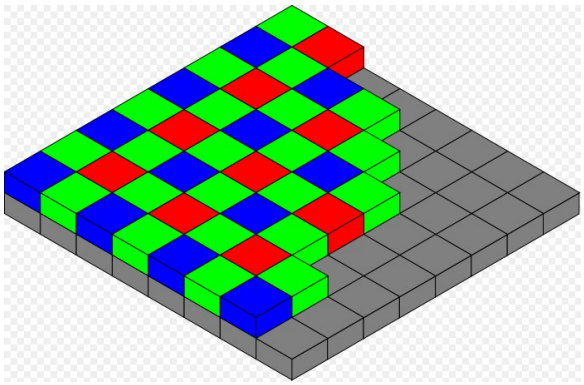


Figure 3.2: Bayer Model.

Figure C.2 shows the most widely-adopted CFA pattern referred to as *Bayer*. Based on the feature of Bayer model, PIM and CG images can be differentiated by the peak value in the frequency domain which is described specifically in Section 3.2. For a large number of PIM images the peak possibly vanishes, hence this hardly permits the distinguishing from CG images. Thus, an improvement of the method proposed in [54] is required.

Our proposed algorithm improves the method of [54], also referring to Gallagher's method, by two means. First, it is proposed to use the variance in the frequency domain on the assumption that image post-processes reduce the variance. Second, a linear parametric model is used to deal with nuisance parameters and based on the residual noise vector, a hypothesis testing model is established. Experimental results show the efficiency and the robustness of our proposed algorithm compared with the algorithm proposed in [54].

This chapter is organized as follows. Section 3.2 recalls Gallagher's method in [54]. Section 3.3 presents the proposed linear parametric model dealing with nuisance parameters. The Likelihood Ratio Test (LRT) is established in Section 3.4. Numerical experiments are presented in Section 3.5 and, finally, Section 3.6 concludes this chapter.

3.2 Outline of Gallagher's Method

In [54], Gallagher proposed to identify CG images by detecting the peak value of the image in the frequency domain. The specific algorithm is summarized below. Let us denote $\mathbf{I}(x, y, c)$ the pixel intensity of a given image with $c = \{r, g, b\}$ the color channel and (x, y) the pixel position. First, to avoid disturbances from low frequencies, $\mathbf{I}(x, y, g)$ is filtered by the following high-pass filter $\mathbf{H}(x, y)$:

$$\mathbf{H}(x, y) = \begin{bmatrix} 0 & 1 & 0 \\ 1 & -4 & 1 \\ 0 & 1 & 0 \end{bmatrix}.$$

Note that only the green channel is used because it carries more information, due to the specific Bayer's CFA pattern illustrated in Figure C.2, but an extension to other channels is straightforward. Then, the mean of all diagonal values, from the filtered image, is calculated to obtain a vector denoted $\mathbf{d} = (d_1, \ldots, d_N)^T$, where $n \in \{1, \cdots, N\}$ is the index number of the diagonal, N is the total number of diagonal and $\mathbf{M}^T$ is the transpose of matrix $\mathbf{M}$. Finally, it is proposed in [54] to use the frequential representation of $\mathbf{d}$, denoted $\mathbf{D}$ and formally defined as follows:

$$\mathbf{D} = |\mathcal{DFT}(\mathbf{d})| \text{ with } d_n = N_n^{-1} \sum_{x+y=n} |\mathbf{H} * \mathbf{I}(x, y, g)| \tag{3.1}$$

where $\mathcal{DFT}[\cdot]$ represents the calculation of DFT and N_n the number of pixels on the n_{th} diagonal.

Some examples of vectors $\mathbf{D}$, obtained from Gallagher's method described in (3.1), are given in Figure 3.3. Roughly speaking, the very simple test proposed in [54] consists in declaring a given image as a PIM image if a peak occurs at $\mathbf{D}(N/2)$. It is obvious that a typical PIM image 3.3a can be efficiently discriminated from CG images 3.3b and 3.3d. On the opposite, a true PIM image 3.3c is likely to be detected as a CG because it has no peak. Since the missing pixels whin each diagonal are padded by using the neighbourhood pixels acquired directly from the camera sensor, the interpolation will result in the periodical property among a vector in the spatial domain. Hence, the peak appears in Fourier domain. In fact, there is a large number of PIM images without peak, see Figure 3.3g, and thus, that may be wrongly detected as CG images by Gallagher's method [54]. Consequently, for reliability and efficiency purposes, it is necessary to improve the detection scheme proposed in [54].

Note that the peak value is not the only characteristic that distinguishes PIM from CG images. For almost every PIM image, see Figure 3.3e and 3.3g, the noise present in vector $\mathbf{D}$ has a much smaller variance than for CG images, see Figure 3.3. Due to image post-processing such as democaicing, the correlation among the neighbouring pixels is enhanced. Hence, the pixels acquired by a real camera result in the noise with smaller variance than its counterpart from the pixels generated by a computer. In the present chapter it is proposed to use this property of noise variance to distinguish PIM from CG images.

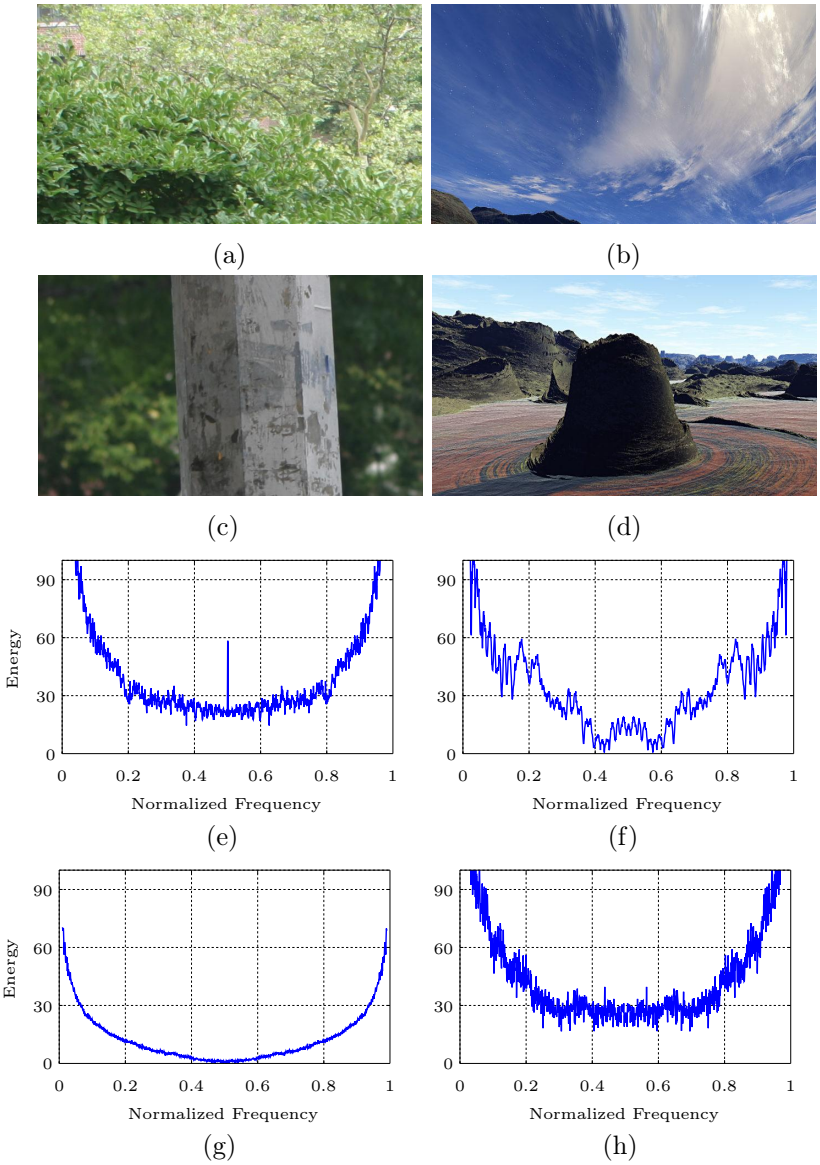


Figure 3.3: PIM (a)-(c) and CG (b)-(d) images together with their diagonal mean spectrum, (e)-(g) and (f)-(h) respectively.

3.3 Dealing with Nuisance Parameters

In this chapter, a linear parametric model is proposed to deal with diagonal mean spectrum $\mathbf{D}$. To this end, $\mathbf{D}$ is split into K non-overlapping vectors, denoted $\mathbf{y}_1, \ldots, \mathbf{y}_K$, of m samples. Let us define

$$\mathbf{y}_k \sim \mathcal{N}(\boldsymbol{\mu}_k, \sigma^2 \mathbf{I}_m) = \boldsymbol{\mu}_k + \xi_k. \tag{3.2}$$

where $\boldsymbol{\mu}_k = (\mu_{k,1}, \ldots, \mu_{k,m})^T$ of expectations, ξ_k is the realization of a Gaussian vector with variance $\sigma^2 \mathbf{I}_m$ and $\mathbf{I}_m$ the identity matrix of size $m \times m$. Hence, the problem

of distinguishing PIM from CG images can be formulated as a choice between the following hypotheses:

$$\begin{cases} \mathcal{H}_0 = \{\mathbf{y}_k \sim \mathcal{N}(\boldsymbol{\mu}_k, \sigma_0^2 \mathbf{I}_m), \forall k = (1, ..., K), \sigma_0 \leq \sigma\} \\ \mathcal{H}_1 = \{\mathbf{y}_k \sim \mathcal{N}(\boldsymbol{\mu}_k, \sigma_1^2 \mathbf{I}_m), \forall k = (1, ..., K), \sigma_1 > \sigma\} \end{cases} \tag{3.3}$$

where σ_0^2 and σ_1^2 respectively represent the variance under each hypothesis $\mathcal{H}_0 = \{\text{the image is PIM}\}$ and $\mathcal{H}_1 = \{\text{the image is CG}\}$ and σ is the threshold. Obviously, the expectation $\boldsymbol{\mu}_k$ is the nuisance parameter without any interest to distinguish PIM from CG images. Furthermore, $\mathbf{y}_k$ can be described with the following linear parametric model:

$$\boldsymbol{\mu}_k = \mathbf{A}\mathbf{x}_k, \tag{3.4}$$

where $\mathbf{A}$ is a known full rank matrix of size $m \times n$, with $m > n$, and $\mathbf{x}_k$ is a $n \times 1$ vector of parameters describing the expectation of $\mathbf{y}_k$.

The idea of using such linear parametric model is that it allows an easy elimination of nuisance parameter $\boldsymbol{\mu}_k$ which can be used in a hypothesis test using invariance theory [20, chap.6]. To apply this theory, let us define $R(\mathbf{A}) \subseteq \mathbb{R}^m$ the column space spanned by $\mathbf{A}$ and $R(\mathbf{A})^\perp \subseteq \mathbb{R}^{m-n}$ its orthogonal complement, sometimes referred to as the "parity space". The projection of observation vector $\mathbf{y}_k$ onto the parity space is obtained by $\mathbf{n}_k = \mathbf{W}\mathbf{y}_k$ where the matrix $\mathbf{W}$ verifies, among others, the following useful properties:

$$\mathbf{W}\mathbf{A} = \mathbf{0} \quad \text{and} \quad \mathbf{W}\mathbf{W}^T = \mathbf{I}_{m-n}. \tag{3.5}$$

Hence, by using the definitions of hypotheses (3.3), the projection of observation vector $\mathbf{y}_k$ onto the parity space $R(\mathbf{A})^\perp$ yields: $\mathbf{n}_k = \mathbf{W}\mathbf{y}_k = \mathbf{W}\xi_k \sim \mathcal{N}(0, \sigma_i^2 \mathbf{I}_{m-n})$ with $i = \{0, 1\}$ depending on the hypothesis.

Note that the use of projection matrix $\mathbf{W}$ can be replaced by using the Maximum Likelihood Estimation (MLE). By rejecting the nuisance parameter, let us define the estimation of $\mathbf{n}_k$ as

$$\widehat{\mathbf{n}}_k = \mathbf{y}_k - \widehat{\mathbf{y}}_k = \mathbf{P}_{\mathbf{A}}^{\perp}\mathbf{y}_k \quad \text{with} \quad \mathbf{P}_{\mathbf{A}}^{\perp} = \mathbf{I}_m - \mathbf{A}(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T. \tag{3.6}$$

In fact, a straightforward calculation, using the properties (7.5), shows that :

$$\| \mathbf{P}_{\mathbf{A}}^{\perp}\mathbf{y}_k \|_2^2 = \mathbf{y}_k^{\mathrm{T}}\mathbf{W}^{\mathrm{T}}\mathbf{W}\mathbf{W}^{\mathrm{T}}\mathbf{W}\mathbf{y}_k = \| \mathbf{W}\mathbf{y}_k \|_2^2.$$

For clarity, in the present chapter the matrix $\mathbf{W}$ is used in all calculus while matrix $\mathbf{P}_{\mathbf{A}}^{\perp}$ is used for illustrations and figures, see Figure 3.4, to keep the same number of observations.

Note that in the present chapter, the chosen linear parametric model is an algebraic polynomial of degree $n-1$; this yields the following matrix $\mathbf{A}$:

$$\mathbf{A} = \begin{pmatrix} 1 & 1 & 1 & \dots & \dots & 1 & 1 \\ 1 & 2 & 4 & \dots & \dots & 2^{n-2} & 2^{n-1} \\ \vdots & \vdots & \vdots & \ddots & \ddots & \vdots & \vdots \\ 1 & m & m^2 & \dots & \dots & m^{n-2} & m^{n-1} \end{pmatrix}$$

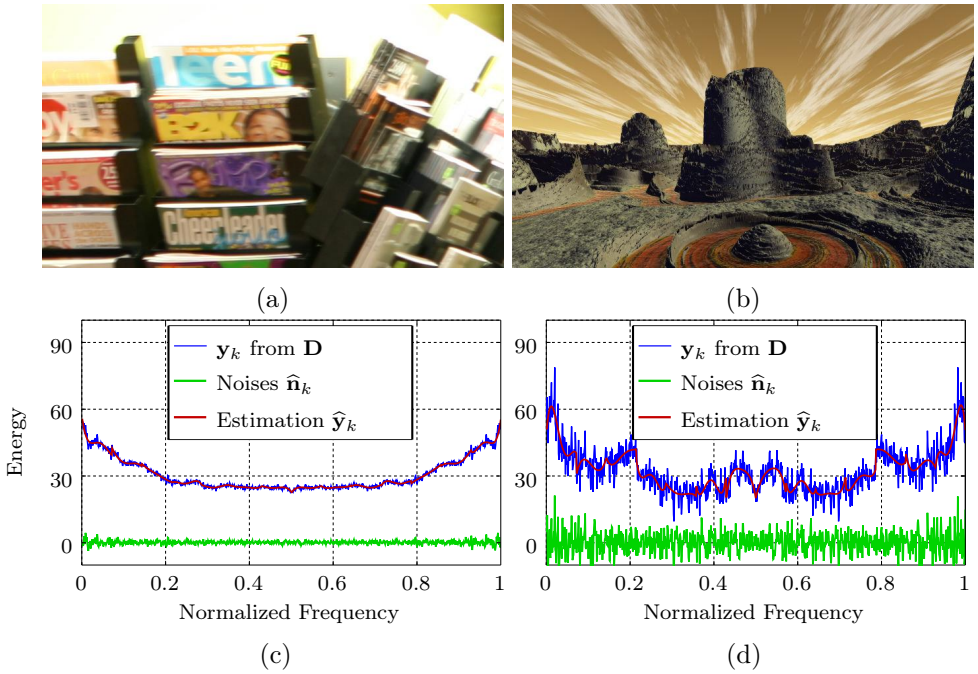


Figure 3.4: PIM (a) and CG (b) images with their diagonal mean spectrum, estimated expectation and residual noises (c)-(d).

In addition, as detailed in Section 3.5, it has been chosen to remove the few first and last samples from vector $\mathbf{D}$ as well as few samples around the $N/2$.

3.4 Likelihood Ratio Test Performances

By using a linear parametric model, as described in the methodology presented in Section 3.3, the problem of detecting PIM and CG images can be formulated as follows:

$$\begin{cases} \mathcal{H}_0 = \{\mathbf{n}_k \sim \mathcal{N}(\mathbf{0}, \sigma_0^2 \mathbf{I}_{m-n}), \forall k = (1, ..., K), \sigma_0 \leq \sigma^\star\} \\ \mathcal{H}_1 = \{\mathbf{n}_k \sim \mathcal{N}(\mathbf{0}, \sigma_1^2 \mathbf{I}_{m-n}), \forall k = (1, ..., K), \sigma_1 > \sigma^\star\}. \end{cases} \tag{3.7}$$

For solving the statistical detection problem such as (3.7), it follows from the Neyman-Pearson lemma [20, Theorem 3.2.1] that the Likelihood Ratio Test (LRT) is optimal in the sense described below. For definition, let

$$\mathcal{K}_\alpha = \left\{ \delta : \sup_{\sigma_0 \leq \sigma^\star} \mathbb{P}_{\mathcal{H}_0}[\delta(\mathbf{D}) = \mathcal{H}_1] \leq \alpha \right\} \tag{3.8}$$

be the class of tests, solving problem (3.7), with an upper-bounded false alarm probability α. Here $\mathbb{P}_{\mathcal{H}_j}[\cdot]$ is the probability under $\mathcal{H}_j, j \in \{0, 1\}$. Among all the tests in $\mathcal{K}_\alpha$ the LRT is the most powerful test, it maximizes the detection power

$$\beta_\delta = \mathbb{P}_{\mathcal{H}_1}[\delta(\mathbf{D}) = \mathcal{H}_1]. \tag{3.9}$$

From the statistical independence of vectors $\mathbf{y}_k$, the LRT is given by the following decision rule:

$$\delta(\mathbf{D}) = \begin{cases} \mathcal{H}_0 \text{ if } \Lambda(\mathbf{D}) = \sum_{k=1}^{K} \Lambda(\mathbf{y}_k) \leq \tau_\alpha \\ \mathcal{H}_1 \text{ if } \Lambda(\mathbf{D}) = \sum_{k=1}^{K} \Lambda(\mathbf{y}_k) > \tau_\alpha \end{cases} \tag{3.10}$$

where the decision threshold τ_α is the solution of equation $\sup_{\sigma_0 \leq \sigma^\star} \mathbb{P}_{\mathcal{H}_0}[\Lambda(\mathbf{D}) > \tau_\alpha] = \alpha$ to guarantee that $\delta(\mathbf{D}) \in \mathcal{K}_\alpha$. From the model of tested hypotheses, a straightforward calculation shows that the Likelihood Ratio (LR) $\Lambda(\mathbf{y}_k)$ is given by

$$\Lambda(\mathbf{y}_k) = \|\mathbf{n}_k\|_2^2.$$

Finally, from the statistical distribution of noise residuals $\mathbf{n}_k$, and from the properties of Gaussian random variables, one immediately obtains that under hypothesis $\mathcal{H}_i, i = \{0; 1\}$:

$$\begin{aligned} & \tfrac{1}{\sigma_i^2}\|\mathbf{n}_k\|_2^2 \sim \chi^2_{m-n} \\ \Leftrightarrow \quad & \Lambda(\mathbf{y}_k) = \|\mathbf{n}_k\|_2^2 \sim \Gamma(\tfrac{m-n}{2}, 2\sigma_i^2) \end{aligned}$$

where $\Gamma(\frac{m-n}{2}, 2\sigma_i^2)$ represents the Gamma distribution with a shape parameter $\frac{m-n}{2}$ and a scale parameter $2\sigma_i^2$. Subsequently, it follows from the stability under summation of Gamma random variables, that the statistical distribution of the LR $\Lambda(\mathbf{D})$ is given under hypothesis $\mathcal{H}_i, i = \{0; 1\}$ by:

$$\Lambda(\mathbf{D}) \sim \Gamma\left(\frac{K(m-n)}{2}, 2\sigma_i^2\right) \tag{3.11}$$

It is thus immediate to establish the statistical properties of the proposed test (3.11) which are given in the following theorems; for clarity, $\mathbf{F}_\Gamma(\cdot)$ and $\mathbf{F}_\Gamma^{-1}(\cdot)$ represent the Gamma cumulative distribution function and its inverse respectively.

Theorem 3.1. *Assume that the model hypothesis (3.3) holds, then for any $\alpha \in (0; 1)$ the decision threshold:*

$$\tau_\alpha = F_\Gamma^{-1}\left(1 - \alpha; \frac{K(m-n)}{2}, 2\sigma_0^2\right) \tag{3.12}$$

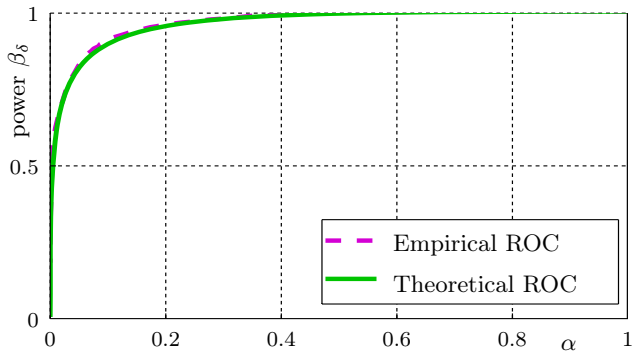
guarantees that the LRT δ (3.10) is in the class $\mathcal{K}_\alpha$.

Theorem 3.2. *Assume that the model hypothesis (3.3) holds, for any decision threshold $\tau_\alpha \in \mathbb{R}$, the power function associated with the test δ (3.10) is given by*

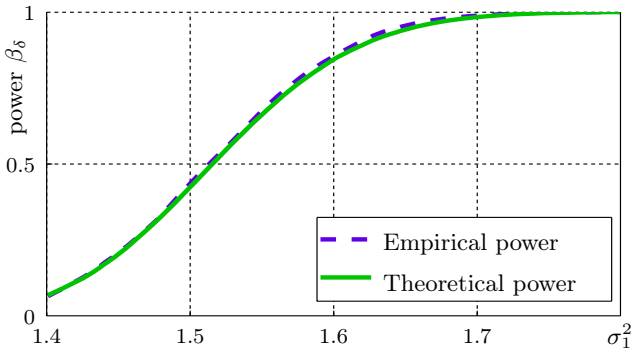
$$\beta_\delta = 1 - F_\Gamma\left(\tau_\alpha; \frac{K(m-n)}{2}, 2\sigma_1^2\right) \tag{3.13}$$

3.5 Numerical Results

To verify the sharpness of the theoretically established results, a Monte-Carlo simulation is performed. Prior to our experiments, it is proposed to use an image



(a) Detection power as a function of FAR α (ROC curve).



(b) Detection power as a function of the variance expectation σ_1^2.

Figure 3.5: Comparison between theoretically established and empirically obtained performance of the proposed test (3.10).

database containing 300 PIM images (with 150 images from Nikon D70 and 150 images from Canon 10D) from Columbia's ADVENT dataset [144] and 300 CG images downloaded from www.pandromeda.com. All these 600 images are cropped to the dimension of 700×400 pixels and finally saved as JPEG format with the quality factor 85.

The parametric linear model (3.4) is defined by a polynomial order $n - 1 = 4$ and the size of vector $\mathbf{y}_k$ is set to $m = 64$ samples. Note that to avoid dealing with different variance, possibly non-uniform, the first, last and middle samples are excluded from analysis. To this end, it is proposed in practice not to consider the first, $\mathbf{y}_1$, the last, $\mathbf{y}_K$, and the two middle vectors, $\mathbf{y}_{K/2}$ and , $\mathbf{y}_{K/2+1}$. From the remaining sample, the variance of $\mathbf{D}$ is calculated using $\mathbf{n}_k$ in each image. Two sets containing 10000 vectors of 768 samples[1] are randomly generated with zero mean and variance $\sigma_0^2 = 1.39$, to simulate residual noises from PIM images, or hypothesis $\mathcal{H}_0$, and with variance $\sigma_1^2 = 1.80$ to simulate CG images, hypothesis $\mathcal{H}_1$.[2] In fact, the

[1] It should be noted that our linear model is established on each 64 samples from the whole vector. Hence the number of each vector for simulation is set as 768, which is a multiple of 64.

[2] The values from σ_0^2 and σ_1^2 are chosen based on the mean value of the practical calculation from two kinds of images.

larger the distance between σ_0^2 and σ_1^2 is, the better the proposed detector performs.

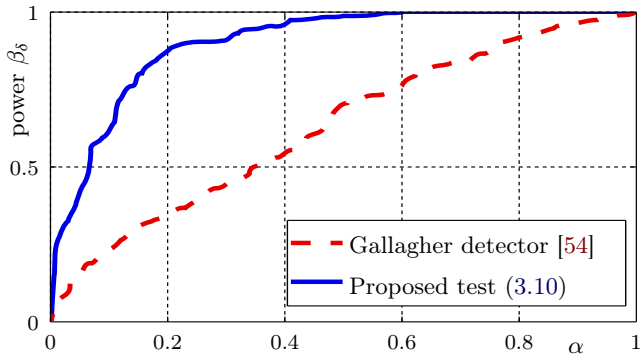
The detection performances obtained with the proposed test are illustrated in Figure 3.5a; the Receiver Operating Characteristic (ROC), that is the detection power β_δ as a function of false alarm probability α, of both empirical and theoretically established results (3.13) are compared. Similarly, Figure 3.5b shows a comparison between empirical and theoretical detection power as a function samples variance, σ_1^2, under alternative hypothesis $\mathcal{H}_1$. The numerical results presented in Figure 3.5b are obtained with a false alarm probability set to $\alpha = 0.05$. From Figure 3.5a and 3.5b it is obvious that the empirical detection powers are almost identical to the theoretically established ones (3.13); this shows the sharpness and the relevance of theoretical findings.

To emphasize the improvement of the proposed test compared with Gallagher's method proposed in [54], Figure **??** presents the detection performance, as ROC curves, of those detectors. It can be noted that, for instance, for $\alpha = 0.2$, the proposed test power is above 0.85 while it is below 0.35 using Gallagher's method. In addition, when the detected images with the low image quality, the proposed test preserves a high detection performance. This is emphasized in Figure **??**, which presents ROC curves of the proposed test for uncompressed images and compressed images with JPEG standard and quality factors ranging from 55 to 95.

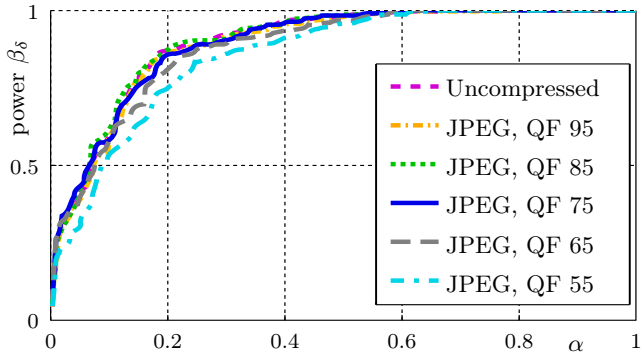
Finally it is proposed to study the detection performance of the proposed test in the context of counter-forensics. First, it is proposed to apply Gaussian blur to both PIM and CG on the assumption that such a blurring process should largely reduce the noise variance. In addition, PIM images are subjected to deterministic degradations, which are well modelled as blurring processes, during its acquisition. Figure 3.7a shows the empirically obtained detection performance, of both the proposed test and the method proposed in [54], after image blurring. Second, knowing the image acquisition pipeline, it is reasonable to assume that one may try to simulate the CFA interpolation in CG images. Hence, it is proposed in this chapter to apply the well known bi-linear demosaicing filter on all the CG images. The empirically obtained results are shown in Figure 3.7b; results particularly highlight that the detection method proposed in [54] performs poorly. In fact since most of the PIM images have a small peak, or no peak at all, the simulation of CFA interpolation artificially creates a periodic pattern which results in a peak in CG diagonal mean spectrum. The method proposed in [54] hence easily classifies CG images as PIM images as soon as the CFA interpolation process is simulated. The results from Figure 3.7 emphasize the lack of the robustness of the detection method proposed in [54] and, on the opposite, highlight the efficiency as well as the good robustness of the proposed statistical test (3.10).

3.6 Conclusion

In this chapter, we describe an approach of distinguishing between PIM and CG images based on statistical decision theory. A linear parametric model is developed



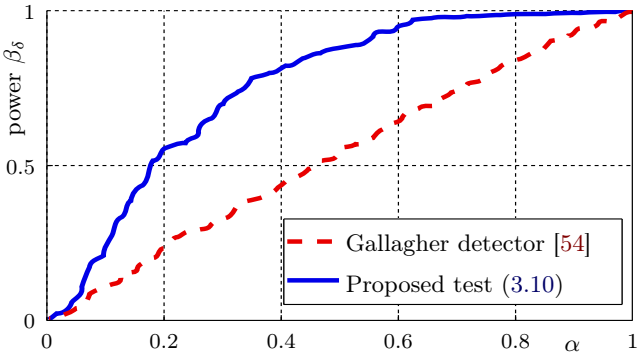
(a) Comparison of the two methodologies detection performance, ROC curves.



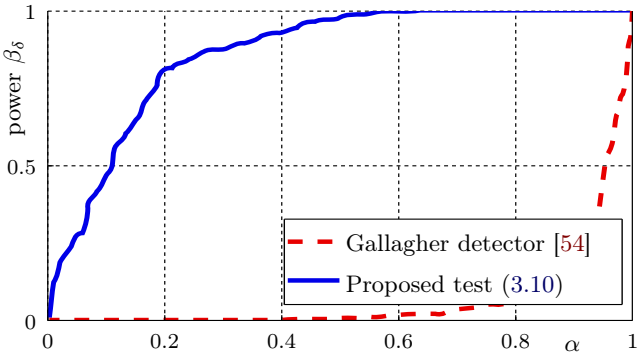
(b) Detection performance for different image quality factor (QF) of JPEG compression.

Figure 3.6: Illustration of the proposed test (3.10) performance for real PIM and CG images.

to deal with nuisance parameters. By using the residual noise $\mathbf{n}_k$ representing the property of each detected image, hypothesis testing model is exploited to categorize two kinds of images. The method proposed in this chapter outperforms the detector proposed in [54] and improves the detection accuracy. Moreover, experimental results also emphasize that the proposed method has a good robustness with respect to basic counter-forensic techniques.



(a) Comparison of robustness with respect to Gaussian blur: detection performance of the two methodologies.



(b) Comparison of robustness with respect to simulated C-FA interpolation.

Figure 3.7: Comparison of detection performance in the presence of counter-forensic process.

Chapter 4

Steganalysis of JSteg Algorithm using Hypothesis Testing Theory

Contents

4.1 Introduction and Contributions

In today's digital world, there exists many steganographic tools available on the Internet. Due to the fact that some of them are readily available and very simple to use, it is necessary to design the most reliable steganalysis methodology to counter against steganography. Although the LSB replacement steganalysis method (see [129, 130, 132, 145–147]) has been studied for many years, it can be noted that most of the prior-art detectors are designed to detect data hidden in the spatial domain. In addition, for only a few detectors the statistical properties have been studied and established, referred to as the optimal detectors. As detailed in [124], a wide range of problems, theoretical as well as practical, remain uncovered and

some prevent the moving of "steganography and steganalysis from the laboratory into the real world". This is especially the case in the field of Optimal Detection, see [124, Sec. 3.1], in which this chapter lies. Roughly speaking, the goal of optimal detection in steganalysis, is to exploit an accurate statistical model of cover source, usually digital images, to design a statistical test whose properties can be established; typically, in order to guarantee a False Alarm Rate (FAR) and to calculate the optimal detection performance one can expect from the most powerful detector.

In 2004, the Weighted Stego-image (WS) method [120] and the test proposed in [121] for LSB replacement steganalysis changed the situation opening the way to optimal detectors. Driven by these pioneer works, the enhanced WS algorithm proposed in [122] improved the detection rate by enhancing pixels predictor, adjusting weighting factor and introducing the concept of bias correction. Nevertheless, the drawback of original WS method is that it can only be applied in the spatial domain. Due to the prevalence of images compressed in the Joint Photographic Experts Group (JPEG) format, how to deal with this kind of images becomes mandatory. Inspired by the prior studies [120, 122], the WS steganalyzer for JPEG covers was proposed in [123]. However, the WS steganalyzer does not allow to get high detection performance for a low FAR, see [148], and its statistical properties remain unknown, which prevents us from guaranteeing a prescribed FAR. In practical forensic cases, since a large database of images needs to be processed, the getting of a very low FAR is crucial.

For the detection of data hidden within the Discrete Cosine Transform (DCT) coefficients of JPEG images, the application of hypothesis testing theory for designing optimal detectors, that are efficient in practice, is facing the problem of accurately modelling statistical distribution of DCT coefficients. It can be noted that several models have been proposed in the literature to model statistically the DCT coefficients. Among those models, the Laplacian distribution is probably the most widely used due to its simplicity and its fairly good accuracy [149]. More accurate models such as the Generalized Gaussian [150] and, more recently, the Generalized Gamma model [151] have provided much more accuracy at the cost of higher complexity. Some of those models have been exploited in the field of steganalysis, see [152, 153] for instance. In the framework of optimal detection, a first attempt has been made to design a statistical test modelling the DCT coefficient with the quantized Laplacian distribution, see [135].

It should be noted that other approaches have been proposed for the detection of data hidden within DCT coefficients of JPEG images, to cite a few, the structural detection [116], the category attack [154], the WS detector [123], and universal or blind detectors [64, 138]. However, establishing the statistical properties of those detectors remains a difficult work which has not been studied yet. In addition, most accurate detectors based on statistical learning are sensitive to the so-called cover source-mismatch [155]: the training phase must be performed with caution.

In this context, the detector proposed in [135] is an interesting alternative; however it is based on the assumption that DCT coefficients are independent and identically distributed (i. i. d.) within a subband and have a zero expectation which

might be inaccurate and hence make the detection performance poor in practice. In practice, this model is not independent of the image content, which performs well only in the case of high-texture image (see Figure C.4a), but hardly holds true in the case of low-texture image (see Figure C.4b). On the opposite, this chapter proposes a statistical model assuming that each DCT coefficient has a different expectation and variance. The use of this model, together with hypothesis theory, allows us to design the most powerful Likelihood Ratio Test (LRT) when the distribution parameters (expectation and variance) are known. Then in the practical case of not knowing those parameters, estimations have to be used instead; this leads to the design of the proposed detector with estimated parameters. By taking into account those distribution parameters as nuisance parameters and using an accurate estimation, it is shown that the loss of power compared with the optimal detector is small.

Therefore, the contributions of this chapter are as follows:

1. First, a novel model of DCT coefficients is proposed; its major originality is that this model does not assume that all the coefficients of the same subband are i. i. d.

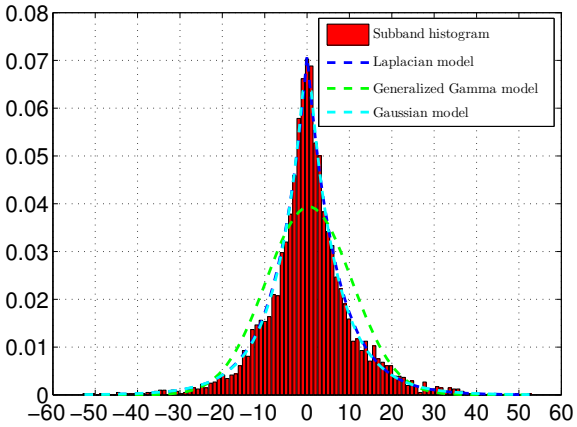
2. Second, assuming that all the parameters are known, this statistical model of DCT coefficients is used to design the optimal test to detect data hidden within JPEG images with JSteg algorithm. This statistical test takes into account distribution parameters of each DCT coefficient as nuisance parameters.

3. Further, assuming that all the parameters are unknown, a simple approach is proposed to estimate the expectation (or location parameter) of each coefficient by using linear properties of DCT as well as estimation of pixel expectation in the spatial domain; the variance (or scale parameter) is also estimated locally.

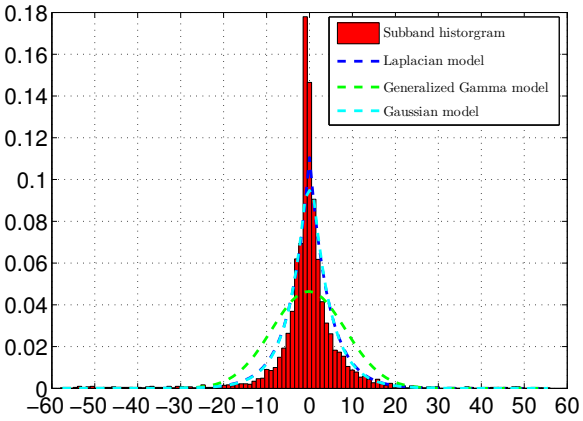
4. The designed detector is improved by exploring a DCT channel selection, which has been proposed very recently [156, 157], that selects only a sub-set of pixels or DCT coefficients in which embedding is most likely. Hence the image within JSteg embedding is easier to detect.

5. Numerical results show the sharpness of the theoretically established results and the good performance of the proposed statistical test. A comparison with the statistical test based on the Laplacian distribution as well as the assumption of i. i. d. coefficient, see [135], shows the relevance of the proposed methodology. In addition, compared with prior-art WS detector [123], experimental results show the efficiency of the proposed detector.

This chapter is organized as follows. Section 4.2 formalizes the statistical problem of detection of information hidden within DCT coefficients of JPEG images. Then, Section 4.3 presents the optimal Likelihood Ratio Test (LRT) for detecting the JSteg algorithm based on the Laplacian distribution model. Section 4.4



(a) *Baboon*



(b) *Sky*

Figure 4.1: Illustration of the quantized DCT coefficient subband (2,1), (a) a high-texture image: *Baboon*, (b) a low-texture image: *Sky*.

presents the proposed approach for estimating the nuisance parameters in practice, and compares our proposed detector with WS detector [123] theoretically. Finally, Section 4.5 presents numerical results of the proposed steganalyzer on simulated and real images and Section 4.6 concludes this chapter.

4.2 Problem Statement

In this chapter, a grayscale digital image is represented, in the spatial domain, by a single matrix $\mathbf{Z} = \{z_{i,j}\}, i \in \{1, \ldots, I\}, j \in \{1, \ldots, J\}$. The present work can be extended to colour image by analyzing each colour channel separately. Most of digital images are stored using the JPEG compression standard. This standard exploits the linear DCT, over blocks of 8×8 pixels to represent an image in the so-called DCT domain. In the present chapter, we avoid the description of the imaging pipeline of a digital still camera; the reader can refer to [41] for a description of the whole imaging pipeline and to [158] for a detailed description of the JPEG

compression standard.

Let us denote DCT coefficients by the matrix $\mathbf{V} = \{v_{i,j}\}$. An alternative representation of those coefficients is usually adopted by gathering the DCT coefficients that corresponds to the same frequency subband. In this chapter, this alternative representation is denoted by the matrix $\mathbf{U} = \{u_{k,l}\}, k \in \{1, \dots, K\}, l \in \{1, \dots, 64\}$ with $K \approx I \times J/64$[1].

The coefficients from the first subband $u_{k,1}$, often referred to as DC coefficients, represent the mean of pixels value over k-th block of 8×8 pixels. The modification of those coefficients may be obvious and creates artifacts that can be detected easily, hence, they are usually not used for data hiding. Similarly, the JSteg algorithm does not use the coefficients from the other subbands, referred to as AC coefficients, if they equal 0 or 1. In fact, it is known that using the coefficients equal to 0 or 1 modifies significantly statistical properties of AC coefficients; this creates a flaw that can be detected.

The JSteg algorithm embeds data within DCT coefficients of JPEG images using the well-known LSB (Least Significant Bit) replacement method, see details in [159]. In brief, this method consists in substituting the LSB of each DCT coefficient by a bit of the message it is aimed to hide. The number of hidden bits per coefficient, usually referred to as the payload, is denoted $R \in (0, 1]$. Since the JSteg algorithm does not use each DCT coefficient, the payload will in fact be measured in this chapter as the number of hidden bits per *usable* coefficients ,that is the number of bits divided by the number of AC coefficients that differ from 0 and 1.

Let us assume that the DCT coefficients are independent and that they all follow the same probability distribution, denoted $\mathcal{P}_\theta$, parameterized by the parameter θ which may change among the coefficients. Since the DCT coefficients can only take value into a discrete set, the distribution $\mathcal{P}_\theta$ may be represented by its probability mass function (pmf) denoted $P_\theta = \{p_\theta[u]\}$; for simplicity[2], it is assumed in this chapter that $u \in \mathbb{Z}$. Let us denote $\mathcal{Q}_\theta^R$ the probability distribution of *usable* DCT coefficients from the stego-image, after embedding a message with payload R. A short calculation shows that, see [120,121,132], the stego-image distribution may be represented with following the pmf $Q_\theta^R = \{q_\theta^R[u]\}_{u \in \mathbb{Z}}$ where

$$q_\theta^R[u] = (1 - \frac{R}{2})p_\theta[u] + \frac{R}{2}p_\theta[\bar{u}], \tag{4.1}$$

and $\bar{u} = u + (-1)^u$ represents the integer u with flipped LSB. For the sake of clarity, let us denote $\theta_{k,l}$ the distribution parameter of k-th DCT coefficient from l-th subband and let $\boldsymbol{\theta} = \{\theta_{k,l}\}, k \in \{1, \dots, K\}, l \in \{2, \dots, 64\}$ represent the distribution parameter of all the AC coefficients.

When inspecting a given JPEG image, more precisely its DCT coefficient matrix $\mathbf{U}$, in order to detect data hidden with the JSteg algorithm, the problem consists in choosing between the two following hypotheses $\mathcal{H}_0$: "*the coefficients $u_{k,l}$ follow the*

[1] In this chapter we assume, without loss of generality, that both width and height of an inspected image are multiples of 8.

[2] In practice, DCT coefficients belong to set $[-1024, \dots, 1023]$, see [135].

distribution $\mathcal{P}_{\theta_{k,l}}$" and $\mathcal{H}_1$: "*the coefficients* $u_{k,l}$ *follow the distribution* $\mathcal{Q}^R_{\theta_{k,l}}$" which can be written formally as:

$$\begin{cases} \mathcal{H}_0 : \left\{ u_{k,l} \sim \mathcal{P}_{\theta_{k,l}}, \forall k \in \{1,\dots,K\}, \forall l \in \{2,\dots,64\} \right\}, \\ \mathcal{H}_1 : \left\{ u_{k,l} \sim \mathcal{Q}^R_{\theta_{k,l}}, \forall k \in \{1,\dots,K\}, \forall l \in \{2,\dots,64\} \right\}. \end{cases} \tag{4.2}$$

A statistical test is a mapping $\delta : \mathbb{Z}^{I \cdot J} \mapsto \{\mathcal{H}_0, \mathcal{H}_1\}$ such that hypothesis $\mathcal{H}_i$ is accepted if $\delta(\mathbf{U}) = \mathcal{H}_i$ (see [127] for details on hypothesis testing). As previously explained, this chapter focuses on the Neyman-Pearson bi-criteria approach: maximising the correct detection probability for a given false alarm probability α_0. Let:

$$\mathcal{K}_{\alpha_0} = \left\{ \delta : \sup_{\boldsymbol{\theta}} \mathbb{P}_{\mathcal{H}_0}[\delta(\mathbf{U}) = \mathcal{H}_1] \leq \alpha_0 \right\}, \tag{4.3}$$

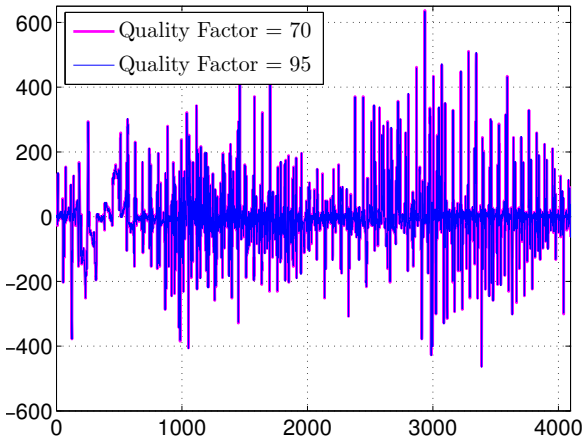
be the class of tests with a false alarm probability upper-bounded by α_0. Here $\mathbb{P}_{\mathcal{H}_i}(A)$ stands for the probability of event A under hypothesis $\mathcal{H}_i, i = \{0, 1\}$, and the supremum over $\boldsymbol{\theta}$ has to be understood as whatever the distribution parameters might be, in order to ensure that the false alarm probability α_0 can not be exceed. Among all the tests in $\mathcal{K}_{\alpha_0}$, it is aimed at finding a test δ which maximizes the power function, defined by the correct detection probability:

$$\beta_\delta = \mathbb{P}_{\mathcal{H}_1}[\delta(\mathbf{U}) = \mathcal{H}_1], \tag{4.4}$$

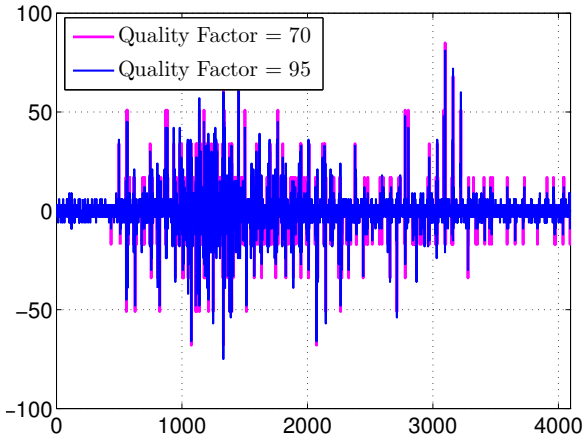
which is equivalent to minimize the missed detection probability $\alpha_1(\delta) = \mathbb{P}_{\mathcal{H}_1}[\delta(\mathbf{U}) = \mathcal{H}_0] = 1 - \beta_\delta$.

In order to design a practical *optimal detector*, as referred in [124], for steganalysis in spatial domain, the main difficulty is to estimate the distribution parameters, that are expectation and variance of each pixel. In the case of DCT coefficients, the application of hypothesis testing theory to design an optimal detector has previously being attempted with the assumption that the distribution parameter remains the same for all the coefficients from a same subband. With this assumption, the estimation of the distribution parameters is not an issue because thousands of DCT coefficients are available. However which distribution model to choose remains an open problem.

The hypothesis testing theory has been applied for the steganalysis of JSteg algorithm in [135] using a Laplacian distribution model and using the assumption that DCT coefficients of each subband are i. i. d. However, this pioneer work does not allow the design of an efficient test because a very important loss of performance has been observed when comparing results on real images and theoretically established ones. Such a result can be explained by the two following reasons: 1) the Laplacian model might be not accurate enough to detect stegonagraphy and 2) the assumption that the DCT coefficients of each frequency subband are i. i. d. may be wrong. Recently, it has been shown that the use of Generalized Gamma model or even more accurate model [136, 160] allows the designing of a test with very good detection performance. In this chapter, it is proposed to challenge the assumption that all the DCT coefficients of a subband are i. i. d.



(a) Subband (1,2)

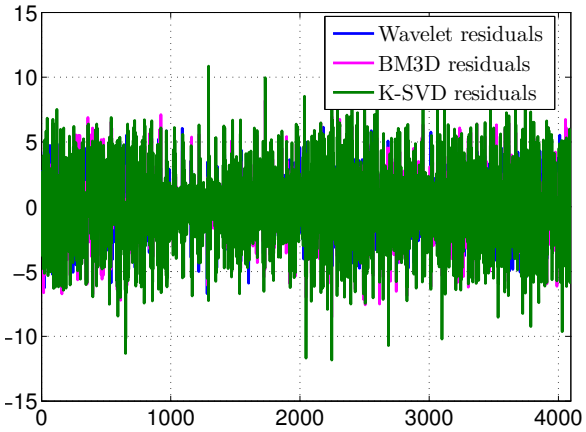


(b) Subband (4,4)

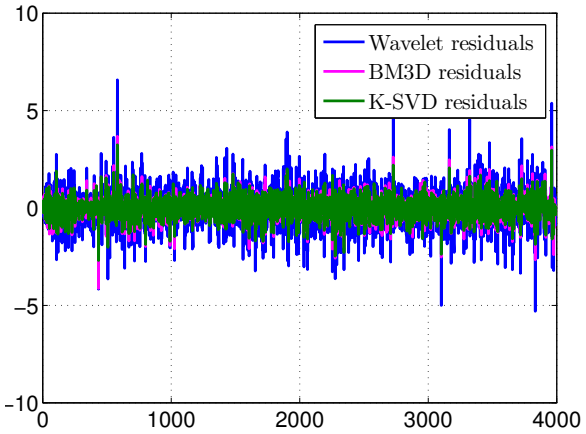
Figure 4.2: Illustrative examples of the value of DCT coefficients of two subbands from *lena* image. Those examples show that the assumption that DCT coefficients are i. i. d. within a subband hardly holds true in practice. Horizontal coordinate: index of the coefficient; vertical coordinate: value of the corresponding coefficient.

A typical example is given by Figure 4.2 and 4.3. Figure 4.2a (resp. Figure 4.2b) represents the DCT coefficients of the subband (1,2) (resp. subband (4,4)) extracted from the image *lena*. Observing those two graphs, it is obvious that the assumption of all those coefficients being i. i. d. is doubtful. However, if it is assumed that each coefficient has a different expectation, one can estimate this expected value and compute the "residual noise", that is the difference between the observation and the computed expectation. Such results are shown in Figure 4.3, with three different models for estimating the expectation of DCT coefficients of the same two subbands from *lena*. Moreover, Figure 4.4 illustrates the distribution of residual noises which are plotted in Figure 4.3. Obviously, residual noises look much more i. i. d. than the original DCT coefficients.

In the following section, we detail the statistical test that takes into account both



(a) Subband (1,2)



(b) Subband (4,4)

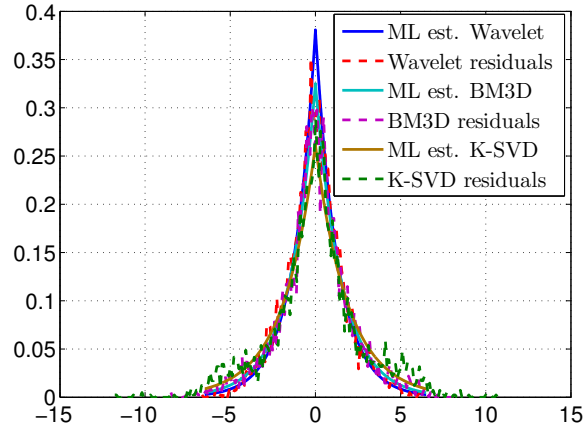
Figure 4.3: Illustrative examples of DCT coefficients of residual noise, obtained by a denoised image. The same two DCT subbands, as in Figure 4.2 are extracted from the residual noise of *lena* image. On those examples, the assumption of i.i.d. distribution seems more realistic.

the expectation and the variance as nuisance parameters and we study the optimal detection when those parameters are known. A discussion on nuisance parameters is also provided in Section 4.4.

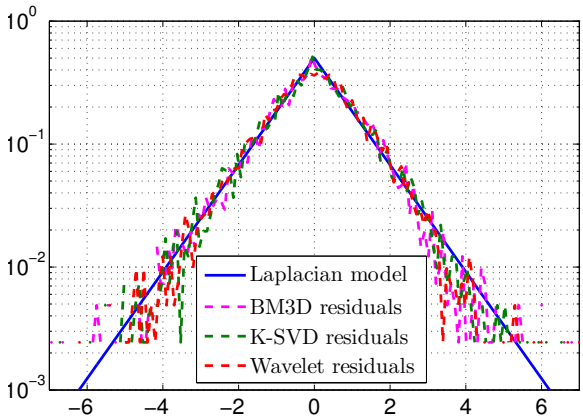
4.3 LRT for Two Simple Hypotheses

4.3.1 Optimal Detection Framework

When the payload R and the distribution parameters $\boldsymbol{\theta} = \{\theta_{k,l}\}, k \in \{1, \ldots, K\}, l \in \{2, \ldots, 64\}$ are known, problem (4.2) is reduced to a statistical test between two simple hypotheses. In such a case, the Neyman-Pearson Lemma [127, theorem 3.2.1] states that the most powerful test in the class $\mathcal{K}_{\alpha_0}$ (4.3) is the LRT defined, on the



(a) Subband (1,2)



(b) Subband (4,4)

Figure 4.4: Statistical distribution of the DCT coefficients of the residual noise plotted in Figure 4.3. For comparison, the Laplacian pdf with parameters estimated by the MLE are also shown in Figure 4.4a. Note that for a meaning comparison, Figure 4.4b shows the results after normalization by the estimated scale parameter $\widehat{b}_k$.

assumption that DCT coefficients are independent, as:

$$\delta^{\mathrm{lr}}(\mathbf{U}) = \begin{cases} \mathcal{H}_0 \ \text{ if } \Lambda^{\mathrm{lr}}(\mathbf{U}) = \displaystyle\sum_{k=1}^{K}\sum_{l=2}^{64} \Lambda^{\mathrm{lr}}(u_{k,l}) < \tau^{\mathrm{lr}}, \\ \mathcal{H}_1 \ \text{ if } \Lambda^{\mathrm{lr}}(\mathbf{U}) = \displaystyle\sum_{k=1}^{K}\sum_{l=2}^{64} \Lambda^{\mathrm{lr}}(u_{k,l}) \geq \tau^{\mathrm{lr}}, \end{cases} \tag{4.5}$$

where the decision threshold τ^{lr} is the solution of the equation $\mathbb{P}_{\mathcal{H}_0}\left[\Lambda^{\mathrm{lr}}(\mathbf{U}) \geq \tau^{\mathrm{lr}}\right] = \alpha_0$, to ensure that the false alarm probability of the LRT equals α_0, and the log Likelihood Ratio (LR) for one observation is given, by definition, by:

$$\Lambda^{\mathrm{lr}}(u_{k,l}) = \log\left(\frac{q^{R}_{\theta_{k,l}}[u_{k,l}]}{p_{\theta_{k,l}}[u_{k,l}]}\right). \tag{4.6}$$

In practice, when the rate R is not known one can try to design a test which is locally optimal around a given payload rate, named Locally Asymptotically Uniformly Most Powerful (LAUMP) test, as proposed in [130, 132] but this lies outside the scope of this chapter.

From the definition of $p_{\theta_{k,l}}[u_{k,l}]$ and $q^R_{\theta_{k,l}}[u_{k,l}]$ (4.1), it is easy to write the LR (4.6) as:

$$\Lambda^{\text{lr}}(u_{k,l}) = \log\left(1 - \frac{R}{2} + \frac{R}{2}\frac{p_{\theta_{k,l}}[\bar{u}_{k,l}]}{p_{\theta_{k,l}}[u_{k,l}]}\right), \tag{4.7}$$

where, as previously defined, $\bar{u}_{k,l} = u_{k,l} + (-1)^{u_{k,l}}$ represents the DCT coefficient $u_{k,l}$ with flipped LSB.

4.3.2 Statistical Performance of LRT

Accepting, for a moment, that one is in this most favorable scenario, in which all the parameters are perfectly known, we can deduce some interesting results. Due to the fact that observations are considered to be independent, the LR $\Lambda^{\text{lr}}(\mathbf{U})$ is the sum of random variables and some asymptotic theorems allow to establish its distribution when the number of coefficients becomes "sufficiently large". This asymptotic approach is usually verified in the case of digital images due to the very large number of pixels or DCT coefficients.

Let us denote $E_{\mathcal{H}_i}(\theta_{k,l})$ and $V_{\mathcal{H}_i}(\theta_{k,l})$ the expectation and the variance of the LR $\Lambda^{\text{lr}}(u_{k,l})$ under hypothesis $\mathcal{H}_i, i = \{0, 1\}$. Those quantity obviously depend on the parameterized distribution $\mathcal{P}_{\theta_{k,l}}$. The Lindeberg's central limit theorem (CLT) [127, theorem 11.2.5] states that as K tends to infinity it holds true that[3]:

$$\frac{\sum_{k=1}^{K}\sum_{l=2}^{64}\Lambda^{\text{lr}}(u_{k,l}) - E_{\mathcal{H}_i}(\theta_{k,l})}{\left(\sum_{k=1}^{K}\sum_{l=2}^{64}V_{\mathcal{H}_i}(\theta_{k,l})\right)^{1/2}} \xrightarrow{d} \mathcal{N}(0,1)\ ,\ i = \{0,1\}\,, \tag{4.8}$$

where $\xrightarrow{d}$ represents the convergence in distribution and $\mathcal{N}(0,1)$ is the standard normal distribution with zero mean and unit variance.

This theorem is of crucial interest to establish the statistical properties of the proposed test [129, 134–136]. In fact, once the moments have been calculated under both $\mathcal{H}_i, i = \{0, 1\}$, one can normalize under hypothesis $\mathcal{H}_0$ the LR $\Lambda^{\text{lr}}(\mathbf{U})$ as follows:

$$\begin{aligned}\overline{\Lambda}^{\text{lr}}(\mathbf{U}) &= \frac{\Lambda^{\text{lr}}(\mathbf{U}) - \sum_{k=1}^{K}\sum_{l=2}^{64}E_{\mathcal{H}_0}(\theta_{k,l})}{\left(\sum_{k=1}^{K}\sum_{l=2}^{64}V_{\mathcal{H}_0}(\theta_{k,l})\right)^{1/2}},\\ &= \frac{\sum_{k=1}^{K}\sum_{l=2}^{64}\Lambda^{\text{lr}}(u_{k,l}) - E_{\mathcal{H}_0}(\theta_{k,l})}{\left(\sum_{k=1}^{K}\sum_{l=2}^{64}V_{\mathcal{H}_0}(\theta_{k,l})\right)^{1/2}}.\end{aligned} \tag{4.9}$$

[3]Note that we refer to the Lindeberg's CLT, whose conditions are easily verified in our case, because the random variable are independent but are not i. i. d.

Since this essentially consists in adding a deterministic value and scaling the L-R, this operation of normalization preserves the optimality of the LRT. It is thus straightforward to define the normalized LRT with $\overline{\Lambda}^{\text{lr}}(\mathbf{U})$ by:

$$\overline{\delta}^{\text{lr}}(\mathbf{U}) = \begin{cases} \mathcal{H}_0 \text{ if } \overline{\Lambda}^{\text{lr}}(\mathbf{U}) < \overline{\tau}^{\text{lr}} \\ \mathcal{H}_1 \text{ if } \overline{\Lambda}^{\text{lr}}(\mathbf{U}) \geq \overline{\tau}^{\text{lr}}. \end{cases} \tag{4.10}$$

It immediately follows from Lindeberg's CLT (4.8) that $\overline{\Lambda}^{\text{lr}}(\mathbf{U})$ asymptotically follows, as K tends to infinity, the normal distribution $\mathcal{N}(0,1)$. Hence, it is immediate to set the decision threshold that guarantees the prescribed false alarm probability:

$$\overline{\tau}^{\text{lr}} = \Phi^{-1}(1-\alpha_0), \tag{4.11}$$

where Φ and Φ^{-1} respectively represent the cumulative distribution function (cdf) of the standard normal distribution and its inverse. Similarly, denoting

$$m_i = \sum_{k=1}^{K}\sum_{l=2}^{64} E_{\mathcal{H}_i}(\theta_{k,l}); \sigma_i^2 = \sum_{k=1}^{K}\sum_{l=2}^{64} V_{\mathcal{H}_i}(\theta_{k,l}), i=\{0,1\},$$

it is also straightforward to establish the detection function of the LRT given by:

$$\beta_{\overline{\delta}^{\text{lr}}} = 1 - \Phi\left(\frac{\sigma_0}{\sigma_1}\Phi^{-1}(1-\alpha_0) + \frac{m_0 - m_1}{\sigma_1}\right). \tag{4.12}$$

Equations (4.11) and (4.12) emphasize the main advantage of normalizing the LR as described in relation (4.9): it allows to set any of threshold that guarantees a false alarm probability independently from any distribution parameters and, this is particularly crucial because digital images are heterogeneous, their properties vary for each image. Second, the normalization allows to easily establish the detection power which again, is achieved, for any distribution parameters and hence, for any inspected image.

4.3.3 Application with Laplacian Distribution

In the case of Laplacian distribution, the framework of hypothesis testing theory has been applied for the steganalysis of JSteg in [135] in which the moments of LR are calculated under the two following assumptions: 1) all the DCT coefficients from the same subband are i. i. d. and 2) the expectation of each DCT coefficient is zero.

The continuous Laplacian distribution has the following pdf:

$$f_{\mu,b}(x) = \frac{1}{2b}\exp\left(-\frac{|x-\mu|}{b}\right) \tag{4.13}$$

where $\mu \in \mathbb{R}$, sometimes referred to as the location parameter, corresponds to the expectation, and $b > 0$ is the so-called scale parameter. During the compression of JPEG images, the DCT coefficients are quantized. Hence, let us define the discrete Laplacian distribution by the following pmf, see details in Appendix A.1:

$$
\begin{aligned}
f_{\mu,b}[k] &\overset{def.}{=} \mathbb{P}\Big[x \in [\Delta(k-\frac{1}{2}), \Delta(k+\frac{1}{2})[\Big] \\
&= \begin{cases} \exp\left(-\frac{|\Delta k-\mu|}{b}\right) \sinh\left(\frac{\Delta}{2b}\right) & \text{if } \frac{\mu}{\Delta} \notin [k-\frac{1}{2}; k+\frac{1}{2}[\\ 1-\exp\left(-\frac{\Delta}{2b}\right) \cosh\left(-\frac{\Delta(k)-\mu}{2b}\right) & \text{otherwise} \end{cases}
\end{aligned} \tag{4.14}
$$

where Δ is the quantization step.

From the expression of the discrete Laplacian distribution (4.14) and from the expression of LR (4.7), one can express the LR for the detection of JSteg under the assumption that DCT coefficients follow a Laplacian distribution, as follows (see Appendix A.2):

$$
\Lambda^{\text{lr}}_{\mu,b}[k] = \log\left(1 - \frac{R}{2} + \frac{R}{2}\exp\left[\frac{\Delta}{b}\,\text{sign}(\Delta k - \mu)(k - \bar{k})\right]\right), \tag{4.15}
$$

where the observed DCT coefficient, referred as $u_{k,l}$ in Eq. (4.7), is denoted k. It can be noted that this expression (4.15) of the LR is almost the same as the one obtained in [135] assuming that all DCT coefficients have a zero-mean, only the sign term $\text{sign}(\Delta k - \mu)$ becomes $\text{sign}(k)$ when assuming a zero-mean. It should also be noted that the log-LR equals 0 for every DCT coefficient whose value is 0 or 1 because the JSteg algorithm does not embed hidden data in those coefficients. In the present chapter, the moments of the LR (4.15) are not analytically established, the reader interested can refer to [135].

4.4 Proposed Approach for Estimating the Nuisance Parameters in Practice

4.4.1 Expectation Estimation of Each DCT Coefficient

As already explained, most of statistical models of DCT coefficients assume that within a subband the coefficients are i. i. d. However, as illustrated in Figure C.4b and 4.2 this assumption is doubtful in practice. Another way to explain why the DCT coefficients may not be i. i. d. is to consider a block of 8×8 pixels in the spatial domain, say the first, $\mathbf{z} = z_{i,j}, i \in \{1, \ldots, 8\}, j \in \{1, \ldots, 8\}$. The value of those pixels can be decomposed as:

$$
z_{i,j} = x_{i,j} + n_{i,j},
$$

where $x_{i,j}$ is a deterministic value that represents the expectation of pixel at location (i, j) and $n_{i,j}$ is the realization of a random variable representing all noises corrupting the inspected image. Clearly, this decomposition can be done for the whole block $\mathbf{z} = \mathbf{x} + \mathbf{n}$, where $\mathbf{x} = \{x_{i,j}\}$ and $\mathbf{n} = \{n_{i,j}\}$. Since the DCT operation is linear the DCT coefficient of any block may be expressed as :

$$
\begin{aligned}
\mathcal{DCT}(\mathbf{z}) &= \mathbf{D}^T\mathbf{z}\mathbf{D} = \mathbf{D}^T(\mathbf{x}+\mathbf{n})\mathbf{D} \\
&= \mathbf{D}^T\mathbf{x}\mathbf{D} + \mathbf{D}^T\mathbf{n}\mathbf{D} = \mathcal{DCT}(\mathbf{x}) + \mathcal{DCT}(\mathbf{n}),
\end{aligned} \tag{4.16}
$$

where $\mathcal{DCT}$ represents the DCT operation and $\mathbf{D}$ is the change of basis matrix from spatial to DCT basis, often referred to as the DCT matrix.

It makes sense to assume that the expectation of the noise component $\mathbf{n}$ has a zero-mean in the spatial and in DCT domain. On the opposite, it is difficult to justify that the DCT of pixels' expectation $\mathbf{x}$ should necessary be around zero. Actually, this assumption holds true if and only if the expectation is the same for of all the pixels from a block: $\forall i \in \{1,\ldots,8\}, \forall j \in \{1,\ldots,8\}, x_{i,j} = x$, see [136, 160, 161] for details.

In the chapter, it is mainly aimed at estimating the expectation of each DCT coefficient. To this end, it is proposed to decompress a JPEG image $\mathbf{V}$ into the spatial domain to obtain $\mathbf{Z}$, then to estimate the expectation of each pixel in the spatial domain $\widehat{\mathbf{Z}}$ by using a denoising filter. Then this denoised image $\widehat{\mathbf{Z}}$ is transformed back into the DCT domain to finally obtain the estimated value of all DCT coefficients, denoted $\widehat{\mathbf{V}} = \{\hat{v}_{i,j}\}, i \in \{1,\ldots,I\}, j \in \{1,\ldots,J\}$. Several methods have been tested to estimate the expectation of pixels in the spatial domain $\widehat{\mathbf{Z}}$, namely, the BM3D collaborative filtering [162], K-SVD sparse dictionary learning [163], non-local weighted averaging method from NL-means [164] and the wavelet denoising filter [42]. The codes used for the methods [162–164] have been downloaded from the Image Processing On-Line website[4]. The codes used for the method [42] have been downloaded from DDE [5].

4.4.2 A Local Estimation of b

In addition, the proposed model also assumes that the scale parameter $b_{k,l}$ is different for each DCT coefficient. The estimation of this parameter, for each DCT coefficient, is based on the WS JPEG method to locally estimate the variance; that is, for coefficients $v_{i,j}$, it simply consists of the sample variance of the DCT coefficients of the same subband from neighbouring blocks:

$$\widehat{\sigma}^2_{i,j} = \frac{1}{7} \sum_{\substack{s=-1 \\ (s,t)\neq(0,0)}}^{1} \sum_{t=-1}^{1} \left(v_{i+8s,j+8t} - \bar{v}_{i,j}\right)^2, \tag{4.17}$$

where $\bar{v}_{i,j}$ is the sample mean: $\frac{1}{8}\sum_{\substack{s=-1 \\ (s,t)\neq(0,0)}}^{1}\sum_{t=-1}^{1} v_{i+8s,j+8t}$. Let us recall that the Maximum Likelihood Estimation (MLE) of the scale parameter of Laplacian distribution from realizations $x_1,\ldots,x_N$ is given by $\hat{b} = N^{-1}\sum_{n=1}^{N}|x_n - \mu|$. The local estimation of the scale parameter it is proposed to use in this chapter is given by:

$$\hat{b}_{i,j} = \frac{1}{8} \sum_{\substack{s=-1 \\ (s,t)\neq(0,0)}}^{1} \sum_{t=-1}^{1} \left|v_{i+8s,j+8t} - \hat{v}_{i+8s,j+8t}\right|, \tag{4.18}$$

[4] Image Processing On-Line journal is available at: http://www.ipol.im

[5] Source codes are available at: http://dde.binghamton.edu

Table 4.1: Ratio (%) comparison before and after embedding.

	Inspected images index										
	No.1	No.2	No.3	No.4	No.5	No.6	No.7	No.8	No.9	No.10	Average
Cover Channel Selection Ratio	0.23	0.17	0.56	0.61	0.21	0.03	0.87	0.41	1.23	0.33	0.63
Stego Channel Selection Ratio	0.23	0.17	0.56	0.62	0.21	0.04	0.88	0.42	1.22	0.34	0.64
Cover DCT coefs. std	0.98	1.01	1.06	1.03	0.90	1.07	1.02	0.93	1.26	1.03	7.45
Stego DCT coefs. std	0.98	1.00	1.06	1.03	0.89	1.07	1.03	0.90	1.27	1.03	7.52
Cover JSteg Selection Ratio	1.12	0.81	2.46	2.49	0.80	0.08	5.07	2.42	7.56	0.34	1.44
Stego JSteg Selection Ratio	4.48	2.85	7.63	7.60	2.27	1.07	17.1	7.92	20.5	1.04	4.95
Cover and Stego Selection Similarity	89.5	91.5	94.2	93.0	80.7	80.7	93.9	93.7	93.3	93.9	92.8

where $\hat{v}_{i+8s,j+8t}$ is the estimation of expectation of each DCT coefficient by using denoising filter previously defined. As in the WS JPEG algorithm, this approach raises the problem of scale parameter estimation for blocks located on the sides of the image. In the present chapter, as in the WS JPEG method, it is proposed not to use those blocks in the test.

4.4.3 A Channel Selection to Improve the Method

Inspired by the channel selection algorithms (See [156, 157]), it is proposed to improve our detector with a Weighting Factor (WF). In practice, WF is generated from the quantized and rounded "residual noise", which is calculated by the following steps:

1. By uncompressing the JPEG format image, we obtain the intensity value of a JPEG image in the spatial domain.
2. By using a denoising filter, we extract the raw "residual noise" in the spatial domain.
3. By using DCT, we transform the raw "residual noise" from the spatial to the frequency domain.
4. By using quantization table, we can obtain the quantized "residual noise".
5. By rounding the quantized "residual noise" in the frequency domain, the quantized and rounded "residual noise" is obtained.
6. If a quantized and rounded "residual noise" takes zero, WF equals 0; If not, WF equals 1.

Thus, based on our proposed WF, it is proposed to categorize "residual noise" set into two sub-sets: "non-zero" sub-set and "zero" sub-set. To verify the effectiveness of our improved algorithm, it is proposed to randomly choose ten exemplary images

which are compressed to JPEG format images with quality factor 70, embedding rate $R = 0.05$. Also, all the images of BossBase database [155] are used for computing the average value. Table 4.1 gives the statistical ratio of the data in which the annotations of the table are as followed:

- **Cover Channel Selection Ratio**: denotes the ratio of "non-zero" sub-set to "residual noise" set of a cover image.
- **Stego Channel Selection Ratio**: denotes the ratio of "non-zero" sub-set to "residual noise" set of a stego image.
- **Cover DCT coefs. std**: denotes the standard deviation of "residual noise" set from a cover image.
- **Stego DCT coefs. std**: denotes the standard deviation of "residual noise" set from a stego image.
- **Cover JSteg Selection Ratio**: denotes the ratio of the DCT coefficients used by JSteg in "non-zero" sub-set to the DCT coefficients used by JSteg in "residual noise" set from a cover image.
- **Stego JSteg Selection Ratio**: denotes the ratio of the DCT coefficients used by JSteg in "non-zero" sub-set to the DCT coefficients used by JSteg in "residual noise" set from a stego image.
- **Cover and Stego Selection Similarity**: denotes the ratio of the same position in "non-zero" sub-set before and after embedding.

In our proposed statistical test, the number of the selected coefficients for the detection should be kept very close before and after embedding. As Table 4.1 illustrated, the ratio of **Cover Channel Selection Ratio** and **Stego Channel Selection Ratio** basically remains the same before and after embedding, which reveals the proportion of the coefficients used for the test nearly the same. Similarly, the ratio of **Cover DCT coefs. std** and **Stego DCT coefs. std** allows us to verify our assumption that the embedding doesn't change much the statistical properties of the "residual noise". In addition, those numbers also show that, after rejection of the content, the "residual noise" standard deviation is very small compared to original DCT coefficients (see also Figures 4.2 and 4.3), which thus permits a better detection of modifications due to JSteg embedding. The ratio of **Cover and Stego Selection Similarity** which is kept at the high value signifies most of "residual noise" are chosen at the same position. Then the only difference is the comparison between **Cover JSteg Selection Ratio** and **Stego JSteg Selection Ratio**. It should be noted that if all DCT coefficients used by JSteg are included in "non-zero" sub-set, then the ratio equals 100%. It is observed that only a few of DCT coefficients used by JSteg algorithm is included in "non-zero" sub-set. Nevertheless, after embedding, the ratio of **Stego JSteg Selection Ratio** is largely

improved, compared with the ratio of **Cover JSteg Selection Ratio**. It can be assumed that by using a WF, more "residual noise" from the embedding positions are counted. Besides, prior to embedding secret information, we never know which position will be embedded, the very low ratio of **Cover JSteg Selection Ratio** is reasonable.

By investigating the "non-zero" and "zero" sub-set, although we can not capture all the embedding positions in the DCT domain, it is totally enough to detect the JSteg steganography. Besides, all the coefficients in "zero" sub-set are not counted in our proposed test. On average, for a cover image with the size of 512×512, 0.63% of the coefficients are kept to compute the test; 0.64% of the coefficients from a stego image are used. As the embedding rate $R = 0.05$, it is obvious that most of DCT coefficients remain the same before and after embedding. Thus, it is not necessary to compute these values. Furthermore, the LR values of these DCT coefficients without embedding any information probably mask or disturb LR from DCT coefficients with JSteg embedding.

4.4.4 Design of Proposed Test

In Section 4.3 the framework of hypothesis testing theory has been presented assuming that distribution parameters are known for each DCT coefficient. To design a practical test, a usual solution consists in replacing the unknown parameter by its ML estimation. This leads to the construction of a Generalized LRT. A similar construction is adopted in this chapter, using the *ad hoc* estimators presented at the beginning of section 4.4, instead of using the ML method to estimate the distribution parameters of each DCT coefficient. The proposed test is thus defined as:

$$\widehat{\delta}(\mathbf{U}) = \begin{cases} \mathcal{H}_0 \text{ if } \widehat{\Lambda}(\mathbf{U}) = \displaystyle\sum_{k=1}^{K}\sum_{l=2}^{64} \widehat{\Lambda}_{cs}(u_{k,l}) < \widehat{\tau}, \\ \mathcal{H}_1 \text{ if } \widehat{\Lambda}(\mathbf{U}) = \displaystyle\sum_{k=1}^{K}\sum_{l=2}^{64} \widehat{\Lambda}_{cs}(u_{k,l}) \geq \widehat{\tau}, \end{cases} \tag{4.19}$$

where the channel selection decision statistic $\widehat{\Lambda}_{cs}(u_{k,l}) = \widehat{\Lambda}(u_{k,l}) \cdot w_{k,l}$ for a single DCT coefficient is given and a weighting factor $w_{k,l}$ selects the DCT channel. Next, let us study the $\widehat{\Lambda}(u_{k,l})$ to verify the effectiveness of our proposed test, where $\widehat{\Lambda}(u_{k,l}) \in \{\widehat{\Lambda_0}(u_{k,l}), \widehat{\Lambda_1}(u_{k,l}), \widehat{\Lambda_2}(u_{k,l}), \widehat{\Lambda_3}(u_{k,l})\}$.

To verify our improvement based on the Laplacian test, see [135], it is proposed to consider the weighing factor $w_{k,l}$ as a constant equal to 1. The scale parameter $\widehat{b}$ is estimated by using MLE and the location parameter is ignored (see details in [135]). The LR is given by:

$$\widehat{\Lambda_0}(u_{k,l}) = \log\left(1 - \frac{R}{2} + \frac{R}{2}\exp\left[\frac{\Delta}{\widehat{b}}\,\mathrm{sign}(\Delta k)(k - \bar{k})\right]\right). \tag{4.20}$$

The first improvement of the previous LR is the consideration of the location pa-

rameter $\widehat{\mu}_{k,l}$ (see subsection 4.4.1). The new LR is designed by:

$$\widehat{\Lambda_1}(u_{k,l}) = \log\left(1 - \frac{R}{2} + \frac{R}{2}\exp\left[\frac{\Delta}{\widehat{b}}\,\mathrm{sign}(\Delta k - \widehat{\mu}_{k,l})(k - \bar{k})\right]\right). \tag{4.21}$$

The second improvement is the estimation of the scale parameter $\widehat{b}_{k,l}$ (see subsection 4.4.2) and ignore the location parameter. The LR is designed by:

$$\widehat{\Lambda_2}(u_{k,l}) = \log\left(1 - \frac{R}{2} + \frac{R}{2}\exp\left[\frac{\Delta}{\widehat{b}_{k,l}}\,\mathrm{sign}(\Delta k)(k - \bar{k})\right]\right). \tag{4.22}$$

The third improvement is to give the assumption that DCT coefficients are i. i. d. The scale parameter $\widehat{b}_{k,l}$ and the location parameter $\widehat{\mu}_{k,l}$ of the distribution are estimated separately by using our proposed algorithms of subsection 4.4.1 and 4.4.2.

$$\widehat{\Lambda_3}(u_{k,l}) = \log\left(1 - \frac{R}{2} + \frac{R}{2}\exp\left[\frac{\Delta}{\widehat{b}_{k,l}}\,\mathrm{sign}(\Delta k - \widehat{\mu}_{k,l})(k - \bar{k})\right]\right). \tag{4.23}$$

Moreover, it is proposed to explore the effectiveness of introducing a weighting factor $w_{k,l}$ which is defined as:

$$w_{k,l} = \begin{cases} 0 & \text{if} \quad \Delta k - \widehat{\mu}_{k,l} \in (-0.5, 0.5) \\ 1 & \text{otherwise.} \end{cases} \tag{4.24}$$

The last LR is obtained by multiplying (4.23) by $w_{k,l}$:

$$\widehat{\Lambda}_{cs}(u_{k,l}) = \widehat{\Lambda_3}(u_{k,l}) w_{k,l}. \tag{4.25}$$

It should be noted that (4.20) is the algorithm from [135]. In Section 4.5, the specific comparison of the detectors is presented. In order to have a normalized decision statistic for the whole image, $\widehat{\Lambda}(\mathbf{U})$ is defined as:

$$\begin{aligned} \widehat{\Lambda}(\mathbf{U}) &= \frac{1}{S_L}\sum_{k=1}^{K}\sum_{l=2}^{64}\widehat{\Lambda}_{cs}(u_{k,l}) - E_{\mathcal{H}_0}(\widehat{\mu}_{k,l}, \widehat{b}_{k,l}) \\ \text{with} \quad S_L^2 &= \sum_{k=1}^{K}\sum_{l=2}^{64} V_{\mathcal{H}_0}(\widehat{\mu}_{k,l}, \widehat{b}_{k,l}). \end{aligned} \tag{4.26}$$

4.4.5 Comparison with Prior-art

The WS JPEG algorithm, as well as the WS for the spatial domain, is based on the underlying assumption that the observations follow a Gaussian distribution. As recently shown [130, 132], the WS implicitly assumes that the quantization step is negligible. Let us rewrite the LR test for JSteg detection based on a Gaussian distribution model of DCT coefficients. Let X be a random variable following a

quantized Gaussian distribution. Exploiting the assumption that the quantization step is negligible compared to noise standard deviation allows the writing of:

$$\begin{aligned}\mathbb{P}[X=k] &= \int_{\Delta(k-1/2)}^{\Delta(k+1/2)} \frac{1}{\sigma\sqrt{2\pi}}\exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)dx \\ &\approx \frac{\Delta}{\sigma\sqrt{2\pi}}\exp\left(-\frac{(\Delta k-\mu)^2}{2\sigma^2}\right). \end{aligned} \tag{4.27}$$

Putting this expression of the pmf under hypothesis $\mathcal{H}_0$ into the LR (4.2), and assuming that the quantization step is negligible compared to the noise standard deviation, $\Delta << \sigma$, it is immediate to obtain the following expression of the LR under the assumption of Gaussian distribution of DCT coefficient

$$\begin{aligned}&\log\left(1-\frac{R}{2}+\frac{R}{2}\frac{\exp\left(-\frac{(\Delta\bar{k}-\mu)^2}{2\sigma^2}\right)}{\exp\left(-\frac{(\Delta k-\mu)^2}{2\sigma^2}\right)}\right) \\ \approx &\frac{R\Delta}{2\sigma^2} \quad (k-\bar{k}) \quad (\Delta k-\mu) \\ = &\overbrace{w_\sigma} \quad \overbrace{\pm 1} \quad \overbrace{(\Delta k-\mu)}\end{aligned} \tag{4.28}$$

see details in Appendix A.3. This expression highlights the well known fact the WS consists in fact of three terms: 1) the term w_σ which is a weight so that pixels or DCT coefficients with highest variance have a smallest importance, 2) the term $(k-\bar{k}) = \pm 1$ according to the LSB of k and 3) the term $(\Delta k - \mu)$.

In comparison, the expression of the LR for a Laplacian distribution model (4.15), as well as the expression of the proposed test with estimates (4.21) can be approximated by (see details in Appendix A.2):

$$\begin{aligned}&\frac{R\Delta}{2b} \quad (k-\bar{k}) \quad \text{sign}(\Delta k-\mu) \\ = &\overbrace{w_b} \quad \overbrace{\pm 1} \quad \overbrace{\text{sign}(\Delta k-\mu)}\end{aligned} \tag{4.29}$$

which is also made of three terms; the two first are roughly similar to the two first terms of the WS : 1) the term w_b is a weight so that DCT coefficients with highest "scale" b have a smallest importance, note that the variance is proportional to b^2, 2) the term $(k-\bar{k}) = \pm 1$ according to the LSB of k. However, in the expression of the LR based on the Laplacian model the term $(\Delta k - \mu)$ of the WS is replaced with its sign. This shows that the statistical tests based on Laplacian model and based on Gaussian model are essentially similar.

4.5 Numerical simulations

4.5.1 Results on Simulated Images

One of the main contributions of this chapter is to show that the hypothesis testing theory can be applied in practice to design a statistical test with known statistical properties for JSteg steganalysis.

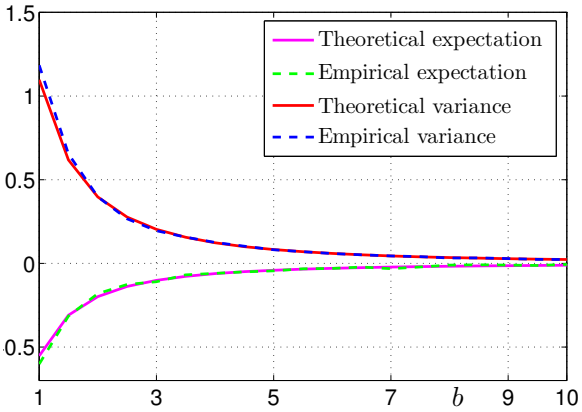


Figure 4.5: Expectation m_0 and variance σ_0^2 as a function of the scale parameter b theoretically and empirically.

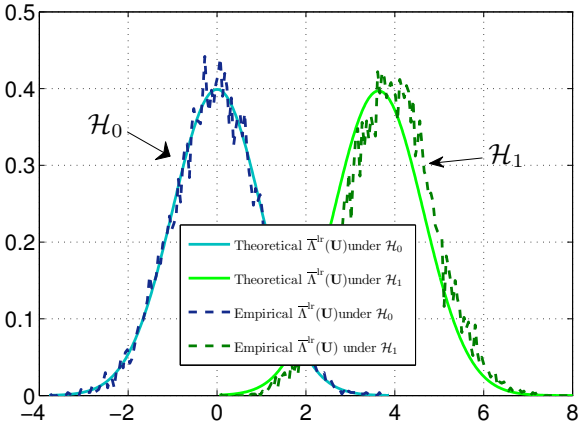


Figure 4.6: Comparison between empirical and theoretical distribution of $\overline{\Lambda}^{\mathrm{lr}}(\mathbf{U})$.

To verify the sharpness of the theoretically established results, we generate 1000 sets of 4000 random variable (a Monte-Carlo simulation) following the Laplacian distribution, where $R = 0.05$, $\mu = 0$ and b distributed from 1 to 10 with a step of 0.5. Then, the expectation and variance values are calculated empirically and theoretically. As shown in Figure 4.5, the empirically calculated moments are almost equal to the analytically established ones.

Subsequently, to verify the effectiveness of the established LRT $\overline{\delta}^{\mathrm{lr}}(\mathbf{U})$, again, a Monte-Carlo simulation is performed by repeating 10000 times using a vector 64×4096 following the Laplacian distribution, in which the scale parameter is selected arbitrarily as 3 and the location parameter 0. Under the hypothesis $\mathcal{H}_0$ and $\mathcal{H}_1$ respectively, Figure 4.6 presents the comparison between empirical and theoretical distribution of $\overline{\Lambda}^{\mathrm{lr}}(\mathbf{U})$. The results highlight the validity of the proposed test (4.10).

Figure 4.7 gives the comparison between the empirical and theoretical FAR α_0

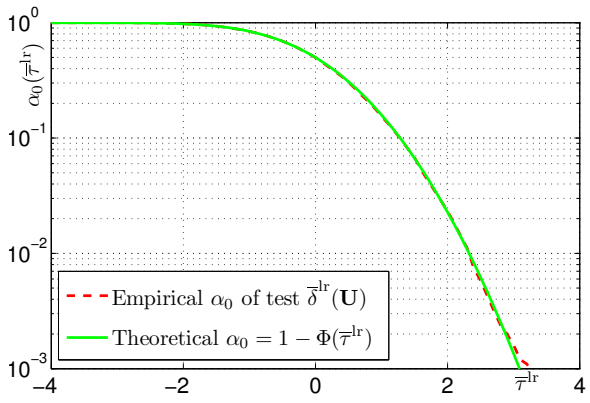


Figure 4.7: FAR α_0 as a function of the threshold $\overline{\tau}^{\mathrm{lr}}$.

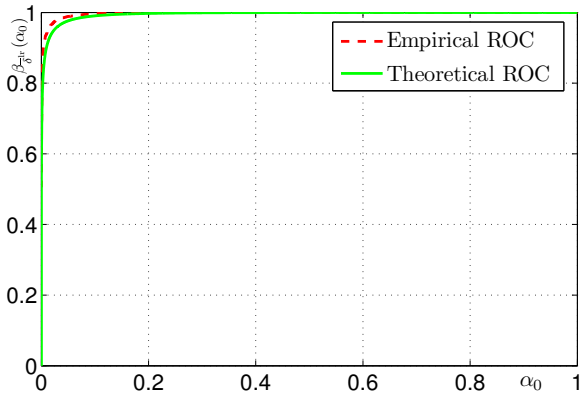


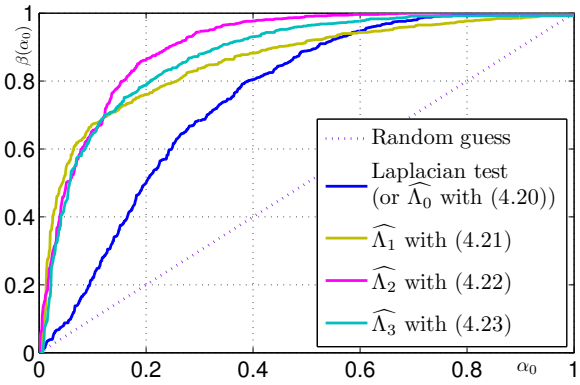
Figure 4.8: Detection power $\beta_{\overline{\delta}^{\mathrm{lr}}}$ as a function of FAR α_0 (ROC curve).

respectively of the test (4.10). This particularly demonstrates that two curves are very close. Figure 4.8 offers the Receiver Operating Characteristic (ROC) comparison, that is the detection power $\beta_{\overline{\delta}^{\mathrm{lr}}}$ as a function of FAR α_0, of both empirical and theoretical established results in (4.11) and (4.12).

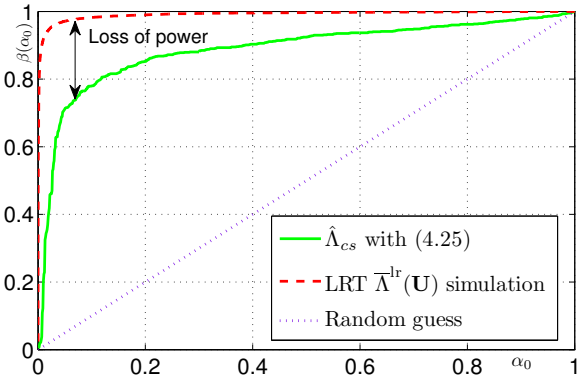
4.5.2 Results on Real Images

Another contribution of this chapter is to design the optimal test with estimated parameters to break JSteg algorithm in practical case.

First, let us investigate our proposed detectors (4.21)-(4.23). It is proposed to perform a numerical simulation over the 1000 images from BossBase [155] which have been compressed in JPEG with quality factor 70. The payload, or embedding rate, R is set at 0.05 for JSteg algorithm. For fairly comparison with Laplacian test from [135], it first shows the improvement provided by the proposed model with $w_{k,l} = 1$. As Figure 4.9a illustrates, all the proposed detectors outperform $\widehat{\Lambda_0}(u_{k,l})$ (4.20) (Laplacian test) proposed by [135]. Moreover, in the following



(a) Comparison of detection performance without channel selection (constant weighting factor $w_{k,l} = 1$).



(b) Comparison of detection performance with channel selection (weighting factor $w_{k,l}$ from (4.24)).

Figure 4.9: ROC curves comparison, detection power as a function of FAR α_0.

investigation, it is proposed to use $\widehat{\Lambda}_{cs}(u_{k,l})$. Then, it is proposed to give the performance of this detector on 1000 simulated images in which a DCT subband is generated by strictly following the Laplacian distribution (see Figure 4.9b). Then a comparison with simulations of the LR test shows the loss of power due to the estimation of expectation and scale parameters. It should be noted that in all our proposed detectors in this chapter, $\widehat{\Lambda}_3(u_{k,l})$ (4.23) with $w_{k,l}$ (4.24) performs best. Thus, it is proposed to use it as our optimal steganalyzer for competing with the state-of-the-art JSteg detectors. It should be emphasized that in Figure 4.9, the wavelet denoising filter [42] is used for estimating the location parameter $\widehat{\mu}_{k,l}$ (see subsection 4.4.1).

To verify the relevance of the proposed methodology, it is proposed to compare the proposed statistical test with two other detectors. The first chosen competitor is the statistical test proposed in [135] as it is also based on Laplacian model but does not take into account the distribution parameters as nuisance parameters; it considers that DCT coefficients are i. i. d. , following a Laplacian distribution with

zero-mean. The comparison with this test is meaningful as it allows us to measure how much the detection performance is improved by removing the assumption that the DCT coefficients of each subband are i. i. d. The second chosen competitor is the WS detector [123] due to its similarity with the proposed statistical test, see details in subsection 4.4.5.

For a large scale verification, it is proposed to use the BOSS database, made of 10 000 grayscale images of size 512×512 pixels, used with payload $R = 0.05$. Prior to our experiments, the images have been compressed in JPEG using the linux command convert which uses the standard quantization table. Note also that all the JSteg steganography was performed using a Matlab source code we developed based on Phil Sallee's JPEG Toolbox[6]. Four denoising methods have been tested to estimate the expectation of each DCT coefficient, namely the K-SVD, the BM3D, the NL-means and the wavelet denoising algorithms.

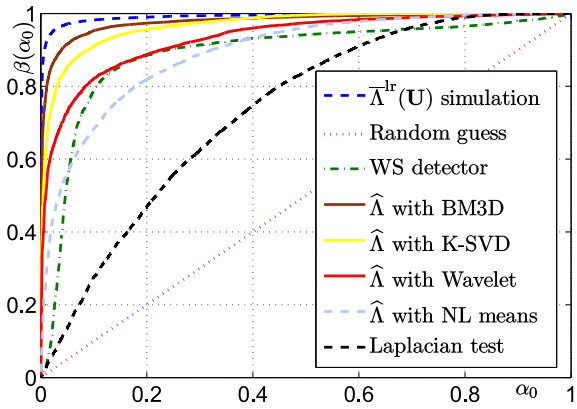
Figure C.5 shows the detection performances obtained over the BOSS database compressed with quality factor 70, in which our proposed test $\widehat{\Lambda}$ (4.26) compares with prior-arts. The detection performances are shown as ROC curves, that is the detection power is plotted as a function of false alarm probability. Figure C.5a particularly emphasizes that the Laplacian test [135] does not perform well while the proposed methodology which takes into account the Laplacian distribution parameters as nuisance parameters allows us to largely improve the performance. Similarly the WS detector achieves overall good detection performance. However, it can be shown on Figure C.5b, which presents the same results using a logarithmic scale, that for low false alarm probabilities, the performance of the WS significantly decreases. On the opposite, the proposed statistical test still performs well.

Among the four denoising algorithms that have been tested, the BM3D achieves the best performance but it can be observed on Figure C.5 that the performance obtained using the K-SVD and using the wavelet denoising methods are also very good. The performance of NL means method is comparable with WS detector [123].

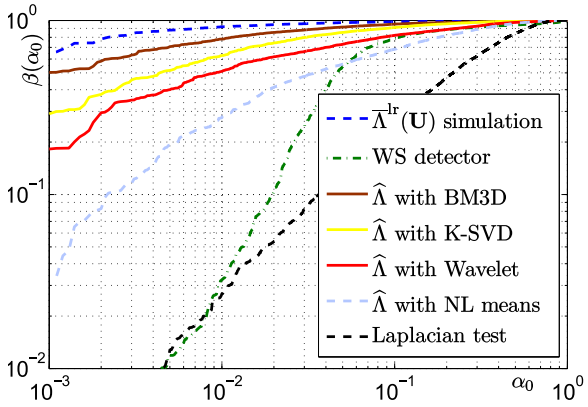
To extend the results previously presented, a similar test has been performed over the BOSS database using the quality factor 85. The detection performance obtained by the proposed test $\widehat{\Lambda}$ (4.26) and by the competitors are presented in Figure 4.11. Again, this figure shows that based on the Laplacian model, the Laplacian test assuming that DCT coefficients of a subband are i. i. d. has an unsatisfactory performance. It can also be noted that even though the WS performs slightly better for low false alarm probability, compared with the results obtained with quality factor 70, it performs much worse than our proposed statistical test.

For a comparison with current steganalysis, Figure 4.12 shows the performance obtained with "universal detectors" using ROC curves. For the same steganalysis problem, let us detect JSteg with payload R=0.05 on BOSS database compressed with quality factor 70. In order to be able to draw a ROC curve, the ensemble classifier [141] is cast within the hypothesis testing framework as described in [142,

[6]Phil Sallee's JPEG Toolbox is available at : http://dde.binghamton.edu/download/jpeg_toolbox.zip



(a) Comparison of detection performance for BOSS database with quality factor 70 (linear scale).



(b) Comparison of detection performance for BOSS database with quality factor 70 (logarithmic scale).

Figure 4.10: Comparison of detection performance for BOSS database with quality factor 70.

165] and the low complexity linear classifier described in [143].

While the results obtained with Spatial Rich Model with unitary quantization step (SRMQ1) [137] are not presented because the classification is perfect on all the testing sets that we have used (10 split of 5000 images have been tested), Figure 4.12 only presents the results obtained with JPEG Rich Model (ccJRM) [166]. Then let us compare the results obtained with those prior-art feature sets and the prior-art classifier with those obtained by the proposed method using the BM3D denoising algorithm (that gives the best empirical performances). It is important to recall that the proposed method is based on hypothesis testing and requires extremely accurate statistical model of DCT coefficients while, on the opposite, the efficient machine learning method tend to extract a tremendous set of features in order to comprehensively describe the statistical properties of images.

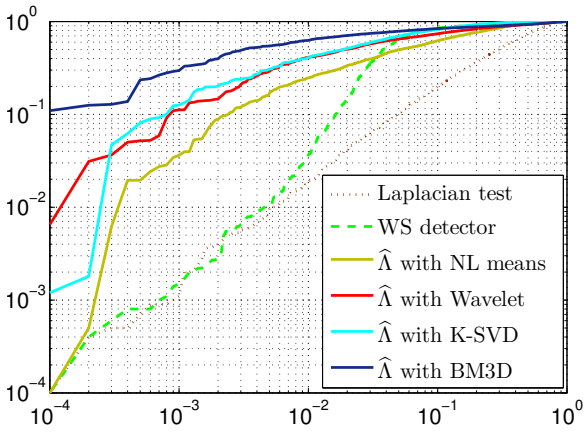


Figure 4.11: Comparison of detection performance for BOSS database with quality factor 85 (logarithmic scale).

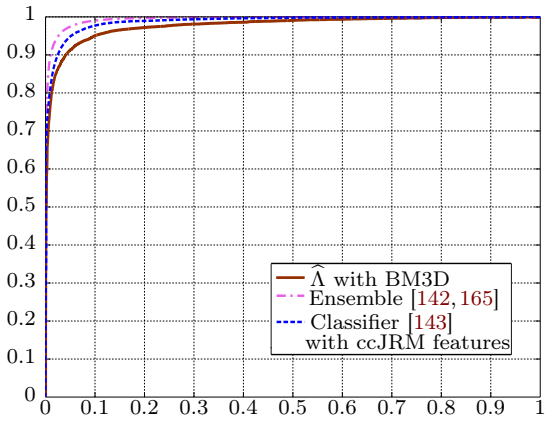


Figure 4.12: Illustration of detection performance from "universal detectors".

It is also more fair to acknowledge that the proposed method performs fairly well compared to those state of the art method only on non-adaptive scheme. It is very likely that on state of the art adaptive steganographic scheme, such as for instance those proposed in [128, 167, 168], for which the state of the art supervised learning have been developed, the gap in terms of performance would be much more important.

4.6 Conclusion

This chapter aims at improving the optimal detection of data hidden within the DCT coefficients of JPEG images. Its main originality is that the usual Laplacian model is used as a statistical model of DCT coefficients but, opposed to what is usually proposed, it is not assumed that all DCT coefficients from a subband are i. i. d. This leads us to consider the Laplacian distribution parameters, namely the

expectation e and the scale parameter b, as nuisance parameters as they have no interest for the detection of hidden data, but must be carefully taken into account to design an efficient statistical test. Numerical results show that by estimating those nuisance parameters, the Laplacian model allows the designing of an accurate statistical test which outperforms the WS detector. Besides, the comparison with the optimal detector based on the Laplacian model and on the assumption that all DCT coefficients of a subband are i. i. d. shows the relevance of the proposed approach.

A possible future work will apply this approach with state-of-the-art statistical model of DCT coefficients, such as the Generalized Gaussian or the Generalized Gamma model. This could provide improvements in the detection performance at the cost of a higher complexity.

Chapter 5

Individual Camera Device Identification Based on RAW Images

Contents

5.1 Introduction and Contributions

Most digital image forensic methods are designed based on the extracted fingerprints from the stages of image acquisition. Hence, let us first generally illustrate the image acquisition pipeline in a digital camera (see Figure 5.1). Photons radiating from a scene pass through the optical system (e.g. lens). Then, Color Filter Array (CFA) filters the incident light spectrum. Subsequently, the image sensor converts it to electrical energy which is then converted to digital signal by an analog-to-digital (A/D) converter inside the camera. In the capture of a RAW image, it

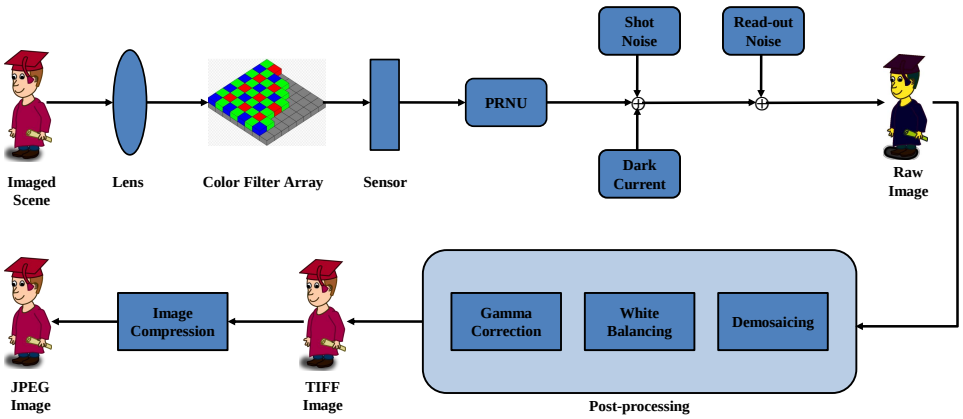


Figure 5.1: Illustration of image acquisition pipeline in a digital camera.

is unavoidable that additive noise (e.g. shot noise, read-out noise, dark current) and multiplicative noise (Photo-Response Non-Uniformity noise, often referred to as PRNU) corrupt the acquired image. Then, after several post-processing stages (e.g. demosaicing, white balancing, gamma correction), an uncompressed image is obtained; this uncompressed high-quality image is referred to as a TIFF (Tagged Image File Format) image in the present chapter. Finally, for storage, an image compression algorithm is applied very often using the Joint Photographic Experts Group (JPEG) standard and, hence, an image in JPEG format is output. An overview of the structure and processing stages in a digital still camera is detailed in [2, 41]. In this chapter, we mainly investigate image origin identification from RAW images.

Source camera or image origin identification, which relies on the camera fingerprints left in the digital images, can generally be classified into the two following categories:

1. The methodologies in the first category rely on the image acquisition pipeline. For instance, using the white balancing as a camera fingerprint, the algorithm proposed in [169] identified the device origin of a given image. By exploiting the fingerprints in the early-acquisition stage such as lens aberration, the forensic detector [16] was proposed to identify the source camera model. This lens aberration-based detector, however, has to face the challenge that interchangeable-lens cameras could change their lens at any time. The methods proposed by [14, 15, 51, 52] utilized CFA and demosaicing algorithms to identify the source camera model. Besides, the intrinsic features extracted by JPEG compression were also considered as influential factors to identify the camera origin, see for instance [57]. These algorithms are cast within the framework of supervised classification (usually based on the popular Support Vector Machine, SVM). The challenging problems are that a few manufacturers share the similar image processing technique which leads to a

high similarity of intrinsic fingerprints, especially in the case of same camera manufacturer and model. Moreover, the supervised classification framework, however, unavoidably involves two main drawbacks which somewhat limit its application. During its training stage, the scheme first requires a large scale of different types of samples (e.g. images) from various sources, which is probably hardly possible in practice. Besides, the statistical performance of the trained classifier is only known empirically from a testing set, and cannot be established, which remains an open problem for all machine learning methods [27].

2. The methods in the second category aims to exploit the unique noise of each camera model and device as an "intrinsic" fingerprint for source camera identification. Due to imperfections during sensor manufacturing process and non-uniformity of photo-electronic conversion caused by inhomogeneity of silicon wafers (see [40, 41]), it is unavoidable that Sensor Pattern Noise (SPN) is generated. SPN extracted from a given image was first used for identifying source camera device [42]. Inspired by this pioneer work, [45, 47–49, 170] improved the prior algorithm. It should be noted that the SPN includes two main components: Fixed Pattern Noise (FPN) and PRNU. The FPN was represented by dark current, which was exploited in [43], can be suppressed typically by subtracting a dark frame from the output image. Thus, the FPN fingerprint is not considered as enough robust and, thus, not used in later works any more. Compared with the FPN, the PRNU is much more robust. [47, 48, 170, 171] directly investigated the algorithms of source camera identification based on the PRNU. However, the counter-forensic algorithm [103] challenged the credibility of the SPN-based fingerprint in a digital camera. Therefore, the main challenge in this category is that the extracted PRNU is not independent of the image content, which can be severely contaminated by the details from scenes. Besides, detectors proposed in the existing literature limitedly investigate the hypothesis theory and statistical image models. Thus, the performance of source camera identification still remains analytically unestablished and is empirically measured on specific databases.

To the best of our knowledge, the proposed tests of [42, 47, 48, 169] could implement camera device identification, which means that the detectors use extracted fingerprints to distinguish the different devices, possibly of the same camera model. In this chapter, we mainly study the source camera device identification, which requires more accurate and robust characteristics of the fingerprints.

The methodology proposed in the present chapter lies in the second category; it exploits the intrinsic fingerprint of digital image noise statistical properties to identify the individual camera device, or instance, from raw images. The proposed methodology is based on hypothesis testing theory and also aims at establishing the statistical properties of the detector.

Let us define an inspected image $\mathbf{Z}$, and two camera sources $\mathcal{S}_0$ and $\mathcal{S}_1$, referred to as individual devices, are required to be identified. Then the problem of source

camera identification can be cast within the following framework. In this scenario, a forensic detector makes a choice between the following bi-criteria hypotheses:

$$\begin{cases} \mathcal{H}_0 : \mathbf{Z} & \text{taken by the camera source } \mathcal{S}_0 \\ \mathcal{H}_1 : \mathbf{Z} & \text{taken by the camera source } \mathcal{S}_1 \text{ different from } \mathcal{S}_0. \end{cases}$$

In a practical context, it is proposed to use the additive noise (e.g. shot noise and read-out noise) as the unique fingerprints for source camera device identification. In our prior research, based on the Poissonian-Gaussian noise model, the statistical test proposed in [22] was designed for identifying source camera model. The fingerprints used in [22], however, hardly discriminated the different devices of the same camera model. By using the enhanced Poissonian-Gaussian noise model proposed in this chapter, we design a new statistical detector for identifying source camera device from RAW format images. Hence, the main contributions are the followings:

- The Poissonian-Gaussian model of [172–174] was proposed to describe the pixel distribution in a natural raw image. However, by challenging the assumption that among a single image all the pixels, regardless of their values, share the same camera parameters, the present chapter proposes to explicitly describe the statistical properties of pixels as a function of the incident among light, or their level, which leads to our design of the enhanced Poissonian-Gaussian noise model.

- In an ideal context where all model parameters are perfectly known, the Likelihood Ratio Test (LRT) is presented and its performance is theoretically established. The statistical performance of LRT serves as an upper bound on the detection power for the camera device identification problem.

- In a practical context, it is proposed to design a Generalized Likelihood Ratio Test (GLRT). Based on the enhanced Poissonian-Gaussian model, the GLRT is designed to identify, among two different devices, the one with which the inspected image has been captured.

- Numerical simulations show the sharpness of the theoretically established results and the good performance of our proposed tests. Comparisons with prior-art methods proposed in [22, 48], experimental results also show the practical efficiency of our proposed detector.

This chapter is organized as follows. Section 5.2 reviews the Poissonian-Gaussian noise model, and proposes its enhanced version for a RAW image. In Section 5.3, based on the enhanced Poissonian-Gaussian noise model, an optimal LRT is presented. This test assumes that the statistical properties of each pixel is known. In this ideal setting the statistical performance of this LRT is established and serves as an upper bound on camera device identification. Then, Section 5.4 presents a practical GLRT which estimates the expectation of each pixel, that roughly speaking represents the image content. Then, Section 5.5 presents numerical results of the

proposed detectors on the simulated and real images. Finally, Section 5.6 concludes this chapter.

5.2 Proposed Camera Fingerprints for RAW Images

5.2.1 Overview on Poissonian-Gaussian Noise Model and Its Limitation

Let us assume that a natural RAW image is a vector $\mathbf{Z} = \{z_i\}$ of I pixels where $i \in \{1, \cdots, I\}$. The photo-electron conversion essentially consists in a counting process modelled as a Poisson process. Then, the number of the collected electrons, denoted as Ne_i, is the sum of the electrons generated by the incident photons Np_i and the dark electrons Nt_i generated by thermal noise. It follows that Ne_i is defined by:

$$Ne_i \sim \mathcal{P}(\eta_i Np_i + Nt_i), \tag{5.1}$$

where $\mathcal{P}(\cdot)$ denotes the Poisson distribution and η_i represents a conversion factor that accounts for filter transmittance and quantum efficiency. For clarity, let us assume that no PRNU or FPN are generated in this stage. Therefore, the photo-sensitivity and thermal noise are constant for each pixel, the index i is therefore omitted from η and Nt. During the read-out process, the recorded signal is corrupted by different sources of electronic noise modelled as a zero-mean Gaussian random variable ϵ_i with variance ω^2. Therefore, z_i is given by:

$$z_i = g \cdot (Ne_i + \epsilon_i), \tag{5.2}$$

where g represents the analog gain controlled by International Standardization Organization (ISO) sensitivity setting. It should be noted that RAW pixels are statistically independent (see details in [172,174]). Since the number of counted electrons is large enough, the Gaussian approximation of the Poisson distribution can be applied, and is represented by the following definition:

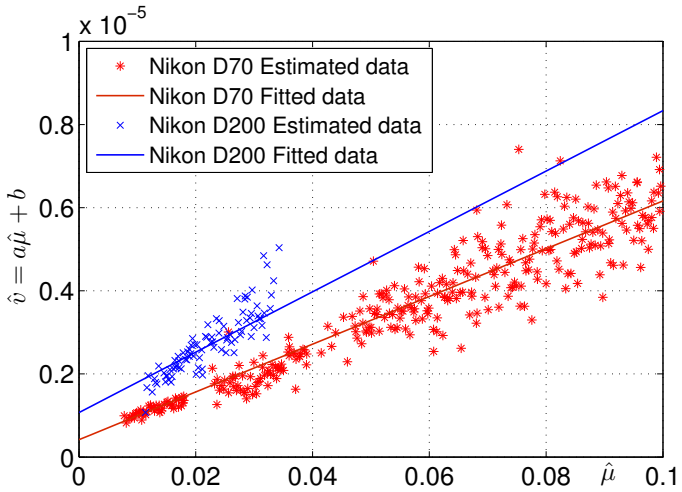
$$z_i \sim \mathcal{N}(\mu_i, a\mu_i + b), \tag{5.3}$$

where $\mathcal{N}(\cdot)$ represents the Gaussian distribution with the pixel expectation $\mu_i = g(\eta Np_i + Nt)$ and variance $v_i = a\mu_i + b$. The parameters (a, b) are respectively represented by:

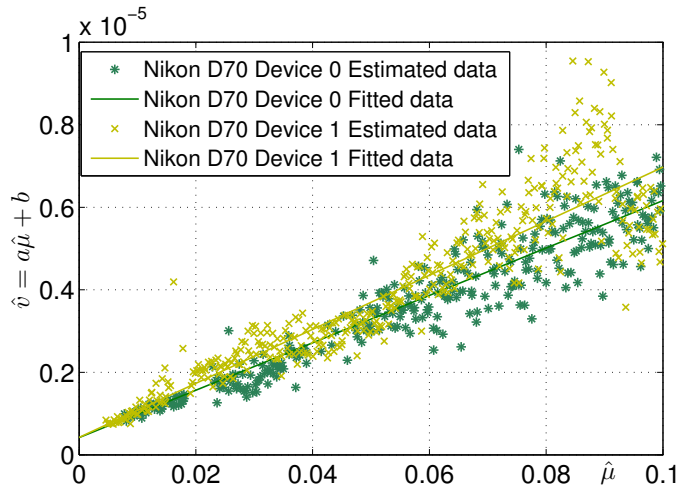
$$a = g \quad \text{and} \quad b = g^2\omega^2. \tag{5.4}$$

In practice, it should be noted that in some digital imaging sensors, since the collected electrons N_{e_i} is compensated by a base pedestal parameter p_0 to establish an offset-from-zero of the output pixel, $b = g^2(\omega^2 - p_0)$ (see details in [22, 172]). Therefore, in the case of $p_0 > \omega^2$, $b < 0$. In [22], it was proposed to define (a, b) as the camera parameters and the pixel expected value μ_i as the image parameter.

The camera parameters (a, b) are mainly controlled by the ISO sensitivity g. Besides, it should be noted that other camera settings such as shutter speed, focal



(a) Results from RAW images of *Nikon D200* and *Nikon D70*.



(b) Results from RAW images of *Nikon D70* device 0 and *Nikon D70* device 1.

Figure 5.2: Illustration of Scatter-plot of pixels' expectation $\hat{\mu}$ and variance $\hat{v}$.

length, and integration time can also slightly impact the parameters (a, b). Nevertheless, compared with the effect of ISO sensitivity, those effects are negligible and hence ignored in this chapter for clarity. Hence, for a RAW format image, we only emphasize the effect of ISO sensitivity. In other words, all the identification should be investigated with the constant ISO value while other camera settings could change. The parameter a is proportional to ISO sensitivity while the parameter b is proportional to its square. Moreover, based on the assumption that ISO sensitivity g is pre-set before acquiring an image, and variance ω^2 is nearly stable in the inspected RAW format image, a linear relationship between pixels' expectation and variance referring to Equation (5.3) can be illustrated in Figure 5.2a.

However, the camera parameters (a, b) are probably not constant for all the pixels within a RAW format image due to the spatial variation in the pixel response, e.g. PRNU. Besides, based on this noise model (5.3), it is hardly possible to implement source camera device identification. Figure 5.2b illustrates the linear relationship from the same camera model but different devices. Most overlapped parts of the estimated data predict that the constant camera parameters (a, b) from two devices of *Nikon D70/D200* can not be discriminated (see Figure 5.3a). Hence, it is proposed to improve this Poissonian-Gaussian noise model in the following subsection by taking into account pixels' response non-uniformity.

5.2.2 Enhanced Poissonian-Gaussian Noise Model

By challenging the assumption that among a single image all the pixels, regardless of their level or expected value, share the same camera parameters, referred to as constant camera parameters (a, b), it is proposed that the pixels that have different expectations follow the Poissonian-Gaussian noise model with different parameters. A physical explanation of this phenomenon is that for each level, the sensor response characterized by the parameters (a, b) is not constant. Then let us define the fingerprints of camera source device as $(\mathbf{a}, \mathbf{b})$, where the vector $\mathbf{a} = \{a_1, \cdots, a_K\}$, $\mathbf{b} = \{b_1, \cdots, b_K\}$ represent the parameters which is used for linking between all pixels' expectations and variances; $k \in \{1, \cdots, K\}$ is an index for pixels' expected value μ_i, with K the number of level sets. Each level set is characterized by its center value u_i and allowed deviation Δ, $\mu_i \in [u_i - \frac{\Delta}{2}, u_i + \frac{\Delta}{2}]$. Hence the following accurate noise model is proposed:

$$z_i \sim \mathcal{N}(\mu_i, a_k \mu_i + b_k), \tag{5.5}$$

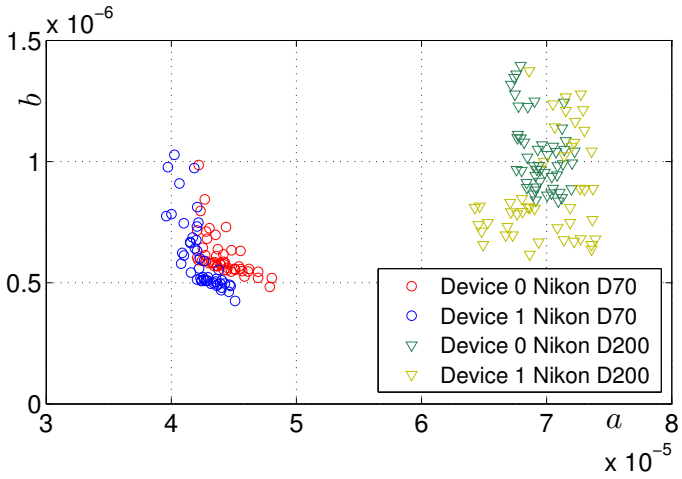
$$\text{with} \quad a_k = a \cdot w_k, b_k = b \cdot w_k^2, \tag{5.6}$$

where $\mathcal{N}(\cdot)$ denotes the Gaussian distribution with the expectation μ_i and variance $a_k \mu_i + b_k$. The camera parameters (a_k, b_k) represent the unique fingerprints originating from the k-th level set. The non-linear relationship between a_k and b_k in the k-th level set can be formulated by:

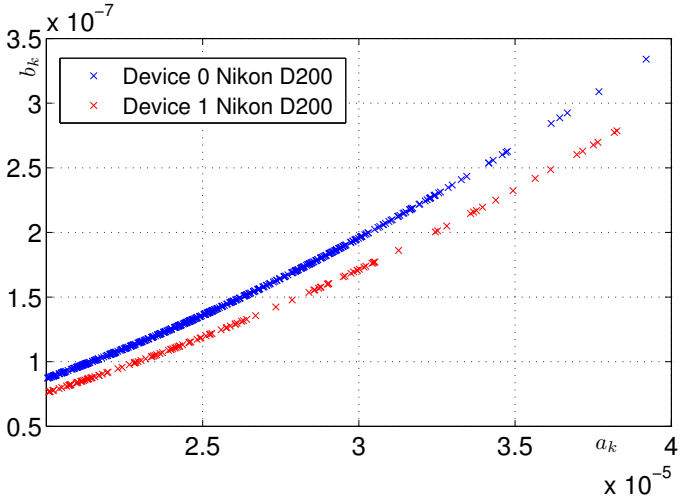
$$b_k = \frac{b}{a^2} a_k^2 \tag{5.7}$$

which is illustrated in Figure 5.3b. It should be noted that in the case of $w_k = 1$, the proposed enhanced Poissonian-Gaussian model (5.5) is reduced back to the noise model (5.3).

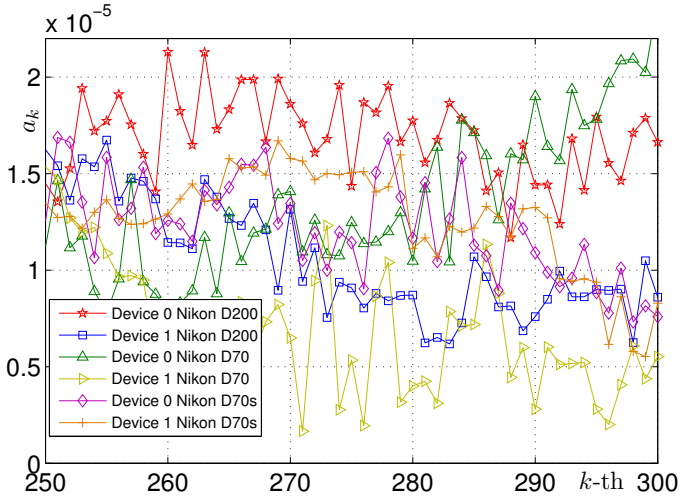
Let us illustrate our estimated camera fingerprints based on the model (5.5) in comparison with the fingerprints based on the model (5.3) proposed in [22]. For a fixed ISO sensitivity, the camera fingerprints (a, b) can distinguish between *Nikon D70* and *Nikon D200*, but are not discriminative for different devices of the same camera model, see Figure 5.3a. On the opposite, Figure 5.3b and 5.3c both illustrate that the fingerprints (a_k, b_k) proposed in this chapter are very distinguishable between different devices of the same model. Then, in Section 5.3, based on this



(a) camera parameters (a, b) estimated using [22] from 50 RAW images.



(b) camera parameters (a_k, b_k) proposed of k-th level set where the non-linear relationship is expressed by $b_k = \frac{b}{a^2} a_k^2$.



(c) camera parameter a_k in k-th level set. For simplicity, only a part of level sets are selected for comparison.

Figure 5.3: Camera fingerprints comparison of different devices per camera model with ISO 200. Natural RAW images for *Nikon D70, Nikon D70s* and *Nikon D200* are from the Dresden image database [1].

enhanced noise model, it is proposed to design the LRT assuming that we know all model parameters, and to establish its statistical properties. In a practical scenario, in Section 5.4, it is proposed to design the GLRT, which can identify among two camera devices the one that captured the image under investigation.

5.3 Likelihood Ratio Test for Two Simple Hypotheses based on RAW Format

5.3.1 Problem Statement

This chapter aims at identifying the source camera device, possibly of the same model, based on the enhanced Poissonian-Gaussian noise model (5.5). The camera device identification problem is cast within the framework of hypothesis testing theory. Supposed that a natural RAW image $\mathbf{Z} = \{z_i\}$ is under investigation, let us analyze which device $\mathcal{S}_0$ or $\mathcal{S}_1$ generated $\mathbf{Z}$. Each camera device $\mathcal{S}_j$, $j \in \{0, 1\}$ is characterized by its specific camera parameters $(a_{k,j}, b_{k,j})$. The problem consists in choosing between the two following hypotheses $\mathcal{H}_0$: "*the pixels z_i follow the Gaussian distribution* $\mathcal{N}(\mu_i, a_{k,0}\mu_i + b_{k,0})$" and $\mathcal{H}_1$: "*the pixels z_i follow the Gaussian distribution* $\mathcal{N}(\mu_i, a_{k,1}\mu_i + b_{k,1})$" which can be written formally as:

$$\begin{cases} \mathcal{H}_0 : \{z_i \sim \mathcal{N}(\mu_i, a_{k,0}\mu_i + b_{k,0})\}, \\ \mathcal{H}_1 : \{z_i \sim \mathcal{N}(\mu_i, a_{k,1}\mu_i + b_{k,1})\}. \end{cases} \tag{5.8}$$

A statistical test is a mapping $\delta : \mathbb{Z}^{I \cdot J} \mapsto \{\mathcal{H}_0, \mathcal{H}_1\}$ such that hypothesis $\mathcal{H}_j$ is accepted if $\delta(\mathbf{Z}) = \mathcal{H}_j$ (see [127] for details on hypothesis testing). As previously explained, this chapter focuses on the Neyman-Pearson bi-criteria approach: maximizing the correct detection probability for a given false alarm probability α_0. Let:

$$\mathcal{K}_{\alpha_0} = \left\{ \delta : \sup_{\boldsymbol{\theta}_0} \mathbb{P}_{\mathcal{H}_0}[\delta(\mathbf{Z}) = \mathcal{H}_1] \leq \alpha_0 \right\}, \tag{5.9}$$

be the class of tests with a false alarm probability upper-bounded by α_0. Here $\mathbb{P}_{\mathcal{H}_j}[A]$ stands for the probability of event A under hypothesis $\mathcal{H}_j, j \in \{0, 1\}$, and the supremum over model parameters $\boldsymbol{\theta}_0$, in which $\boldsymbol{\theta}_0 = (a_{k,0}, b_{k,0}, \mu_i)$, has to be understood as whatever the distribution parameters might be, in order to ensure that the false alarm probability α_0 can not be exceeded. Among all the tests in $\mathcal{K}_{\alpha_0}$, it is aimed at finding a test δ which maximizes the power function, often referred to as the true positive probability:

$$\beta_\delta = \mathbb{P}_{\mathcal{H}_1}[\delta(\mathbf{Z}) = \mathcal{H}_1], \tag{5.10}$$

which is equivalent to minimize the false positive probability $\alpha_1(\delta) = \mathbb{P}_{\mathcal{H}_1}[\delta(\mathbf{Z}) = \mathcal{H}_0] = 1 - \beta_\delta$.

The main difficulty of the problem (5.8) is to estimate the camera parameters $(a_{k,j}, b_{k,j})$ where $k \in \{1, \cdots, K\}$ and $j \in \{0, 1\}$, and image parameter μ_i where

$i \in \{1, \cdots, I\}$. In the following subsection, we will detail the statistical test that takes into account nuisance parameters, referred to as the expectations of pixels and camera fingerprints. It is proposed to study the optimal detection when those parameters are perfectly known, and meanwhile to analyze its statistical performance. When the nuisance parameters are unknown, A discussion on a practical statistical test will be presented in Section 5.4.

5.3.2 Optimal Detection Framework

When the camera and image parameters are known, the problem (5.8) is reduced to a statistical test between two simple hypotheses. In this scenario, the Neyman-Pearson Lemma [127, theorem 3.2.1] states that the most powerful test in the class $\mathcal{K}_{\alpha_0}$ (5.9) is the LRT defined, assuming that pixels z_i are independent, as:

$$\delta^{\text{lr}}(\mathbf{Z}) = \begin{cases} \mathcal{H}_0 \text{ if } \Lambda^{\text{lr}}(\mathbf{Z}) = \sum_{i=1}^{I} \Lambda^{\text{lr}}(z_i) < \tau^{\text{lr}}, \\ \mathcal{H}_1 \text{ if } \Lambda^{\text{lr}}(\mathbf{Z}) = \sum_{i=1}^{I} \Lambda^{\text{lr}}(z_i) \geq \tau^{\text{lr}}, \end{cases} \tag{5.11}$$

where the decision threshold τ^{lr} is the solution of the Equation $\mathbb{P}_{\mathcal{H}_0}\left[\Lambda^{\text{lr}}(\mathbf{Z}) \geq \tau^{\text{lr}}\right] = \alpha_0$, to ensure that the false alarm probability of the LRT equals α_0. Based on the Gaussian distribution from Equation (5.8), similar to the definition of $\boldsymbol{\theta}_0$ under hypothesis $\mathcal{H}_0$, let us define the distribution parameter $\boldsymbol{\theta}_1 = (a_{k,1}, b_{k,1}, \mu_i)$ under hypothesis $\mathcal{H}_1$. Then the probability density function (pdf) under two hypotheses can be respectively given as: $\mathcal{P}_{\boldsymbol{\theta}_0}$ and $\mathcal{P}_{\boldsymbol{\theta}_1}$. Thus, the log Likelihood Ratio (LR) [1] for one observation is given by:

$$\Lambda^{\text{lr}}(z_i) = \log \frac{\mathcal{P}_{\boldsymbol{\theta}_1}[z_i]}{\mathcal{P}_{\boldsymbol{\theta}_0}[z_i]}. \tag{5.12}$$

From the definition of (5.8), it is easy to write the LR (5.12) as:

$$\Lambda^{\text{lr}}(z_i) = \log\left(\frac{\sigma_{i,0}}{\sigma_{i,1}}\right) + \frac{\sigma_{i,1}^2 - \sigma_{i,0}^2}{2\sigma_{i,1}^2\sigma_{i,0}^2}(z_i - \mu_i)^2, \tag{5.13}$$

where the variance $\sigma_{i,j}^2 = a_{k,j}\mu_i + b_{k,j}$, $j \in \{0, 1\}$ and level set index $k \in \{1, \cdots, K\}$.

5.3.3 Statistical Performance of LRT

Due to the fact that observations are considered to be independent, the LR $\Lambda^{\text{lr}}(\mathbf{Z})$ is the sum of random variables and some asymptotic theorems allow us to establish its distribution when the number of pixels becomes "sufficiently large". Let us denote $E_{\mathcal{H}_j}(\Lambda^{\text{lr}}(z_i))$ and $V_{\mathcal{H}_j}(\Lambda^{\text{lr}}(z_i))$ as the expectation and the variance of the LR $\Lambda^{\text{lr}}(z_i)$ under hypothesis $\mathcal{H}_j, j = \{0, 1\}$. The Lindeberg's central limit theorem (CLT) [127,

[1] For simplicity in this chapter, the term LR refers to the log value of likelihood ratio.

theorem 11.2.5] states that as the number of pixels I tends to infinity, it holds true that[2]:

$$\frac{\sum_{i=1}^{I} \Lambda^{\text{lr}}(z_i) - E_{\mathcal{H}_j}(\Lambda^{\text{lr}}(z_i))}{\left(\sum_{i=1}^{I} V_{\mathcal{H}_j}(\Lambda^{\text{lr}}(z_i))\right)^{1/2}} \xrightarrow{d} \mathcal{N}(0,1)\ ,\ j \in \{0,1\}\,, \tag{5.14}$$

where $\xrightarrow{d}$ represents the convergence in distribution and $\mathcal{N}(0,1)$ is the standard normal distribution with zero mean and unit variance. This theorem is of crucial interest to establish the statistical properties of the proposed test [133,134,142,175–178]. Under hypothesis $\mathcal{H}_0$, one can normalize the LR $\Lambda^{\text{lr}}(\mathbf{Z})$ as follows:

$$\overline{\Lambda}^{\text{lr}}(\mathbf{Z}) = \frac{\Lambda^{\text{lr}}(\mathbf{Z}) - \sum_{i=1}^{I} E_{\mathcal{H}_0}(\Lambda^{\text{lr}}(z_i))}{\left(\sum_{i=1}^{I} V_{\mathcal{H}_0}(\Lambda^{\text{lr}}(z_i))\right)^{1/2}}. \tag{5.15}$$

It is thus straightforward to define the normalized LRT with $\overline{\Lambda}^{\text{lr}}(\mathbf{Z})$ by:

$$\overline{\delta}^{\text{lr}}(\mathbf{Z}) = \begin{cases} \mathcal{H}_0 & \text{if } \overline{\Lambda}^{\text{lr}}(\mathbf{Z}) < \overline{\tau}^{\text{lr}} \\ \mathcal{H}_1 & \text{if } \overline{\Lambda}^{\text{lr}}(\mathbf{Z}) \geq \overline{\tau}^{\text{lr}}. \end{cases} \tag{5.16}$$

For simplicity, it is proposed to denote

$$m_j = \sum_{i=1}^{I} E_{\mathcal{H}_j}(\Lambda^{\text{lr}}(z_i)), \tag{5.17}$$

$$\sigma_j^2 = \sum_{i=1}^{I} V_{\mathcal{H}_j}(\Lambda^{\text{lr}}(z_i)), \tag{5.18}$$

as the expectation and variance for LR $\Lambda^{\text{lr}}(\mathbf{Z})$ where $j \in \{0,1\}$. The expectation and variance for each LR $\Lambda^{\text{lr}}(z_i)$ can be expressed by:

$$\begin{aligned} E_{\mathcal{H}_j}(\Lambda^{\text{lr}}(z_i)) &= \frac{1}{2}\log\frac{a_{k,0}\mu_i + b_{k,0}}{a_{k,1}\mu_i + b_{k,1}} \\ &+ \frac{1}{2}\left(\frac{\sigma_{i,j}^2}{a_{k,0}\mu_i + b_{k,0}} - \frac{\sigma_{i,j}^2}{a_{k,1}\mu_i + b_{k,1}}\right) \end{aligned} \tag{5.19}$$

$$V_{\mathcal{H}_j}(\Lambda^{\text{lr}}(z_i)) = \frac{\sigma_{i,j}^4}{2}\left(\frac{1}{a_{k,0}\mu_i + b_{k,0}} - \frac{1}{a_{k,1}\mu_i + b_{k,1}}\right)^2. \tag{5.20}$$

Let us now establish the statistical properties of the proposed test (5.16), referred to as the decision threshold $\overline{\tau}^{\text{lr}}$ and the detection power $\beta_{\overline{\delta}^{\text{lr}}}$, which are given in the following propositions. For clarity, Φ and Φ^{-1} respectively represent the cumulative distribution function (cdf) of the standard normal distribution and its inverse.

[2]Note that we refer to the Lindeberg's CLT, whose conditions are easily verified in our case, because the random variable are independent but are not identically distributed.

Proposition 5.1. *Assuming that for the camera device identification problem as case within the two simple hypotheses (5.8), in which both the camera parameters $(a_{k,j}, b_{k,j})$ and the image parameters μ_i are known, then for any $\alpha_0 \in (0,1)$ the decision threshold:*

$$\overline{\tau}^{\text{lr}} = \Phi^{-1}(1-\alpha_0), \tag{5.21}$$

guarantees that test $\overline{\delta}^{\text{lr}}(\mathbf{Z})$ (5.16) is in the class $\mathcal{K}_{\alpha_0}$.

Proposition 5.2. *Assuming that for the camera device identification problem as case within the two simple hypotheses (5.8), in which both the camera parameters $(a_{k,j}, b_{k,j})$ and the image parameters μ_i are known, for any decision threshold $\overline{\tau}^{\text{lr}}$, the power function associated with the proposed test $\overline{\delta}^{\text{lr}}(\mathbf{Z})$ (5.16) is given by:*

$$\beta_{\overline{\delta}^{\text{lr}}} = 1 - \Phi\left(\frac{\sigma_0}{\sigma_1} \cdot \overline{\tau}^{\text{lr}} + \frac{m_0 - m_1}{\sigma_1}\right). \tag{5.22}$$

Equations (5.21) and (5.22) emphasize the main advantage of normalizing the LR as described in the relation (5.15): it allows to set any threshold that guarantees a false alarm probability independently from any distribution parameter. Besides, the detection power $\beta_{\overline{\delta}^{\text{lr}}}$ can serve as an upper bound for any statistical detector that aims at identifying source camera device. However, this ideal scenario is not practical. In fact, it is unrealistic to assume that when testing one perfectly knows both camera parameters $(a_{k,j}, b_{k,j})$ of two individual devices and image parameters μ_i whatever inspected image may be. Hence, in the next Section 5.4, the proposed test will deal with unknown camera and image parameters. The problem, thus, becomes to estimate, from the inspected image $\mathbf{Z}$, the expectation of pixels and uses those to estimate the camera parameters and finally to study the statistical properties of the GLRT that is based on those estimates.

5.4 Generalized Likelihood Ratio Test with Estimated Camera Parameters

5.4.1 Camera Parameters Estimation

This chapter employs the segmentation algorithm proposed in our prior method [22]. The image $\mathbf{Z}$ is first transformed into the wavelet domain and then segmented into K non-overlapping homogeneous level sets, denoted S_k, of size n_k, $k \in \{1, \cdots, K\}$. Among each level set S_k, all the pixels are realization of i.i.d. random variable, that is pixels of the same level set are assumed to be statistically independent and to have the same expectation and the same variance due to the proposed enhanced Poissonian-Gaussian noise model (5.5). Then, let us denote $\mathbf{z}_k^{\text{wapp}} = \{z_{k,i}^{\text{wapp}}\}_{i=1}^{n_k}$ and $\mathbf{z}_k^{\text{wdet}} = \{z_{k,i}^{\text{wdet}}\}_{i=1}^{n_k}$ as the vector of wavelet approximation coefficients and detail coefficients respectively. Since the wavelet transformation is linear, the proposed noise model in the spatial domain can be used in the wavelet domain (see details

in [172]). It is thus immediate to establish that the coefficients $z_{k,i}^{\text{wapp}}$ and $z_{k,i}^{\text{wdet}}$ follow the Gaussian distribution:

$$z_{k,i}^{\text{wapp}} \sim \mathcal{N}(\mu_k, \|\phi\|_2^2 \sigma_k^2) \tag{5.23}$$

$$z_{k,i}^{\text{wdet}} \sim \mathcal{N}(0, \sigma_k^2) \tag{5.24}$$

where $\sigma_k^2 = a_k \mu_k + b_k$ denotes the linear relationship between the expectation and variance and ϕ denotes the 2D normalized wavelet scaling function. Then in k-th level set, the Maximum Likelihood (ML) estimations of the local pixels' expectation:

$$\widehat{\mu}_k = \frac{1}{n_k} \sum_{i=1}^{n_k} z_{k,i}^{\text{wapp}}, \tag{5.25}$$

and of the local pixels' variance:

$$\widehat{v}_k = \frac{1}{n_k - 1} \sum_{i=1}^{n_k} (z_{k,i}^{\text{wdet}} - \frac{1}{n_k} \sum_{i=1}^{n_k} z_{k,i}^{\text{wdet}})^2 \tag{5.26}$$

can be averaged over several images by:

$$\widehat{\mu}_k^{\star} = \frac{1}{N \cdot n_k} \sum_{n=1}^{N} \sum_{i=1}^{n_k} z_{k,i}^{\text{wapp}} \tag{5.27}$$

$$\widehat{v}_k^{\star} = \frac{1}{N \cdot (n_k - 1)} \sum_{n=1}^{N} \sum_{i=1}^{n_k} (z_{k,i}^{\text{wdet}} - \frac{1}{n_k} \sum_{i=1}^{n_k} z_{k,i}^{\text{wdet}})^2 \tag{5.28}$$

where the vector $n \in \{1, \cdots, N\}$ denotes the index of the images used to average the estimates. Then, by using the algorithm proposed by [22], it is proposed to estimate $(\widehat{a}, \widehat{b})$ first which denotes that parameters that characterized the overall relation between pixels' expectation and variance, see (5.3). Then, based on the linear relationship between expectation $\widehat{\mu}_k^{\star}$ and variance $\widehat{v}_k^{\star}$ (5.5), the quadratic equation that relates the estimated variance $\widehat{v}_k^{\star}$ and the variance from the proposed enhanced Poissonian-Gaussian noise model is given by:

$$\widehat{v}_k^{\star} = \widehat{\mu}_k^{\star} \cdot \widehat{a}_k + \frac{\widehat{b}}{\widehat{a}^2} \cdot \widehat{a}_k^2. \tag{5.29}$$

Subsequently, the solution of Equation (5.29) can be used to propose a fast and accurate estimation of camera parameters a_k as follows:

$$\widehat{a}_k = \frac{\widehat{a}^2 \sqrt{\widehat{\mu}_k^{\star 2} + \frac{4\widehat{v}_k^{\star} \cdot \widehat{b}}{\widehat{a}^2}} - \widehat{a}^2 \cdot \widehat{\mu}_k^{\star}}{2\widehat{b}}. \tag{5.30}$$

Based on the non-linear relationship between b_k and a_k (5.7), it follows that the estimates of a_k can be use to estimate camera parameters b_k by:

$$\widehat{b}_k = \widehat{b} \cdot w_k^2 \tag{5.31}$$

where the weighting vector is given by $w_k = \frac{\widehat{a}_k}{\widehat{a}}$.

5.4.2 Design of the Practical GLRT

In Section 5.3 the framework of hypothesis testing theory has been presented assuming that all the model parameters are known for each pixel. In the practical scenario, it is much more realistic to assume that only the camera parameters $(a_{k,0}, b_{k,0})$ and $(a_{k,1}, b_{k,1})$ are known, while image parameter μ_k are unknown. The proposed practical test aims at identifying the given image $\mathbf{Z}$ which has been acquired either by device $\mathcal{S}_0$ or device $\mathcal{S}_1$. A usual solution consists in replacing the unknown parameter by its ML estimation. This leads to the construction of the following practical GLRT:

$$\widehat{\delta}_1(\mathbf{Z}) = \begin{cases} \mathcal{H}_0 \ \text{ if } \widehat{\Lambda}_1(\mathbf{Z}) = \displaystyle\sum_{k=1}^{K}\sum_{i=1}^{n_k} \widehat{\Lambda}_1(z_{k,i}^{\text{wapp}}) < \widehat{\tau}_1, \\ \mathcal{H}_1 \ \text{ if } \widehat{\Lambda}_1(\mathbf{Z}) = \displaystyle\sum_{k=1}^{K}\sum_{i=1}^{n_k} \widehat{\Lambda}_1(z_{k,i}^{\text{wapp}}) \geq \widehat{\tau}_1, \end{cases} \tag{5.32}$$

where $\widehat{\tau}_1$ represents the solution of equation:

$$\mathbb{P}_{\mathcal{H}_0}[\widehat{\Lambda}_1(\mathbf{Z}) \geq \widehat{\tau}_1] = \alpha_0, \tag{5.33}$$

and the log Generalized Likelihood Ratio (GLR) [3], denoted $\widehat{\Lambda}_1(z_{k,i}^{\text{wapp}})$ which is given, for each pixel, by:

$$\begin{aligned} \widehat{\Lambda}_1(z_{k,i}^{\text{wapp}}) = & \frac{1}{2}\log\frac{a_{k,0}\widehat{\mu}_k + b_{k,0}}{a_{k,1}\widehat{\mu}_k + b_{k,1}}, \\ & + \frac{1}{2}\Big(\frac{1}{a_{k,0}\widehat{\mu}_k + b_{k,0}} - \frac{1}{a_{k,1}\widehat{\mu}_k + b_{k,1}}\Big)\cdot\frac{(z_{k,i}^{\text{wapp}} - \widehat{\mu}_k)^2}{\|\phi\|_2^2}, \end{aligned} \tag{5.34}$$

where $\widehat{\mu}_k$ denotes the estimated value of pixels' expectation from the k-th level set, as given in Equation (5.25). In order to have a normalized GLR $\widehat{\Lambda}_2(\mathbf{Z})$ for the entire image, one can redefine Equation (5.34) as:

$$\widehat{\Lambda}_2(\mathbf{Z}) = \frac{\sum_{k=1}^{K}\sum_{i=1}^{n_k}\widehat{\Lambda}_1(z_{k,i}^{\text{wapp}}) - E_{\mathcal{H}_0}(\widehat{\Lambda}_1)}{\Big(\sum_{k=1}^{K}\sum_{i=1}^{n_k} V_{\mathcal{H}_0}(\widehat{\Lambda}_1)\Big)^{1/2}}. \tag{5.35}$$

Hence, the corresponding normalized GLRT is formulated by:

$$\widehat{\delta}_2(\mathbf{Z}) = \begin{cases} \mathcal{H}_0 \ \text{ if } \ \widehat{\Lambda}_2(\mathbf{Z}) < \widehat{\tau}_2, \\ \mathcal{H}_1 \ \text{ if } \ \widehat{\Lambda}_2(\mathbf{Z}) \geq \widehat{\tau}_2. \end{cases} \tag{5.36}$$

For simplicity, let us denote

$$m'_j = \sum_{k=1}^{K}\sum_{i=1}^{n_k} E_{\mathcal{H}_j}(\widehat{\Lambda}_1(z_{k,i}^{\text{wapp}})), \tag{5.37}$$

$$\sigma'^2_j = \sum_{k=1}^{K}\sum_{i=1}^{n_k} V_{\mathcal{H}_j}(\widehat{\Lambda}_1(z_{k,i}^{\text{wapp}})), \tag{5.38}$$

[3] For simplicity in this chapter, the term GLR refers to the log value of generalized likelihood ratio.

as the expectation and variance for GLR $\widehat{\Lambda}_1(\mathbf{Z})$ where $j \in \{0,1\}$. The expectation and variance for each GLR $\widehat{\Lambda}_1(z_{k,i}^{\text{wapp}})$ can be expressed by:

$$\begin{aligned} E_{\mathcal{H}_j}(\widehat{\Lambda}_1) &= \frac{1}{2}\log\frac{a_{k,0}\mu_k + b_{k,0}}{a_{k,1}\mu_k + b_{k,1}} \\ &+ \frac{1}{2}\left(\frac{1}{a_{k,0}\mu_k + b_{k,0}} - \frac{1}{a_{k,1}\mu_k + b_{k,1}}\right) \\ &\cdot \left(\sigma_{k,j}^2 + \frac{\sigma_{k,j}^2}{n_k}\right), \end{aligned} \tag{5.39}$$

$$\begin{aligned} V_{\mathcal{H}_j}(\widehat{\Lambda}_1) &= \frac{\sigma_{k,j}^2\|\phi\|_2^2}{4n_k}\left(\frac{a_{k,0}b_{k,1} - a_{k,1}b_{k,0}}{(a_{k,0}\mu_k + b_{k,0})(a_{k,1}\mu_k + b_{k,1})}\right)^2 \\ &+ \frac{\sigma_{k,j}^4}{2}\left(\frac{1}{a_{k,0}\mu_k + b_{k,0}} - \frac{1}{a_{k,1}\mu_k + b_{k,1}}\right)^2\left(1 + \frac{1}{n_k}\right)^2 \\ &+ \frac{3\|\phi\|_2^2\sigma_{k,j}^6}{4n_k}\left(\frac{a_{k,1}}{(a_{k,1}\mu_k + b_{k,1})^2} - \frac{a_{k,0}}{(a_{k,0}\mu_k + b_{k,0})^2}\right)^2 \\ &\cdot \left(1 + \frac{1}{n_k}\right)^2, \end{aligned} \tag{5.40}$$

where since the image parameter μ_k is unknown in our GLRT, it is proposed to use $\widehat{\mu}_k$ instead of μ_k. The mathematical deduction of $E_{\mathcal{H}_j}(\widehat{\Lambda}_1)$ and $V_{\mathcal{H}_j}(\widehat{\Lambda}_1)$ is referred to [22, Appendix B].

Similar to the optimal detection framework presented in Section 5.3, one can establish the statistical properties of the proposed GLRT, that analytically express the decision threshold $\widehat{\tau}_2$ for a prescribed false alarm probability and the detection power $\beta_{\widehat{\delta}_2}$ in the following two propositions.

Proposition 5.3. *Assuming that the pixels are modelled by the proposed enhanced Poissonian-Gaussian model (5.5), when the camera parameters $(a_{k,j}, b_{k,j})$ are known and the pixels' expectations $\widehat{\mu}_k$ are estimated as in (5.25), then any $\alpha_0 \in (0,1)$ the decision threshold of the proposed GLRT $\widehat{\delta}_2(\mathbf{Z})$ is given by:*

$$\widehat{\tau}_2 = \Phi^{-1}(1 - \alpha_0), \tag{5.41}$$

Proposition 5.4. *Assuming that the pixels are modelled by the proposed enhanced Poissonian-Gaussian model (5.5), when the camera parameters $(a_{k,j}, b_{k,j})$ are known and the pixels' expectations $\widehat{\mu}_k$ are estimated as in (5.25), for any decision threshold $\widehat{\tau}_2$, the power function associated with test $\widehat{\delta}_2$ (5.36) is given by*

$$\beta_{\widehat{\delta}_2} = 1 - \Phi\left(\frac{\sigma_0'}{\sigma_1'} \cdot \widehat{\tau}_2 + \frac{m_0' - m_1'}{\sigma_1'}\right). \tag{5.42}$$

Again, the main advantages of the proposed GLRT $\widehat{\delta}_2(\mathbf{Z})$ are 1) that its statistical performance can be analytically established and, hence, one can easily compute the decision threshold that guarantees a prescribed false alarm probability along with the ensuing power function and 2) thanks to the normalization the decision threshold only depends on the prescribed false alarm probability α_0.

Table 5.1: Image database statistic in RAW format

Database	Camera Device	Alias	Sensor size	Resolution	Bit depth	ISO	NO. images
Dresden [1]	Nikon D70 # 0	N_D70_0	23.7 × 15.6 mm CCD	2014×3039	12	200	180
	Nikon D70 # 1	N_D70_1	23.7 × 15.6 mm CCD	2014×3039	12	200	189
	Nikon D70s # 0	N_D70s_0	23.7 × 15.6 mm CCD	2014×3039	12	200	178
	Nikon D70s # 0	N_D70s_1	23.7 × 15.6 mm CCD	2014×3039	12	200	189
	Nikon D200 # 0	N_D200_0	23.6 × 15.8 mm CCD	2616×3900	12	200	372
	Nikon D200 # 1	N_D200_1	23.6 × 15.8 mm CCD	2616×3900	12	200	380
Our own data	Canon 100D # 0	C_100D_0	22.3 × 14.9 mm CMOS	3528×5280	14	200	317
	Canon 100D # 1	C_100D_1	22.3 × 14.9 mm CMOS	3528×5280	14	200	389
	Pentax K-50 # 0	P_K50_0	23.7× 15.7 mm CMOS	3272×4936	12	200	269
	Pentax K-50 # 1	P_K50_1	23.7× 15.7 mm CMOS	3272×4936	12	200	211
	Nikon D5200 # 0	N_D5200_0	23.5× 15.6 mm CMOS	4020×6036	14	200	300
	Nikon D5200 # 1	N_D5200_1	23.5× 15.6 mm CMOS	4020×6036	14	200	755
$\sum 2$	12	12	5	5	2	1	3729

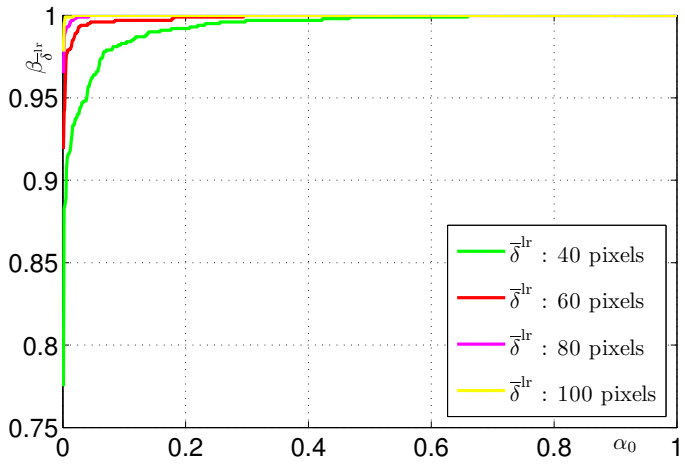
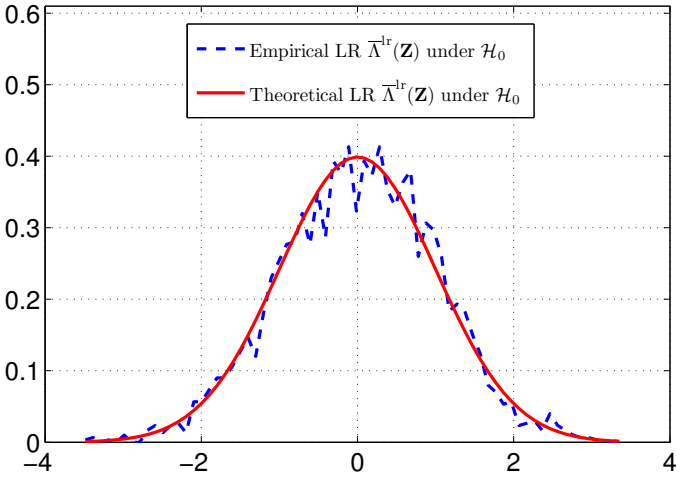


Figure 5.4: Detection performance of the test $\overline{\delta}^{\mathrm{lr}}$ on the assumption of knowing camera and image parameters on simulated images.

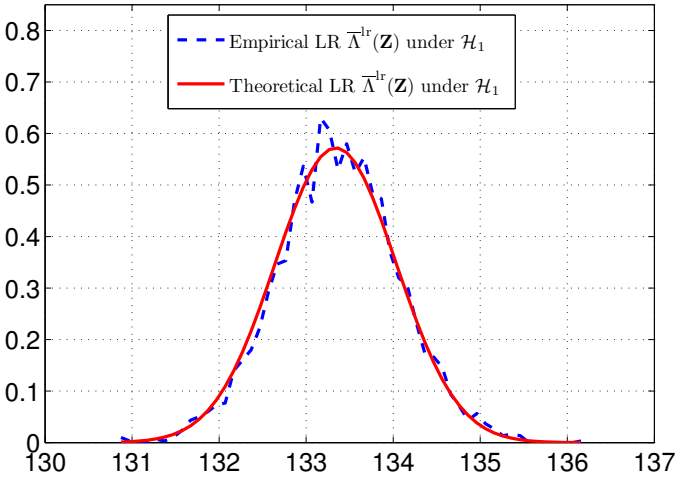
5.5 Numerical Experiments

5.5.1 Results on Simulated Images for RAW Format

To verify the sharpness of the theoretically established results, it is proposed to use a Monte Carlo simulation on a 8-bit synthetic image of size 512×512 with 3000 repetitions. It should be noted that the synthetic image is normalized into the interval $[0, 1]$. In our simulated experiments, it is proposed to simulate 256 level sets, thus 256 pairs of camera parameters of $(a_{k,0}, b_{k,0})$ and $(a_{k,1}, b_{k,1})$, $k \in \{1, \cdots, 256\}$ respectively characterize device $\mathcal{S}_0$ and $\mathcal{S}_1$. Figure 5.4 illustrates the detection performance as a Receiver Operating Character (ROC), that is present the detection power $\beta_{\overline{\delta}^{\mathrm{lr}}}$ as a function of the false alarm probability α_0, for $I = \{40, 60, 80, 100\}$, the number of pixels used in our proposed test $\overline{\delta}^{\mathrm{lr}}$ (5.16). Before testing, the camera parameters and the pixels' expectations are estimated previously by using our proposed algorithms (see details in Section 5.4). As Figure 5.4 shows,



(a) Under $\mathcal{H}_0$



(b) Under $\mathcal{H}_1$

Figure 5.5: Comparison between empirical and theoretical distribution of $\overline{\Lambda}^{\mathrm{lr}}(\mathbf{Z})$ under hypothesis $\mathcal{H}_0$ and hypothesis $\mathcal{H}_1$.

the test performs better and better with increasing the number of pixels.

To verify the efficiency of the established LRT $\overline{\delta}^{\mathrm{lr}}$, again, a Monte-Carlo simulation is performed by repeating 3000 times using all the pixels for each synthetic image. Under hypothesis $\mathcal{H}_0$, the distribution of the empirical $\overline{\Lambda}^{\mathrm{lr}}(\mathbf{Z})$ is illustrated in Figure 5.5a. For comparison, we plot the theoretical pdf of $\overline{\Lambda}^{\mathrm{lr}}(\mathbf{Z})$ following the standard normal distribution with zero mean and unit variance. Similarly, under hypothesis $\mathcal{H}_1$, empirical and theoretical distribution of $\overline{\Lambda}^{\mathrm{lr}}(\mathbf{Z})$ are illustrated in Figure 5.5b. It should be noted that the theoretical pdf of $\overline{\Lambda}^{\mathrm{lr}}(\mathbf{Z})$ following the Gaussian distribution with the mean $\frac{m_1-m_0}{\sigma_0}$ and the variance $\frac{\sigma_1^2}{\sigma_0^2}$. The results of Figure 5.5 highlight the accuracy of the theoretically established distribution of the proposed LR and, hence, the statistical performance of the LRT (5.16).

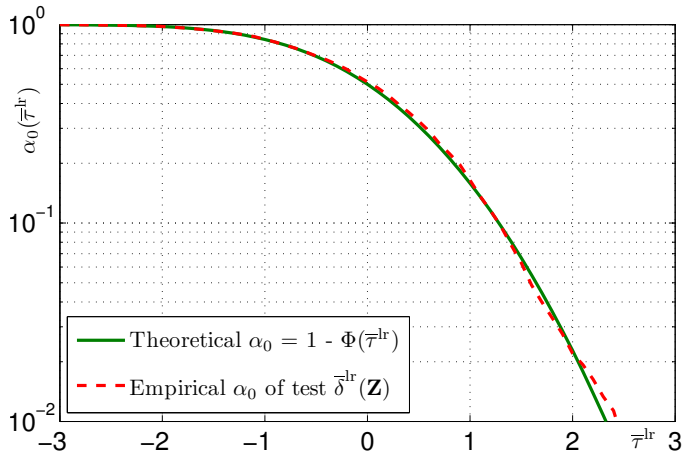


Figure 5.6: Comparison between the theoretical false alarm probability α_0 and its empirical values, plotted as a function of the threshold $\overline{\tau}^{\mathrm{lr}}$.

The goal of this chapter is to design a statistical test that can warrant a prescribed false alarm probability. Thus, it is proposed to compare the empirical false alarm probability with the theoretically established one for the test (5.16) as a function of the threshold $\overline{\tau}^{\mathrm{lr}}$ (see Figure 5.6). The results presented in Figure 5.6 show that the empirical result perfectly matches the theoretical false alarm probability, which confirms the ability of the proposed LRT to guarantee a prescribed false alarm probability in practice. In some cases ($\overline{\tau}^{\mathrm{lr}} \geq 2$), it should be noted that the slight differences of two cures in Figure 5.6 are due to the inaccuracy of the CLT for modelling tails.

5.5.2 Results on Real Images for RAW Format

Finally, let us demonstrate the detection performance on the real images. In our experiments, the Dresden image database [1] and our own database are both used. From Dresden database, all full-resolution images from the three following camera models *Nikon D70, Nikon D70s* and *Nikon D200* are selected; for all those models the Dresden database proposes images captured by two different devices. In fact, the database of [1] only provides these three models capturing the images in RAW format. To diversify camera models in the database, we add our own database of three models: *Canon 100D, Pentax K-50* and *Nikon D5200* into the experiments. The specific parameter settings of the cameras used in our experiments are illustrated in Table 5.1. Each RAW format image is converted to an uncompressed format using the software *Dcraw* (with parameters -D -4 -j -v -r 1 1 1 1, one can obtain a full-resolution image on 12 or 14 bits, depending on the camera model, without any processing.) and decompressed into 4 sub-images. It is proposed to use only the red color channel. Besides, it should be noted that our proposed model is sensitive to the ISO which is pre-set as the same value (200 in our experiments) before capturing an image (see details in Section 5.2).

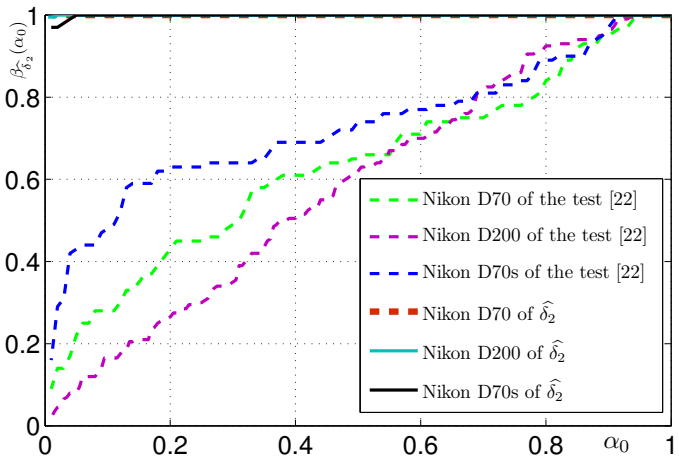


Figure 5.7: Detection performance of the test $\widehat{\delta}_2$: parameter $(a_{k,0},\ b_{k,0})$ of $\mathcal{S}_0$ and $(a_{k,1},\ b_{k,1})$ of $\mathcal{S}_1$ are known, μ_k are unknown on real images. Note that for each camera model, the test $\widehat{\delta}_2$ is designed for identifying the inspected image $\mathbf{Z}$ acquired by $\mathcal{S}_0$ or $\mathcal{S}_0$, which both belongs to the same model.

First, it is proposed to divide each set of images from each device into two subsets: "*Learning Subset*" and "*Testing Subset*". Images of "*Learning Subset*" are used to extract camera fingerprints from each device; images of "*Testing Subset*" are used to identify the origin of a given image. It should be noted that "*Learning Subset*" and "*Testing Subset*" are disjoints, that is none of images are used for estimated camera device parameters and for testing. The number of "*Learning Subset*" is set as 50, which is a good compromise between estimation accuracy and computation efficiency. In the experiments of camera device identification, each model has two devices: $\mathcal{S}_0$ and $\mathcal{S}_1$ which respectively represents the hypotheses $\mathcal{H}_0$ and $\mathcal{H}_1$.

To the best of our knowledge, few camera identification methods work on images in RAW format. The test proposed by [22] based on the Poissonian-Gaussian noise model (5.3) opens the way of investigating the problem of identifying source camera model based on RAW data. Thus, it is proposed to compare our test with the detector of [22]. As Figure 5.7 illustrates, our proposed detector has the considerable ability of identifying source camera device with a high performance. On the contrary, the ROC curves of the detector [22] perform, roughly speaking, a "random guess", since the power function approximatively equals the false alarm probability. Therefore, it can not identify the different devices of the same model. Moreover, it is proposed to present the detection performance by using all the images in the database, which is illustrated in Figure C.6. Based on the enhanced Poissonian-Gaussian noise model (5.5), ROC curves show the relevance of our designed detector.

Moreover, it is proposed to compare our proposed test with the state-of-the-art detector of [48]. To this end, 50 images from "*Learning Subset*" has been used to calculate a reference PRNU as the fingerprint of the camera device $\mathcal{S}_0$. Then the

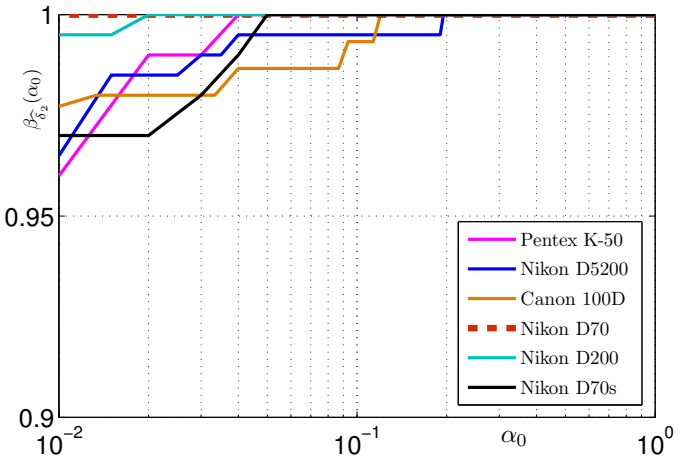


Figure 5.8: Comparison of detection performance on a large scale database (logarithmic scale).

Peak to Correlation Energy (PCE), used for "*Testing Subset*", is computed to detect whether or not the inspected image is captured with the device $\mathcal{S}_0$. For clarity of the comparison, the usual criterion of minimal Probability of Error (or P_E for short) under equal priors is used. This performance criterion corresponds to the minimal value of false alarm and missed detection probability and is formally defined as:

$$P_E = \min_{\alpha_0 \in (0,1)} \frac{\alpha_0 + (1 - \beta_{\hat{\delta}_2})}{2}. \tag{5.43}$$

Table 5.2 compares the empirical performance, through P_E criterion, of the proposed test, the test proposed in [22] and the state-of-the-art test proposed in [48]. Note that the results presented in Table 5.2 are for identification of two different devices of each tested model. This table obviously shows that the test proposed in [22] for camera model identification, and based on the Poissonian-Gaussian noise model, fails to identify the instance, while our proposed test based on the enhanced Poissonian-Gaussian noise performs similarly as the test [48].

To complete the results from Table 5.2, the results presented in Table 5.3 compares, for the same experiment and the same tests, the empirical power for a fixed false alarm probability of α_0. This table also shows that the proposed methodology and the state-of-the-art detector of [48] exhibit roughly the same performance.

Our proposed detector does not only identify the source camera device from the same model, but also from different camera models. Then, the detection power comparison is illustrated in Table 5.4 and 5.5. By comparison, one can observe that our proposed test obviously outperforms the detector of [22], and very comparable to the prior-art detector of [48].

Table 5.2: Minimal P_E comparison, assuming that two devices $\mathcal{S}_0$ and $\mathcal{S}_1$ from the same camera model

Camera Model	Proposed Test $\widehat{\delta}_2$	Test [22]	Test [48]
Nikon D70	0.005	0.380	0.015
Nikon D70s	0.020	0.325	0.005
Nikon D200	0.005	0.417	0.002
Canon 100D	0.016	0.347	0.020
Pentax K-50	0.015	0.355	0.020
Nikon D5200	0.015	0.560	0.005
Average	0.012	0.397	0.011

Table 5.3: Detection power comparison at the given false alarm probability α_0, assuming that two devices $\mathcal{S}_0$ and $\mathcal{S}_1$ from the same camera model

(a) $\alpha_0 = 0.05$

Camera Model	Proposed Test $\widehat{\delta}_2$	Test [22]	Test [48]
Nikon D70	1.00	0.25	1.00
Nikon D70s	1.00	0.34	1.00
Nikon D200	1.00	0.11	1.00
Canon 100D	0.98	0.22	0.96
Pentax K-50	1.00	0.19	0.97
Nikon D5200	0.99	0.28	0.99
Average	0.99	0.23	0.98

(b) $\alpha_0 = 0.01$

Camera Model	Proposed Test $\widehat{\delta}_2$	Test [22]	Test [48]
Nikon D70	1.00	0.09	0.96
Nikon D70s	0.97	0.15	1.00
Nikon D200	1.00	0.01	1.00
Canon 100D	0.97	0.12	0.96
Pentax K-50	0.96	0.11	0.97
Nikon D5200	0.97	0.01	0.99
Average	0.98	0.08	0.98

Table 5.4: Detection power comparison between N_D70_0 under hypothesis $\mathcal{H}_0$ and the selected devices from the same brand under hypothesis $\mathcal{H}_1$ at the given false alarm probability α_0

(a) $\alpha_0 = 0.05$

Camera Model	Proposed Test $\widehat{\delta}_2$	Test [22]	Test [48]
N_D70s_0	1.00	0.11	0.99
N_D70s_1	0.97	0.25	1.00
N_D200s_0	1.00	0.66	1.00
N_D200s_1	0.99	0.57	1.00
Average	0.99	0.39	1.00

(b) $\alpha_0 = 0.01$

Camera Model	Proposed Test $\widehat{\delta}_2$	Test [22]	Test [48]
N_D70s_0	0.99	0.11	0.97
N_D70s_1	0.90	0.25	0.99
N_D200s_0	0.98	0.66	0.99
N_D200s_1	0.99	0.57	0.99
Average	0.97	0.24	0.99

Table 5.5: Detection power comparison between N_D70_1 under hypothesis $\mathcal{H}_0$ and the selected devices from different brands under hypothesis $\mathcal{H}_1$ at the given false alarm probability α_0

(a) $\alpha_0 = 0.05$

Camera Model	Proposed Test $\widehat{\delta}_2$	Test [22]	Test [48]
P_K50_0	1.00	0.56	0.99
P_K50_1	1.00	0.59	1.00
C_100D_0	1.00	0.93	1.00
C_100D_1	1.00	0.91	1.00
Average	1.00	0.75	0.99

(b) $\alpha_0 = 0.01$

Camera Model	Proposed Test $\widehat{\delta}_2$	Test [22]	Test [48]
P_K50_0	1.00	0.44	0.96
P_K50_1	1.00	0.47	0.97
C_100D_0	1.00	0.63	1.00
C_100D_1	1.00	0.25	1.00
Average	1.00	0.45	0.98

5.6 Conclusion and Discussion

This chapter studies the problem of source camera device identification. In the most literature, the proposed detectors mainly focus on camera fingerprints extracted from PRNU. In this context, it is proposed to use the enhanced Poissonian-Gaussian noise model, characterized by the camera fingerprints $(\mathbf{a}, \mathbf{b})$ where the vector $\mathbf{a} = \{a_1, \cdots, a_K\}$, $\mathbf{b} = \{b_1, \cdots, b_K\}$. Based on this noise model, the problem is cast in the framework of hypothesis testing theory. The main strength of this chapter is the designing of the optimal LRT and the practical GLRT. First, assuming that the camera fingerprints and image parameters are perfectly known in advance, the LRT is designed and its statistical performance is analytically established. Then, in the practical case, based on the estimated image parameters, the practical GLRT is established and performs very well at the prescribed false alarm probability.

To the best of our knowledge, our proposed test in this chapter is the only one detector of identifying individual camera device based on RAW images. Although the detector proposed in [22] are also designed for RAW images, it can only identify source camera model, which is different from the source camera device. In fact, numerical experiments in this chapter verify that the detector of [22] fails to identify source camera device. In the forensic community of source camera identification, most detectors are designed based on the prior-art PRNU [42, 47, 48]. Our proposed test based on the enhanced Poissonian-Gaussian noise definitely opens the new way of identifying source camera device.

Besides, many prior-art detectors mainly focus on the source camera identification based on JPEG/TIFF images. Our proposed test in this context is the only one practical detector which deals with the problem of identifying source camera device of the same model based on RAW images. Therefore, this chapter extends the forensic research of this field. To improve the application of our proposed detector, in our future work, we will extend our approach to design another practical detector based on JPEG images. The main limitation of the proposed algorithm is that camera fingerprints from RAW format require the constant ISO value.

Chapter 6

Individual Camera Device Identification Based on JPEG Images

Contents

6.1 Introduction and Contributions

In general, when a photographer captures a digital image by using a camera, the image is stored with the header files (e.g. Exchangeable Image File, EXIF and Joint Photographic Experts Group, JPEG headers), which contains all recording and compression history. Thus, forensic investigators can potentially access to all the information on the inspected image by extracting its header files. Header files as

extrinsic fingerprints can be used to solve the problem of identifying source camera. The original header files, however, can be easily removed or replaced due to post-camera operations, that is referred to as all possible operations that one can apply on a digital image after its acquisition; nowadays free and open source softwares for image header files are readily available such as exiftool software or libexif library. Besides, a mass of photos shared on the Internet, especially on the social web-site such as Facebook or Twitter or on photo-sharing platform such as Flickr web-site, do not always have their header files. Therefore, passive forensic investigators do not use header files as fingerprints for identifying source camera as it is usually considered as unreliable. In the image acquisition pipeline (see details in [2, 41]), each step probably leaves some traces in the image. The passive forensics utilizes the traces as intrinsic fingerprints to acquire the concerning information of the camera source[1]. The methods that investigate the problem of image origin identification aim at addressing the following questions:

- Which intrinsic/in-camera[2] fingerprint can be exploited for identifying with reliability and accuracy the image origin?
- How to extract accurately the proposed fingerprint from a single image or several images that may have a textured context and an important noise level in some areas?
- What are the statistical performance of the proposed detector, e.g. what is the probability that may classify two fingerprints extracted from different images as coming from the same source?

Many forensic detectors have been proposed, which can be generally formulated into two schemes: *supervised classification* and *unsupervised classification*, see [97, 179, 180] for a detailed review:

1. The methodologies in the category of supervised classification mainly rely on identifying unique traces of the imaging pipeline and exploiting the classification approach from a supervised machine. Many different traces (or adapted features set) have been proposed and many statistical learning frameworks have been used. Only a few works have been proposed to identify the camera device from early-acquisition stages such as lens distortion/aberration [16, 181, 182]. Forensic investigators can also utilize the traces left by the stage of post-processing for source camera identification. For instance, using the white balancing as a camera fingerprint, the algorithm proposed in [169] identified the device origin of a given image. The methodologies proposed by [14, 15, 51, 52] made use of Color Filter Array (CFA) and demosaicing algorithms to identify the camera model. Besides, the features extracted by JPEG compression

[1]The term *source* means an individual camera device, a camera model, or a camera brand. Other sources such as cell-phone cameras, scanners are not addressed in this chapter.

[2]The term *in-camera* is referred to as all the processing steps that are unavoidably carried out during image acquisition pipeline, which do not include post-camera operations.

were also considered as influential factors to identify the camera origin [57]. Based on extracted features from the acquisition pipeline, these algorithms use supervised learning methods (such as Support Vector Machine, SVM) to identify source camera device from a digital image. However, the challenging problems are that a few manufacturers share the similar image processing techniques and most of the time partially the same components produced by a few manufacturers lead to the similarity when extracting fingerprints from given images, especially in the case of the same camera manufacturer and model. Moreover, the application of supervised learning is time-consuming. Besides, several problems such as accuracy of feature selection, robustness to a mismatch between training and testing sets, and the establishment of detection performance remain open since with any statistical learning approach the performance are only measured empirically on a validation dataset.

2. The methods in the category of unsupervised classification aim at identifying the unique noise as fingerprints of the acquisition device. Indeed, the level of noise that affects each pixel individually can be considered as a unique fingerprint for each camera device. The origin of noise level variation from pixel to pixel is due to imperfections during sensor manufacturing process and non-uniformity of photo-electronic conversion caused by inhomogeneity of silicon wafer (see [40, 41] for details). Hence, Sensor Pattern Noise (SPN) was generated, which can be extracted from a given image, and was first used for identifying source camera device [42]. Then, [45,47–49,170] improved the prior algorithm. It should be noted that the SPN includes two main components: Fixed Pattern Noise (FPN) and Photo-Response Non-Uniformity noise (PRNU). The FPN represented by dark current, which was used in [43]. Since the FPN can be suppressed by subtracting a dark frame from the output image, it was proposed not to use the FPN any more for source camera identification. On the contrary, the PRNU is more robust than the FPN. Based on the PRNU, the algorithms proposed in [47,48,170,171] directly investigated the problem of source camera identification. Although those PRNU-based detectors performed efficiently, the counter-forensic algorithm [103] challenged the credibility of their detection power. Therefore, the main challenge in this category is to extract relevantly reliable noise-based fingerprints from a given image. Besides, in the literature, few detectors investigate the hypothesis testing theory and are designed based on the statistical image model. Hence, the performance of detectors still remains analytically unestablished.

In a practical context, it is proposed to use the various sources of random noise (e.g. shot noise and read-out noise) as the unique fingerprints for individual device identification. In our prior research, based on the generalized signal-dependent noise model [183], a statistical test that can identify source camera model, from a given image, was designed in [21]. However, those prior works have the indisputable disadvantage to be unable of distinguishing different devices from the same camera model. By improving the signal-dependent noise model and extracting block fin-

gerprints, we develop a new statistical test that aims at identifying the individual camera device from JPEG images. Hence, the main contributions of this chapter are summarized below:

- By taking into account the impact of the main in-camera post-processing stage (such as white balancing, gamma correction) on the variance of each pixel and by studying the non-linearity of pixel's response, this chapter proposes an improved signal-dependent noise model describing more accurately the relation between pixel's expectation and variance.
- This novel noise model is used in the present chapter over blocks of several pixels to extract camera device fingerprints. Moreover, it is shown that the camera parameters describing our improved signal-dependent noise model have a linear relation, which can be used for designing our proposed tests.
- In an ideal context, that is when all the parameters are perfectly known, the Likelihood Ratio Test (LRT) is presented, and its properties are theoretically established. The statistical performance of LRT serves as an upper bound on the detection power of any test that aims at identifying source camera device using individual pixel's noise properties.
- In a practical context, it is proposed to design two Generalized Likelihood Ratio Tests (GLRT)s. In the first scenario, in which both camera parameters characterizing the improved signal-dependent noise model and model parameters describing the relation of camera fingerprints are known, and the statistical parameter of estimate errors is unknown, the first GLRT is designed. In another scenario, when both camera and model parameters are unknown, as well as the statistical parameter of estimate errors, the second GLRT is established.
- Numerical simulations show the sharpness of the theoretically established results and the good performance of our proposed tests. Comparisons with prior-art detector [48, 183] on the real image dataset show the efficiency of our proposed method for source camera device identification.

This chapter is organized as follows. Section 6.2 reviews the generalized signal-dependent noise model for a JPEG image, then improves this noise model. Based on this novel noise model, this section presents the parameters that are used as fingerprints for camera device identification, details their extraction over blocks of pixels, and presents the linear relation between those parameters. Based on the proposed linear model for block fingerprints extracted from several JPEG images, the LRT is established and its statistical performance is presented in Section 6.3. While the optimal LRT requires knowledge of camera fingerprints, model parameters and estimate error's parameter, Section 6.4 presents the first practical GLRT for JPEG images that estimates the error's parameter, but still assumes both camera and model parameters known. Section 6.5 proposes even more practical GLRT that

estimates the camera fingerprints, model parameters and error's parameter. Section 6.6 presents numerical results of the proposed detectors on the simulated and real image dataset, and also presents comparison with current art. Finally, Section 6.7 concludes this chapter.

6.2 Proposed Camera Fingerprints for JPEG Images

6.2.1 Overview on Generalized Signal-Dependent Noise Model and Its Limitation

Digital image acquisition pipeline actually consists in various processing stages briefly described below (the reader interested may find more details in [2,41]). First, light intensity measured at each pixel generates an electrical signal that is read out as a RAW format image. This RAW image is then subjected to several post-acquisition processes, such as demosaicing, white balancing and gamma correction, to issue a full resolution, colored and uncompressed image, that is referred to as a TIFF image in this chapter. Besides, other processing operations, that widely differ among different camera models, such as camera denoising and edge enhancement are not discussed in this chapter. Then, for storage, an image compression algorithm is applied very often using the JPEG standard and, hence, a JPEG image is finally output.

Let us denote $\mathbf{Z} = \{z_i\}$, $i \in \{1, \cdots, I\}$ a RAW image, made of I pixels. The noise corrupting a RAW image is usually described using the Poissonian-Gaussian noise model [172], which includes Gaussian read-out noise and Poissonian photo-counting shot noise, as follows:

$$z_i = \mu_{z_i} + \xi_{z_i} \sim \mathcal{N}(\mu_{z_i}, a\mu_{z_i} + b). \tag{6.1}$$

This model describes the pixel noise variance as an affine function of the pixel's expectation, characterized by the parameters (a, b). Recently, this model has been extended in [22,25] to describe a RAW image.

In fact, since demosaicing and white balancing are linear transformation, see details in [2], the application of those processes does not modify the Gaussian distribution of pixels up to the parameters (a, b):

$$y_i = \mu_{y_i} + \eta_{y_i} \sim \mathcal{N}(\mu_{y_i}, \widetilde{a}\mu_{y_i} + \widetilde{b}) \tag{6.2}$$

where y_i denotes a pixel after white balancing and demosaicing, but before gamma correction, and μ_{y_i} the expectation of y_i which still characterizes the variance through the affine relation $\widetilde{a}\mu_{y_i} + \widetilde{b}$. In addition, $\eta_{y_i} \sim \mathcal{N}(0, \widetilde{a}\mu_{y_i} + \widetilde{b})$ represents the signal-dependent noise with zero mean. Next, a usual processing that occurs, to adjust the brightness with former CTR (Cathode Ray Tube) display device, is the gamma correction, which is a pixel-wise non-linear operation defined as:

$$x_i = y_i^{\frac{1}{\gamma}} \tag{6.3}$$

where x_i denotes a TIFF pixel after gamma correction, and γ the correction factor. The statistical distribution of pixels after gamma correction is not very simple. However, based on the first order of Taylor series expansion of $(1+p)^{\frac{1}{\gamma}}$ at $p=0$, x_i can be approximated by:

$$x_i \approx \mu_i + \frac{1}{\gamma}\mu_i^{1-\gamma}\eta_i \sim \mathcal{N}(\mu_i, \sigma_i^2) \tag{6.4}$$

$$\text{with } \sigma_i^2 = \frac{1}{\gamma^2}\mu_i^{2-2\gamma}(\widetilde{a}\mu_i^{\gamma} + \widetilde{b}) \tag{6.5}$$

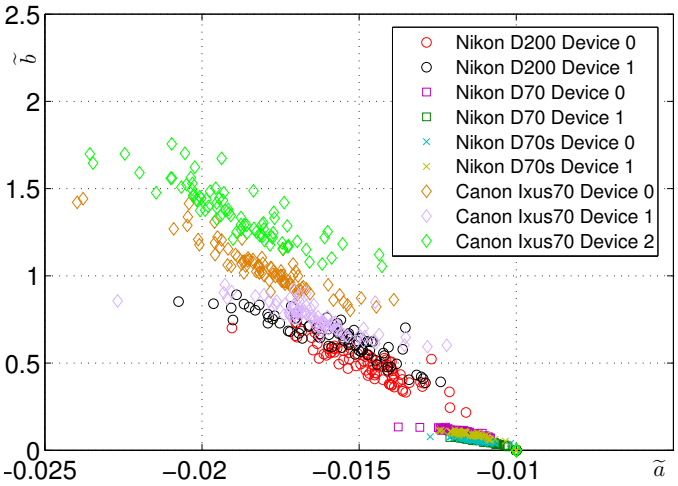
where μ_i and σ_i^2 represent the expectation and the variance of pixel x_i. Finally, an image in TIFF format is usually transformed to a JPEG one for storage. Thus, let us consider the quantization noise modelled as an addictive noise that is uniformly distributed and independent of the input signal, see [184]. Then Equation (6.5) can be rewritten as follows:

$$\sigma_i^2 = \frac{1}{\gamma^2}\mu_i^{2-2\gamma}(\widetilde{a}\mu_i^{\gamma} + \widetilde{b}) + \frac{\Delta^2}{12} \tag{6.6}$$

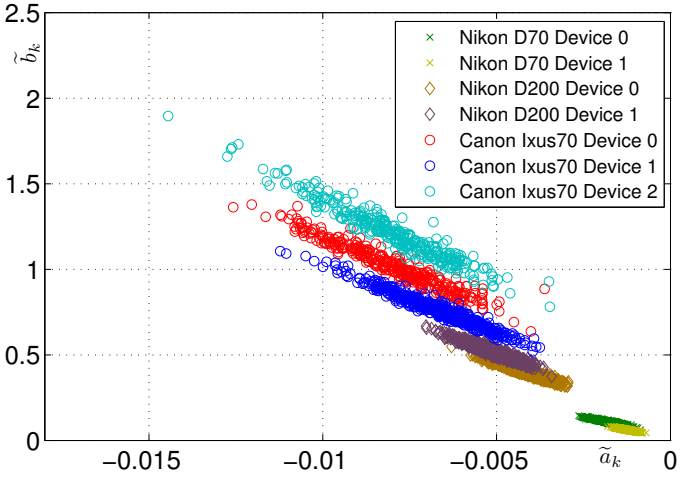
where $\frac{\Delta^2}{12}$ is the term modelling the variance of additive quantization noise. Besides, in this chapter we assume, for clarity, that quantization step Δ equals 1. It should be noted that based on Equation (6.6), x_i of the generalized signal-dependent model (6.4) is redefined as the pixel from an image in JPEG format. It also should be noted that in this model, camera parameters $(\widetilde{a}, \widetilde{b})$ are constant for the whole inspected image.

Compared with the Poissonian-Gaussian noise model for a RAW image, the extra parameter γ of the model (6.4) is very difficult to estimate since it has a non-linear impact on pixel's statistical properties. Finally, it should also be noted that the parameters $(\widetilde{a}, \widetilde{b}, \gamma)$ estimated from JPEG images are relatively insensitive to ISO sensitivity parameters, unlike the fingerprints used for RAW images, see more details in [22, 25]. This allows the design of fingerprints and a statistical test for camera device identification which are independent of ISO sensitivity settings. It should be noted that in the present chapter the pixel's value of the inspected image is used ; hence we do not intend in this chapter to model the DCT coefficient, see details in [136, 149, 161] for instance.

In our prior research [21], this generalized signal-dependent noise model (6.4) has been used for extracting the camera model fingerprints $(\widetilde{a}, \widetilde{b})$ from a single JPEG image. However, $(\widetilde{a}, \widetilde{b})$ can only be used to identify source camera model. Unfortunately they can not identify different devices of the same camera model. Figure 6.1a provides a simple example showing the parameters $(\widetilde{a}, \widetilde{b})$ of the noise model (6.4) extracted from several images captured with a few devices of the same models. It is obvious that those estimated parameters overlap and, hence, lead to a poor statistical test, whose detection performance is weak. In fact, even though in the test of [21] the average $(\widetilde{a}, \widetilde{b})$ is used for detection, the very similar $(\widetilde{a}, \widetilde{b})$ of each device unavoidably leads to unsatisfying detection results, see Section 6.6. Thus, in the following subsection 6.2.2, it is proposed to improve the generalized



(a) Estimated camera parameters $(\widetilde{a}, \widetilde{b})$ proposed in [21]. Each result represents the parameters estimated from a single JPEG image.



(b) Estimated camera parameter $(\widetilde{a}_k, \widetilde{b}_k)$ proposed in this chapter. Each result represents the parameters estimated from a single block.

Figure 6.1: Camera fingerprints comparison with several devices for each camera model. Natural JPEG images for *Nikon D70, Nikon D70s*, *Canon Ixus70* and *Nikon D200* are from Dresden dataset [1].

signal-dependent noise model (6.4) and extract the novel device block fingerprints whose linear relationship is exposed for establishing the statistical tests.

6.2.2 Description of Improved Noise Model and Block Fingerprints

By challenging the assumption that all the noise corrupting all the pixels from a given digital image follows the same statistical distribution, modelled in (6.4), with the same distribution parameters, the improved signal dependent noise model

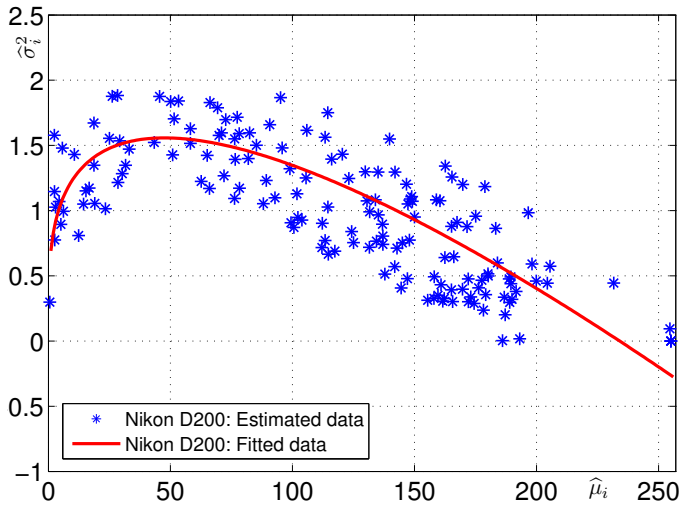


Figure 6.2: Scatter-plot of pixels' estimated expectation $\widehat{\mu}_i$ and estimated variance $\widehat{\sigma}_i^2$ in k-th block from several JPEG format images of *Nikon D200*.

proposes to take into account the non-uniformity of pixel's response by assuming that the parameters $(\widetilde{a}_k, \widetilde{b}_k)$ change for each pixel, or for each block of pixels for feasibility. This model is thus given by:

$$x_i \approx \mu_i + \frac{1}{\gamma}\mu_i^{1-\gamma}\eta_i \sim \mathcal{N}(\mu_i, \sigma_i^2) \tag{6.7}$$

$$\text{with } \sigma_i^2 = \frac{1}{\gamma^2}\mu_i^{2-2\gamma}(\widetilde{a}_k\mu_i^{\gamma} + \widetilde{b}_k) + \frac{\Delta^2}{12} \tag{6.8}$$

where $k \in \{1, \cdots, K\}$ denotes the block index and block fingerprints $(\widetilde{a}_k, \widetilde{b}_k)$ are estimated by using the same block, at the same location, from several JPEG images. Figure 6.2 demonstrates on an example the accuracy of the proposed model and the relevance of the block fingerprints by showing the non-linear relationship (6.8) between pixel's expectation and variance in k-th block.

Let us describe the specific process of extracting the device block fingerprints from images as follows:

1. Let $\mathbf{X} = \{x_{i,j}\}$ be a single matrix representing a JPEG image of size $I \times J$, $i \in \{1, \cdots, I\}$, $j \in \{1, \cdots, J\}$ and let $\mathbf{X}^{(N)} = (\mathbf{X}_1, \dots \mathbf{X}_N)$ be a set of N images. By using the BM3D denoising filter [162, 185], each image $\mathbf{X}_n$, $n \in \{1, \cdots, N\}$ is first decomposed into two components: $\mathbf{X}_n^{\text{app}}$ an estimation of pixel's expectation $\mu_{i,j}$ and $\mathbf{X}_n^{\text{res}}$ the residual noise, that is: $\mathbf{X}_n^{\text{res}} = \mathbf{X}_n - \mathbf{X}_n^{\text{app}}$.

2. Due to the block artifact of JPEG image, it is proposed to decompose $\mathbf{X}_n^{\text{app}}$ into blocks of size 8×8 pixels, denoted $\{\mathbf{x}_{n,k}^{\text{app}}\}$ where $\mathbf{x}_{n,k}^{\text{app}} = \{x_{n,k,l}^{\text{app}}\}$, $k \in \{1, \cdots, K\}$, $l \in \{1, \cdots, 64\}$ with $K \approx \frac{I \times J}{64}$.

3. Similarly, we decompose $\mathbf{X}_n^{\text{Res}}$ into blocks of size 8×8 pixels, denoted $\{\mathbf{x}_{n,k}^{\text{Res}}\}$ where $\mathbf{x}_{n,k}^{\text{Res}} = \{x_{n,k,l}^{\text{Res}}\}$.

4. By computing the standard deviation of $\{\mathbf{x}_{n,k}^{\text{app}}\}$, it is proposed to set the threshold $\tau = 2$ to exclude non-homogeneous blocks in each image. In fact, those blocks correspond to highly textured areas for which estimations of pixel's expectation and of block variance are likely to be inaccurate.

5. For the k-th block of all N images, it is proposed to utilize the selected N_k with $N_k \leq N$ homogeneous blocks to estimate the expectation of $\mathbf{x}_{n,k}^{\text{app}}$ and the variance of $\mathbf{x}_{n,k}^{\text{res}}$.

6. Finally, based on the estimations of pixel's expectation and noise variance, the parameters of the proposed improved signal-dependent noise model $(\widetilde{a}_k, \widetilde{b}_k)$ are jointly estimated using a Least Square algorithm (LS) (see details in Appendix B.1). Those parameters are later used as fingerprints for the camera device identification problem cast with hypothesis testing theory, see Sections 6.3-6.5.

Figure 6.1 illustrates the comparison of the camera fingerprints extracted by using the algorithm of [21] and our proposed device block fingerprints. In Figure 6.1a, each scatter point denotes parameters $(\widetilde{a}, \widetilde{b})$ estimated from a single image using the method proposed in [21]. By contrast, the proposed improved signal-dependent noise model parameters whose estimated values $(\widetilde{a}_k, \widetilde{b}_k)$ are presented in Figure 6.1b and used within the statistical test proposed in this chapter as camera device fingerprints. Obviously, the proposed improved signal-dependent noise model is more relevant for extracting camera device fingerprints. This may be explained by the fact that this novel noise model takes into account the non-uniformity of pixels and hence estimates the proposed fingerprint by block; this leads to an increasing number of parameters for camera device fingerprints and, hence, helps distinguishing two camera devices of the same model.

6.2.3 Exposing Linear Relationship of Block Fingerprints

In fact, even though block fingerprints proposed in this chapter, that are parameters $(\widetilde{a}_k, \widetilde{b}_k)$, allows the distinguishing of camera devices much more accurately than those proposed in [21], it is still very difficult to directly establish the optimal statistical test based on the improved signal-dependent noise model. This can be explained by the following main reasons: the variance of each image block is very difficult to estimate and, hence, can not always perfectly satisfies the improved signal-dependent noise model (6.7). Thus, we propose another solution for detection.

Based on the relationship between camera parameters (a, b) proposed in the the Poissonian-Gaussian noise model [22, 172]:

$$b = \omega^2 a^2 \tag{6.9}$$

where ω^2 is a depending parameter of the camera device, it is proposed to approximate Equation (6.9) due to the low variations of a by using the first order Taylor

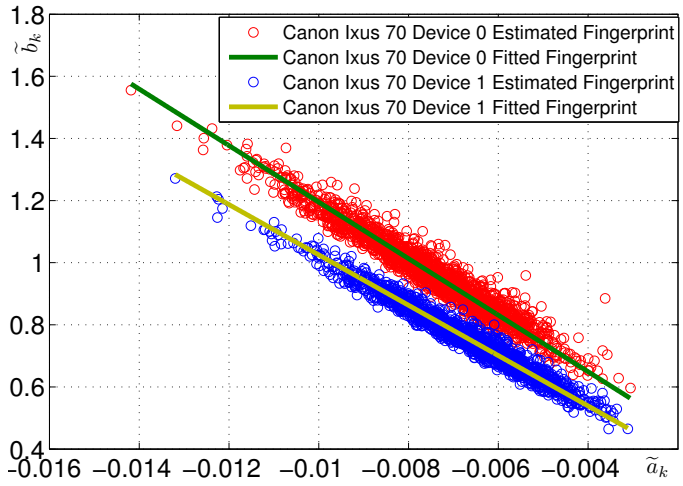


Figure 6.3: Illustration of linear relationship between the parameters $(\widetilde{a}_k, \widetilde{b}_k)$ which are used as fingerprint for camera device identification; the results have been obtained following the process described in Section 6.2.2 fitted by using Least Square algorithm (LS).

series at $\bar{a}$:

$$\begin{aligned} b &\approx \omega^2\bar{a}^2 + 2\omega^2\bar{a}(a-\bar{a}) \\ &= 2\omega^2\bar{a}a - \omega^2\bar{a}^2 \end{aligned} \tag{6.10}$$

where $\bar{a}$ denotes the mean of a. Then, for clarity, let us rewrite Equation (6.10) as:

$$a \approx cb + d \tag{6.11}$$

where $c = {}^1\!/_{2\omega^2\bar{a}}$ and $d = {}^{\bar{a}}\!/_2$ approximately link the linear relationship between parameters (a, b) from the noise model (6.1) of RAW data. The white balancing, demosaicing and gamma correction operations do not change this linear relationship (see Equations (6.2) to (6.8)). Thus, the relationship is still valid for JPEG images. Figure 6.3 illustrates the linear relationship between the parameters $(\widetilde{a}_k, \widetilde{b}_k)$ for two *Canon Ixus70* devices. Based on the linear relationship of the parameters, let us replace $\widetilde{a}_k$ by $c\widetilde{b}_k + d$ and, hence, use $(c\widetilde{b}_k + d, \widetilde{b}_k)$ as the parameters that are used as fingerprints for camera device identification. As Figure 6.3 illustrates, for images captured with the same device, the estimated parameters are well modelled using the proposed linear relation between $\widetilde{a}_k$ and $\widetilde{b}_k$, which tends to confirm our assumption. Thus, let us define the estimate error by:

$$e_k = \widetilde{a}_k - (c\widetilde{b}_k + d). \tag{6.12}$$

where e_k denotes the estimate error in k-th block, c and d the linear model parameters which can be estimated by using LS algorithm. It should be noted that c and d are constant for the whole image. Then, the problem of identifying source individual device from JPEG images is solved by using the estimate error between

camera fingerprint $\widetilde{a}_k$ and its linear relation with $\widetilde{b}_k$. Because numerous sources of noise corrupt a digital image during its acquisition, and because within the proposed approach many parameters have to be estimated, and that each of those estimations generates errors, we can model accurately the statistical distribution of the error e_k, see (6.12), as the realization of a Gaussian random variable.

By using this statistical model for the error e_k, together with hypothesis testing theory, we first establish the optimal statistical test for camera device identification and study its statistical performance. One can note that this optimal statistical test, namely the Likelihood Ratio Test (LRT), assumes that all distribution parameters are known, which are the framework adopted first in Section 6.3. Then, for a practical application, it is proposed to investigate two GLRTs (Generalized Likelihood Ratio Test). In the first GLRT, it is proposed to study the case when the linear model parameters namely, (c, d), of two devices and the camera fingerprints namely, $(\widetilde{a}_k, \widetilde{b}_k)$, of two devices are known, but the statistical distribution parameters of the error e_k are unknown. In such a case Section 6.4 presents the proposed GLRT and especially studies the impact of distribution parameters on the statistical performance of this GLRT. In the second GLRT, presented in Section 6.5, it is assumed that, under hypothesis $\mathcal{H}_1$, neither the linear model parameters (c_1, d_1) nor the camera fingerprint parameters $(\widetilde{a}_{k,1}, \widetilde{b}_{k,1})$ are known, the proposed GLRT deals with this practical case.

6.3 Likelihood Ratio Test for Two Simple Hypotheses Based on JPEG Format

6.3.1 Problem Statement

This section aims at presenting the optimal LRT and, more important, at studying its statistical performance; this statistical test is based on the error e_k, see (6.12) which follows the Gaussian distribution. Each camera device j, $j \in \{0, 1\}$ is characterized by its linear parametric model with parameters (c_j, d_j), which are obtained from the estimated camera device fingerprints $(\widetilde{a}_{k,j}, \widetilde{b}_{k,j})$. In fact, when testing an inspected image $\mathbf{X}$, two sets of errors e_k are known, that are $e_{k,0}$ based on parameters (c_0, d_0); $e_{k,1}$ based on parameters (c_1, d_1). Hence, for simplicity, let us define $\mathbf{e}_k = (e_{k,0}, e_{k,1})$. Within the framework of the LRT, hence, the problem of identifying source camera device is reduced to a choice between the two following simple hypotheses:

$$\begin{cases} \mathcal{H}_0 : \left\{ e_{k,0} = \widetilde{a}_k - (c_0 \widetilde{b}_k + d_0) \sim \mathcal{N}(0, \sigma_0^2) \right\} \\ \mathcal{H}_1 : \left\{ e_{k,1} = \widetilde{a}_k - (c_1 \widetilde{b}_k + d_1) \sim \mathcal{N}(0, \sigma_1^2) \right\}, \end{cases} \tag{6.13}$$
$$\text{where } \forall k = (1, ..., K), (c_1, d_1) \neq (c_0, d_0), \sigma_0^2 \neq \sigma_1^2.$$

Formally, a statistical test is a mapping $\delta : \mathbf{X} \to \delta(x)$, such that hypothesis $\mathcal{H}_j$ is accepted if and only $\delta(\mathbf{X}) = \mathcal{H}_j$. For solving this statistical detection problem such as (6.13), it follows from the Neyman-Pearson lemma [20, Theorem 3.2.1] that

the LRT is optimal in the sense described below. For definition, let

$$\mathcal{K}_{\alpha_\tau} = \{\delta : \mathbb{P}_0[\delta(\mathbf{X}) = \mathcal{H}_1] \leq \alpha_\tau\} \tag{6.14}$$

be the class of tests, solving problem (6.13), with an upper-bounded False Alarm Rate (FAR) α_τ. Here $\mathbb{P}_j[\cdot]$ is the probability under $\mathcal{H}_j, j \in \{0, 1\}$. Among all tests in $\mathcal{K}_{\alpha_\tau}$ the LRT is the most powerful test, which maximizes the detection power defined as:

$$\beta_\delta = \mathbb{P}_1[\delta(\mathbb{E}) = \mathcal{H}_1]. \tag{6.15}$$

In the following subsection, the LRT is first described in details and then its statistical performance is analytically established.

6.3.2 Optimal Detection Framework

For testing simple hypotheses the Neyman-Pearson Lemma [127, theorem 3.2.1] states that the LRT is the most powerful test in $\mathcal{K}_{\alpha_\tau}$ (7.23). Because of the statistical independence of pixels, and so is error $\mathbf{e}_k$, the LRT for camera device identification can be written:

$$\delta^{\text{lr}}(\mathbf{X}) = \begin{cases} \mathcal{H}_0 \ \text{ if } \Lambda^{\text{lr}}(\mathbf{E}) = \displaystyle\sum_{k=1}^{K} \Lambda^{\text{lr}}(\mathbf{e}_k) < \tau^{\text{lr}}, \\ \mathcal{H}_1 \ \text{ if } \Lambda^{\text{lr}}(\mathbf{E}) = \displaystyle\sum_{k=1}^{K} \Lambda^{\text{lr}}(\mathbf{e}_k) \geq \tau^{\text{lr}}, \end{cases} \tag{6.16}$$

where the decision threshold τ^{lr} is the solution of Equation $\mathbb{P}_{\mathcal{H}_0}\left[\Lambda^{\text{lr}}(\mathbf{E}) \geq \tau^{\text{lr}}\right] = \alpha_\tau$, to ensure that the FAR of the LRT equals α_τ. Based on the Gaussian distribution from Equation (6.13), it is proposed to define statistical parameters $\boldsymbol{\theta}_0 = (0, \sigma_0^2)$ and $\boldsymbol{\theta}_1 = (0, \sigma_1^2)$ under hypothesis $\mathcal{H}_0$ and $\mathcal{H}_1$. Then the probability density function (pdf) is given as: $\mathcal{P}_{\boldsymbol{\theta}_0}$ and $\mathcal{P}_{\boldsymbol{\theta}_1}$. Thus one can obtain the log Likelihood Ratio (LR) for one observation given by:

$$\Lambda^{\text{lr}}(\mathbf{e}_k) = \log \frac{\mathcal{P}_{\boldsymbol{\theta}_1}[e_{k,1}]}{\mathcal{P}_{\boldsymbol{\theta}_0}[e_{k,0}]}. \tag{6.17}$$

From the definition of (6.13), it is easy to rewrite the LR (6.17) as:

$$\Lambda^{\text{lr}}(\mathbf{e}_k) = \log\left(\frac{\sigma_0}{\sigma_1}\right) + \frac{1}{2}\left(\frac{e_{k,0}^2}{\sigma_0^2} - \frac{e_{k,1}^2}{\sigma_1^2}\right), \tag{6.18}$$

where the error $e_{k,j} = \widetilde{a}_k - (c_j\widetilde{b}_k + d_j)$, $j \in \{0, 1\}$ and block index $k \in \{1, \cdots, K\}$.

6.3.3 Statistical Performance of LRT

The study of the proposed LRT is made easier thanks to some asymptotic theorems, which are relevant in this context as the number of blocks is chosen very large

(each block is made of a small number of pixels). Let us denote $E_{\mathcal{H}_j}(\Lambda^{\text{lr}}(\mathbf{e}_k))$ and $V_{\mathcal{H}_j}(\Lambda^{\text{lr}}(\mathbf{e}_k))$ the expectation and the variance of the LR $\Lambda^{\text{lr}}(\mathbf{e}_k)$ under hypothesis $\mathcal{H}_j, j = \{0, 1\}$. Note that those moments can be analytically calculated, the details are provided in Appendix B.2. The Lindeberg's central limit theorem (CLT) [127, theorem 11.2.5] states that as the number of blocks K tends to infinity it holds true that[3]:

$$\frac{\sum_{k=1}^{K}\left(\Lambda^{\text{lr}}(\mathbf{e}_k) - E_{\mathcal{H}_j}(\Lambda^{\text{lr}}(\mathbf{e}_k))\right)}{\left(\sum_{k=1}^{K} V_{\mathcal{H}_j}(\Lambda^{\text{lr}}(\mathbf{e}_k))\right)^{1/2}} \xrightarrow{d} \mathcal{N}(0,1)\ ,\ j = \{0,1\}\,, \tag{6.19}$$

where $\xrightarrow{d}$ represents the convergence in distribution and $\mathcal{N}(0,1)$ is the standard normal distribution with zero mean and unit variance. This theorem is of crucial interest to establish the statistical properties of the proposed test [129,134,135,142,177]. In fact, once the two first moments of the LR have been calculated analytically under hypothesis $\mathcal{H}_0$, which again is detailed in Appendix B.2, one can normalize under hypothesis $\mathcal{H}_0$ the LR $\Lambda^{\text{lr}}(\mathbf{E})$ as follows:

$$\overline{\Lambda}^{\text{lr}}(\mathbf{E}) = \frac{\Lambda^{\text{lr}}(\mathbf{E}) - \sum_{k=1}^{K} E_{\mathcal{H}_0}(\Lambda^{\text{lr}}(\mathbf{e}_k))}{\left(\sum_{k=1}^{K} V_{\mathcal{H}_0}(\Lambda^{\text{lr}}(\mathbf{e}_k))\right)^{1/2}}.$$

It is thus straightforward to define the normalized LRT with $\overline{\Lambda}^{\text{lr}}(\mathbf{E})$ by:

$$\overline{\delta}^{\text{lr}} = \begin{cases} \mathcal{H}_0 \ \text{ if } \ \overline{\Lambda}^{\text{lr}}(\mathbf{E}) < \overline{\tau}^{\text{lr}} \\ \mathcal{H}_1 \ \text{ if } \ \overline{\Lambda}^{\text{lr}}(\mathbf{E}) \geq \overline{\tau}^{\text{lr}}. \end{cases} \tag{6.20}$$

Thus, let us establish the statistical properties of the LRT (6.20).

Proposition 6.1. *Assuming that for the camera device identification problem as case within the two simple hypotheses (6.13), in which both parameters (c_j, d_j) and σ_j^2 are known, then for any $\alpha_\tau \in (0,1)$ the decision threshold:*

$$\overline{\tau}^{\text{lr}} = \Phi^{-1}\left(1 - \alpha_\tau\right), \tag{6.21}$$

guarantees that the LRT is in the class $\mathcal{K}_{\alpha_\tau}$, see (7.23). Here Φ and Φ^{-1} respectively represent the cumulative distribution function (cdf) of the standard normal distribution and its inverse.

Proposition 6.2. *Assuming that for the camera device identification problem as case within the two simple hypotheses (6.13), in which both parameters (c_j, d_j) and*

[3] Note that we refer to the Lindeberg's CLT, whose conditions are easily verified in our case as the pixels have a bounded value.

σ_j^2 are known, for any decision threshold $\overline{\tau}^{\mathrm{lr}}$, the power function associated with the proposed test $\overline{\delta}^{\mathrm{lr}}$ (6.20) is given by:

$$\beta_{\overline{\delta}^{\mathrm{lr}}} = 1 - \Phi\left(\sqrt{\frac{v_0}{v_1}}\Phi^{-1}\left(1-\alpha_\tau\right) + \frac{m_0 - m_1}{\sqrt{v_1}}\right). \tag{6.22}$$

where

$$m_j = \sum_{k=1}^{K} E_{\mathcal{H}_j}(\Lambda^{\mathrm{lr}}(\mathbf{e}_k)) \tag{6.23}$$

$$v_j = \sum_{k=1}^{K} V_{\mathcal{H}_j}(\Lambda^{\mathrm{lr}}(\mathbf{e}_k)), j = \{0,1\}. \tag{6.24}$$

Equations (6.21) and (6.22) emphasize the main advantage of normalizing the LR as described in relation (6.20): it allows to set any of threshold that guarantees a FAR independently from any distribution parameter. One can also note that, for practical application, it is considered in the present chapter that a given image under investigation was captured with device 0. Hence the case "false alarm" occurs when an alarm is raised because the given image is authenticated as captured with another device 1. On the opposite, the "correct detection" which defines the power function, corresponds to the probability of accurately identifying an image captured from camera device 1.

6.4 GLRT for JPEG Format with Knowing Camera Fingerprints and Linear Model Parameters

While the previous section studies the optimal LRT assuming that all the statistical parameters are known, this section proposed a less restrictive GLRT. It is indeed assumed in the present section that statistical parameters of the Gaussian distribution, see (6.13), are unknown. Hence it is proposed in this section to estimate those parameters and then, to design and study statically the GLRT which replaces the Gaussian distribution parameters with those estimates. Similar to the prior LRT, let us defines the GLRT $\widehat{\delta}_1^{\mathrm{lr}}$ as follows:

$$\widehat{\delta}_1^{\mathrm{lr}} = \begin{cases} \mathcal{H}_0 \text{ if } \widehat{\Lambda}_1^{\mathrm{lr}}(\mathbf{E}) = \sum_{k=1}^{K} \widehat{\Lambda}_1^{\mathrm{lr}}(\mathbf{e}_k) < \widehat{\tau}_1^{\mathrm{lr}}, \\ \mathcal{H}_1 \text{ if } \widehat{\Lambda}_1^{\mathrm{lr}}(\mathbf{E}) = \sum_{k=1}^{K} \widehat{\Lambda}_1^{\mathrm{lr}}(\mathbf{e}_k) \geq \widehat{\tau}_1^{\mathrm{lr}}. \end{cases} \tag{6.25}$$

$\widehat{\Lambda}_1^{\mathrm{lr}}(\mathbf{e}_k)$ ensures that the proposed GLRT $\widehat{\delta}_1^{\mathrm{lr}}$ is in the class $\mathcal{K}_{\alpha_\tau}$. The associated decision threshold $\widehat{\tau}_1^{\mathrm{lr}}$ is defined as the solution of the following Equation:

$$\mathbb{P}_0[\widehat{\Lambda}_1^{\mathrm{lr}}(\mathbf{E}) \geq \widehat{\tau}_1^{\mathrm{lr}}] = \alpha_\tau. \tag{6.26}$$

Again, here α_τ is the prescribed probability of false alarm.

Again, the Generalized Likelihood Ratio (GLR) essentially consists in replacing the unknown parameters by the Maximum Likelihood (ML) estimation; hence, it follows from the definition of the LR (6.18), that the GLR is defined as:

$$\begin{aligned}\widehat{\Lambda}_1^{\mathrm{lr}}(\mathbf{e}_k) &= \log \frac{\mathcal{P}_{\widehat{\boldsymbol{\theta}}_1}[e_{k,1}]}{\mathcal{P}_{\widehat{\boldsymbol{\theta}}_0}[e_{k,0}]} \\ &= \log\left(\frac{\widehat{\sigma}_0}{\widehat{\sigma}_1}\right) + \frac{1}{2}\left(\frac{e_{k,0}^2}{\widehat{\sigma}_0^2} - \frac{e_{k,1}^2}{\widehat{\sigma}_1^2}\right),\end{aligned} \tag{6.27}$$

where $e_{k,j} = \widetilde{a}_k - (c_j\widetilde{b}_k + d_j)$, with $j \in \{0,1\}$, represents the estimate error and $(\widetilde{a}_k, \widetilde{b}_k)$ stand for the device fingerprint estimated from the inspected image $\mathbf{X}$. $\widehat{\boldsymbol{\theta}}_0$ and $\widehat{\boldsymbol{\theta}}_1$ denote respectively the estimates of statistical parameters $\boldsymbol{\theta}_0$ and $\boldsymbol{\theta}_1$. It should be noted that in this GLR, the linear model parameters (c_j, d_j) and camera device fingerprints $(\widetilde{a}_{k,j}, \widetilde{b}_{k,j})$ are known. Our aim is to estimate the Gaussian distribution parameter σ_j^2. Using ML estimation, the variance is estimated by:

$$\widehat{\sigma}_j^2 = \frac{1}{K-1}\sum_{k=1}^{K}(\widetilde{a}_{k,j} - c_j\widetilde{b}_{k,j} - d_j)^2. \tag{6.28}$$

By invoking again the Lindeberg's CLT [127, theorem 11.2.5] under hypothesis $\mathcal{H}_j$, $j \in \{0,1\}$, immediately the statistical distribution of $\widehat{\Lambda}_1^{\mathrm{lr}}(\mathbb{E})$ is given as:

$$\widehat{\Lambda}_1^{\mathrm{lr}}(\mathbb{E}) \xrightarrow{d} \mathcal{N}(m_j^{(1)}, v_j^{(1)}) \tag{6.29}$$

where the expectation $m_j^{(1)}$ and the variance $v_j^{(1)}$ of the GLR $\widehat{\Lambda}_1^{\mathrm{lr}}(\mathbb{E})$ are respectively defined by:

$$m_j^{(1)} = \sum_{k=1}^{K} E_{\mathcal{H}_j}(\widehat{\Lambda}_1^{\mathrm{lr}}(\mathbf{e}_k)) \tag{6.30}$$

$$v_j^{(1)} = \sum_{k=1}^{K} V_{\mathcal{H}_j}(\widehat{\Lambda}_1^{\mathrm{lr}}(\mathbf{e}_k)), \tag{6.31}$$

where the specific calculation of the expectation $E_{\mathcal{H}_j}(\widehat{\Lambda}_1^{\mathrm{lr}}(\mathbf{e}_k))$ and the variance $V_{\mathcal{H}_j}(\widehat{\Lambda}_1^{\mathrm{lr}}(\mathbf{e}_k))$ of $\widehat{\Lambda}_1^{\mathrm{lr}}(\mathbf{e}_k)$ is expanded in Appendix B.3. Let us straightforwardly define the normalized GLR as follows:

$$\widehat{\Lambda}_1^{\star}(\mathbb{E}) = \frac{\widehat{\Lambda}_1^{\mathrm{lr}}(\mathbb{E}) - m_j^{(1)}}{\sqrt{v_j^{(1)}}}. \tag{6.32}$$

It finally follows that the normalized GLRT, based on $\widehat{\Lambda}_1^{\star}(\mathbb{E})$, is given by:

$$\widehat{\delta}_1^{\star} = \begin{cases} \mathcal{H}_0 \text{ if } \widehat{\Lambda}_1^{\star}(\mathbb{E}) < \tau_1^{\star} \\ \mathcal{H}_1 \text{ if } \widehat{\Lambda}_1^{\star}(\mathbb{E}) \geq \tau_1^{\star}. \end{cases} \tag{6.33}$$

Hence, again, one can immediately establish the statistical properties of the GLRT (6.33).

Proposition 6.3. *Assuming that the pixels are modelled by the proposed improved signal-dependent noise model (6.8), when both linear model parameters (c_j, d_j) and camera device fingerprints $(\widetilde{a}_{k,j}, \widetilde{b}_{k,j})$ are known, and the unknown statistical parameter σ_j^2 is estimated as in (6.28), then any $\alpha_\tau \in (0,1)$ the decision threshold of the proposed GLRT $\widehat{\delta}_1^\star$ is given by:*

$$\tau_1^\star = \Phi^{-1}\left(1 - \alpha_\tau\right). \tag{6.34}$$

Proposition 6.4. *Assuming that the pixels are modelled by the proposed improved signal-dependent noise model (6.8), when both linear model parameters (c_j, d_j) and camera device fingerprints $(\widetilde{a}_{k,j}, \widetilde{b}_{k,j})$ are known, and the unknown statistical parameter σ_j^2 is estimated as in (6.28), for any decision threshold $\widehat{\tau}_1^\star$, the power function associated with test $\widehat{\delta}_1^\star$ (6.33) is given by*

$$\beta_{\widehat{\delta}_1^\star} = 1 - \Phi\left(\sqrt{\frac{v_0^{(1)}}{v_1^{(1)}}}\Phi^{-1}\left(1 - \alpha_\tau\right) + \frac{m_0^{(1)} - m_1^{(1)}}{\sqrt{v_1^{(1)}}}\right). \tag{6.35}$$

It is important to contact the optimal LRT presented in Section 6.3 that requires the knowledge of both noise model through camera device fingerprints $(\widetilde{a}_{k,j}, \widetilde{b}_{k,j})$ and the linear relation between those fingerprints characterized by (c_j, d_j) as well as the error $\mathbf{e}_k$ statistical parameter, that is the expectation and the variance under both hypotheses. On the opposite, the practical GLRT proposed in this Section 6.4 is essentially based on the estimation of noise model parameters, typically from a set of images from each device. Then the proposed GLRT consists of estimating the noise model parameters from given images, using those estimates to extract the error $\mathbf{e}_k$ and testing, for each block, whether the error is more likely distributed according to what is expected under each hypothesis $\mathcal{H}_0$ or $\mathcal{H}_1$. Besides, the impact of those estimations is taken into account in the calculation of the proposed GLRT statistical performance. In practice, before testing, we have known σ from BM3D denoising filter for decomposing an inspected image which controls the strength of denoising and γ which is the value of gamma correction.

6.5 GLRT for JPEG Format without Knowing Camera Fingerprints or Linear Model Parameters

While the GLRT proposed in Section 6.4 can be used in many practical applications in which one has to decide if a given image was acquired either with device 0 or with device 1, it is also possible to imagine some practical case that does not fit within this framework. This would be typically the case if one has only access to a given device 0; and hence he may be interested to test whether a given image has been acquired with this known device, or with any other unknown device, for which

the noise parameters are unknown. In such a practical context, the hypothesis $\mathcal{H}_1$ becomes composite. In this scenario, it is proposed to design a test which would allow us to identify whether or not the inspected JPEG image $\mathbf{X}$ is taken by the camera device 0. Hence, it is important to note that the given image $\mathbf{X}$ may have been acquired by any other unknown device characterized by noise parameters that may take any values. The proposed second GLRT, denoted as $\widehat{\delta}_2^{\text{lr}}$, if defined as follows:

$$\widehat{\delta}_2^{\text{lr}} = \begin{cases} \mathcal{H}_0 \text{ if } \widehat{\Lambda}_2^{\text{lr}}(\mathbf{E}) = \sum_{k=1}^{K} \widehat{\Lambda}_2^{\text{lr}}(\widehat{\mathbf{e}}_k) < \widehat{\tau}_2^{\text{lr}}, \\ \mathcal{H}_1 \text{ if } \widehat{\Lambda}_2^{\text{lr}}(\mathbf{E}) = \sum_{k=1}^{K} \widehat{\Lambda}_2^{\text{lr}}(\widehat{\mathbf{e}}_k) \geq \widehat{\tau}_2^{\text{lr}}, \end{cases} \tag{6.36}$$

where the decision threshold $\widehat{\tau}_2^{\text{lr}}$ is the solution of Equation

$$\mathbb{P}_0[\widehat{\Lambda}_2^{\text{lr}}(\mathbf{E}) \geq \widehat{\tau}_2^{lr}] = \alpha_\tau \tag{6.37}$$

to ensure that the GLRT is in the class $\mathcal{K}_{\alpha_\tau}$. The GLR $\widehat{\Lambda}_2^{\text{lr}}(\widehat{\mathbf{e}}_k)$ is defined by:

$$\widehat{\Lambda}_2^{\text{lr}}(\widehat{\mathbf{e}}_k) = \log \frac{\mathcal{P}_{\widehat{\boldsymbol{\theta}}_1}[\widehat{e}_{k,1}]}{\mathcal{P}_{\widehat{\boldsymbol{\theta}}_0}[e_{k,0}]} \tag{6.38}$$

$$= \log\left(\frac{\widehat{\sigma}_0}{\widehat{\sigma}_1}\right) + \frac{1}{2}\left(\frac{e_{k,0}^2}{\widehat{\sigma}_0^2} - \frac{\widehat{e}_{k,1}^2}{\widehat{\sigma}_1^2}\right), \tag{6.39}$$

where $\widehat{\mathbf{e}}_k$ consists of the estimate error $e_{k,0} = \widetilde{a}_k - c_0\widetilde{b}_k - d_0$ and $\widehat{e}_{k,1} = \widetilde{a}_k - \widehat{c}_1\widetilde{b}_k - \widehat{d}_1$. It should be noted that in this GLR, the linear model parameters (c_0, d_0) and camera device fingerprints $(\widetilde{a}_{k,0}, \widetilde{b}_{k,0})$ are known under hypothesis $\mathcal{H}_0$; (c_1, d_1) and $(\widetilde{a}_{k,1}, \widetilde{b}_{k,1})$ are unknown, which can be estimated by using several tested images. It is proposed to estimate parameters (c_1, d_1) by using LS algorithm, see Appendix B.1, and estimate $(\widetilde{a}_{k,1}, \widetilde{b}_{k,1})$ by using the proposed algorithm of block fingerprints estimation in subsection 6.2.2. Then the variance $\widehat{\sigma}_j^2$ can be estimated by using Equation (6.28).

Once again one can use Lindeberg's CLT to establish the statistical distribution of the proposed normalized $\widehat{\Lambda}_2^{\star}(\mathbb{E})$, defined follows as:

$$\widehat{\Lambda}_2^{\star}(\mathbb{E}) = \frac{\widehat{\Lambda}_2^{\text{lr}}(\mathbb{E}) - m_j^{(2)}}{\sqrt{v_j^{(2)}}} \xrightarrow{d} \mathcal{N}(0,1). \tag{6.40}$$

where the expectation $m_j^{(2)}$ and the variance $v_j^{(2)}$ of $\widehat{\Lambda}_2^{\text{lr}}(\mathbb{E})$ is defined by:

$$m_j^{(2)} = \sum_{k=1}^{K} E_{\mathcal{H}_j}(\widehat{\Lambda}_2^{\text{lr}}(\widehat{\mathbf{e}}_k)) \tag{6.41}$$

$$v_j^{(2)} = \sum_{k=1}^{K} V_{\mathcal{H}_j}(\widehat{\Lambda}_2^{\text{lr}}(\widehat{\mathbf{e}}_k)). \tag{6.42}$$

For clarity, the calculations of the expectation $E_{\mathcal{H}_j}(\widehat{\Lambda}_2^{\mathrm{lr}}(\widehat{\mathbf{e}}_k))$ and the variance $V_{\mathcal{H}_j}(\widehat{\Lambda}_2^{\mathrm{lr}}(\widehat{\mathbf{e}}_k))$ of $\widehat{\Lambda}_2^{\mathrm{lr}}(\widehat{\mathbf{e}}_k)$ are detailed in Appendix B.4.

Finally, it is proposed to define the normalized GLRT with $\widehat{\Lambda}_2^{\star}(\mathbb{E})$ by:

$$\widehat{\delta}_2^{\star} = \begin{cases} \mathcal{H}_0 \text{ if } \widehat{\Lambda}_2^{\star}(\mathbb{E}) < \tau_2^{\star} \\ \mathcal{H}_1 \text{ if } \widehat{\Lambda}_2^{\star}(\mathbb{E}) \geq \tau_2^{\star}. \end{cases} \tag{6.43}$$

Again, it is proposed to establish the statistical properties of the GLRT (6.43) for testing a known camera device against any other unknown device.

Proposition 6.5. *Assuming that the pixels are modelled by the proposed improved signal-dependent noise model (6.8), when parameters (c_0, d_0) and $(\widetilde{a}_{k,0}, \widetilde{b}_{k,0})$ for device 0 are known, parameters (c_1, d_1) and $(\widetilde{a}_{k,1}, \widetilde{b}_{k,1})$ for device 1 are not known, and the unknown parameter σ_j^2 is estimated as in (6.28), then any $\alpha_\tau \in (0,1)$ the decision threshold of the proposed GLRT $\widehat{\delta}_2^{\star}$ is given by:*

$$\tau_2^{\star} = \Phi^{-1}(1 - \alpha_\tau), \tag{6.44}$$

Proposition 6.6. *Assuming that the pixels are modelled by the proposed improved signal-dependent noise model (6.8), when parameters (c_0, d_0) and $(\widetilde{a}_{k,0}, \widetilde{b}_{k,0})$ for device 0 are known, parameters (c_1, d_1) and $(\widetilde{a}_{k,1}, \widetilde{b}_{k,1})$ for device 1 are not known, and the unknown parameter σ_j^2 is estimated as in (6.28), for any decision threshold $\widehat{\tau}_2^{\star}$, the power function associated with test $\widehat{\delta}_2^{\star}$ (6.43) is given by*

$$\beta_{\widehat{\delta}_2^{\star}} = 1 - \Phi\left(\sqrt{\frac{v_0^{(2)}}{v_1^{(2)}}}\Phi^{-1}(1 - \alpha_\tau) + \frac{m_0^{(2)} - m_1^{(2)}}{\sqrt{v_1^{(2)}}}\right). \tag{6.45}$$

The main advantages of the proposed GLRT $\widehat{\delta}_2^{\star}$ are 1) that its statistical performance can be analytically established and, hence, one can easily compute the decision threshold that guarantees a prescribed FAR along with the ensuing power function and 2) thanks to the normalization the decision threshold only depends on the prescribed FAR α_τ. In practice, the parameters σ and γ for camera device 0 are both known; on the contrary, those two parameters for device 1 are estimated by using the algorithms proposed in [183] before applying source camera identification.

6.6 Numerical Experiments

6.6.1 Results on Simulated Images for JPEG Format

One of the main contributions of this chapter is to show that hypothesis testing theory can be applied to design a statistical test with known statistical properties for identifying source camera device on simulated images.

To verify the sharpness of the theoretically established results, it is first proposed to use the Monte-Carlo simulation on a generated dataset. The camera device 0 and 1 are respectively characterized by $(c_0, d_0) = (-0.0142, 0.0015)$

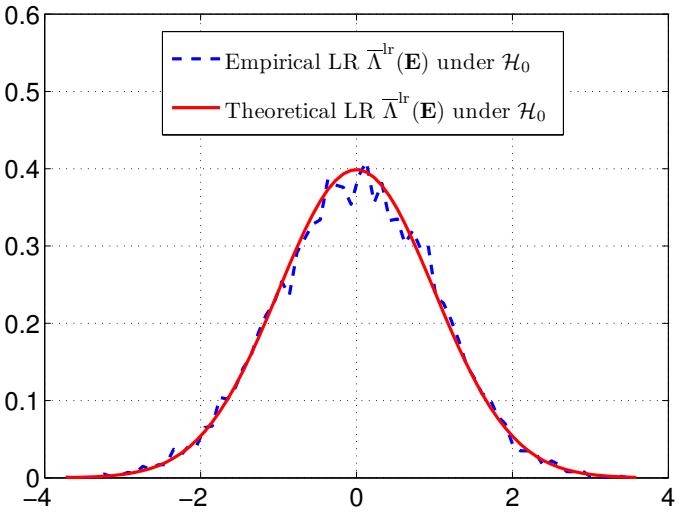
and $(c_1, d_1) = (-0.0132, 0.0018)$. Those values correspond to the estimated parameters for two different devices of camera model *Nikon D200* from the Dresden dataset [1]. With those settings the statistical parameters $(\mu_0, \sigma_0) = (0, 1.46 \times 10^{-8})$ and $(\mu_1, \sigma_1) = (0, 1.98 \times 10^{-8})$ of the Gaussian distribution are also known using estimates from Dresden dataset images. All the required parameters have been detailed, thus, one can now build two sets of random variables by repeating 10000 simulation with those distribution parameters. Then, the first set of 10000 simulated images from camera device 0 consisting of 5000 realization of random variables $(\widetilde{a}_0, \widetilde{b}_0)$ for each image; the second set of 10000 simulated images from device 1 consisting of 5000 realization of random variables $(\widetilde{a}_1, \widetilde{b}_1)$ for each image. Under hypothesis $\mathcal{H}_0$ and $\mathcal{H}_1$ respectively, Figure 6.4 presents the comparison between empirical and theoretical distribution of the optimal LR $\overline{\Lambda}^{\text{lr}}(\mathbf{E})$. Under hypothesis $\mathcal{H}_0$, the empirical distribution of the LR $\overline{\Lambda}^{\text{lr}}(\mathbf{E})$ approximately follows the standard normal distribution with zero mean and unit variance, which directly verifies the accuracy of the theoretically established statistical performance for the proposed LRT (6.20). Similarly, under hypothesis $\mathcal{H}_1$, the empirical and theoretical distribution of the optimal LR $\overline{\Lambda}^{\text{lr}}(\mathbf{E})$ are nearly overlapped, which also verifies the correctness of the established statistical performance.

Another contribution of the present chapter lies in the design of a statistical test that can warrant the prescribed FAR. Thus, it is proposed to compare the empirical and theoretical FAR α_τ of the optimal LR $\overline{\Lambda}^{\text{lr}}(\mathbf{E})$ as a function of the decision threshold $\overline{\tau}^{\text{lr}}$. The results of this comparison are presented in Figure 6.5. This figure emphasizes that the proposed LRT (6.20) has in practice the ability of guaranteeing the prescribed FAR. In some cases ($\overline{\tau}^{\text{lr}} \geq 3$), it should be noted that the slight differences of two cures are due to the inaccuracy of the CLT which can hardly model the tails of the distribution with accuracy.

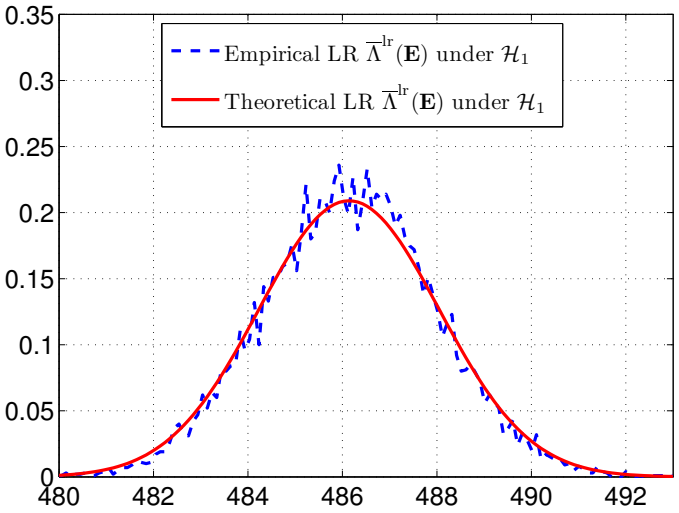
6.6.2 Results on Real Images for JPEG Format

Another major contribution of this chapter is to design the detectors with estimated parameters to identify source camera device in the practical cases. Therefore, it is proposed to verify on numerical experiments, using a real image dataset, the accuracy of both the GLRT $\widehat{\delta}_1^\star$ (6.33), that aims at distinguishing between two camera devices whose parameters are known, as well as the accuracy of the second GLRT $\widehat{\delta}_2^\star$ (6.43), whose goal is to distinguish between a known camera device and any other unknown device with unknown parameters.

To this end, it is proposed in this chapter to exploit the reference images from Dresden dataset [1] for showing the accuracy of the proposed GLR tests as well as for comparison with the prior-art detectors. Note that the images from Dresden dataset are colored image and that in our tests, only the red color channel is selected. Besides, all the images are acquired with the highest available JPEG quality setting and maximum available resolution. Table 6.1 gives the specific number of each device in our experiments. It also displays the comparison of the estimated parameters where σ denotes the parameter of BM3D denoising filter [162, 185] which, roughly



(a) Under hypothesis $\mathcal{H}_0$ between the empirical distribution of the LR $\overline{\Lambda}^{\text{lr}}(\mathbf{E})$ and its theoretically established distribution (here Gaussian distribution with zero mean and unit variance).



(b) Under hypothesis $\mathcal{H}_1$ between the empirical distribution of the LR $\overline{\Lambda}^{\text{lr}}(\mathbf{E})$ and its theoretically established distribution (here Gaussian distribution with expectation $\frac{m_1-m_0}{\sqrt{v_0}}$ and variance $\frac{v_1}{v_0}$, see subsection 6.3.3).

Figure 6.4: Comparison between empirical and theoretical distribution of $\overline{\Lambda}^{\text{lr}}(\mathbf{E})$.

speaking, determines denoising strength, and the presented value of parameter γ represents the value of gamma correction mean parameter in the post-processing of digital imaging. It should be noted that in this chapter, these two parameters are not the device fingerprint, and that they can be estimated accurately, for instance, by using the algorithm proposed in [183] before applying the identification step. In

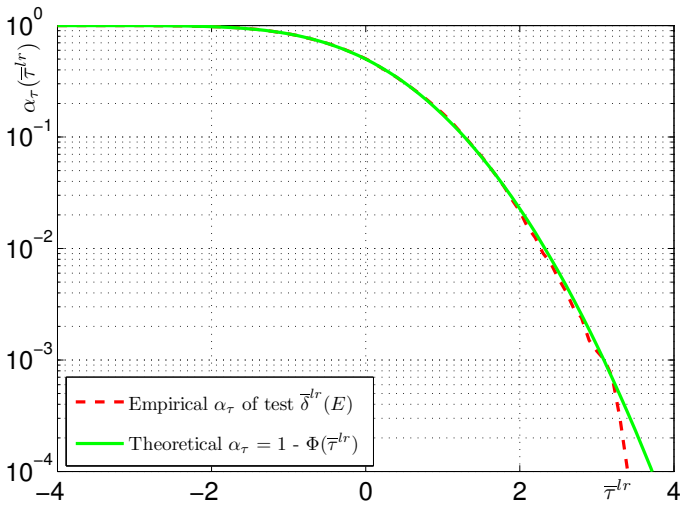


Figure 6.5: Comparison between the theoretical FAR α_τ and the empirical results, plotted as a function of the threshold $\overline{\tau}^{\text{lr}}$.

Table 6.1: Images statistic from Dresden dataset with mean estimated parameters (std: standard deviation)

Camera Device	Alias	σ	γ	Resolution	No. images
Nikon D200 # 0	N_D200_0	3.4000	0.8464	3872×2592	372
Nikon D200 # 1	N_D200_1	3.3294	0.8591	3872×2592	380
Nikon D70 # 0	N_D70_0	2.2365	0.7494	3008×2000	180
Nikon D70 # 1	N_D70_1	2.1823	0.7024	3008×2000	189
Nikon D70s # 0	N_D70s_0	2.2322	0.7001	3008×2000	178
Nikon D70s # 1	N_D70s_1	2.0998	0.7506	3008×2000	189
Canon Ixus70 # 0	C_I70_0	4.3573	0.8596	3072×2304	187
Canon Ixus70 # 1	C_I70_1	4.0456	0.8425	3072×2304	194
Canon Ixus70 # 2	C_I70_2	4.5535	0.8819	3072×2304	186
$\sum$ 11	$\sum$ 11	std: 1.0025	std: 0.0726	$\sum$ 3	$\sum$ 2055

other words, parameters (σ, γ) are assumed to be known in our proposed practical test (6.33) and unknown in the test (6.43).

For the practical application of the proposed GLRT, the dataset available for each image is split into two subsets: one "*Learning Subset*" and one "*Testing Subset*"; those subsets are disjoint as none of images is used in both subsets. Images of "*Learning Subset*" are used to extract device fingerprints from each device; images of "*Testing Subset*" are used to identify the origin of a given image. The number of "*Learning Subset*" is set as 100 where the images are randomly selected with high quality factor (QF $\geq$ 70). Besides, unlike the fingerprints of RAW images [22, 25], we have found that the ISO sensitivity has no impact on the proposed fingerprints and, hence, on the camera device identification methodology.

Experiments are realized on a large dataset to verify the sharpness of the proposed algorithms. It should be noted that in the proposed tests GLRT $\widehat{\delta}_1^\star$ and $\widehat{\delta}_2^\star$,

Table 6.2: Detection performance comparison of the test $\widehat{\delta}_2^{\star}$ at the given FAR $\alpha_\tau = 0.02$ with increasing the number n of tested images for each group, assuming that two devices 0 and 1 from the same camera model.

Camera Device	n = 10	n = 20	n = 30	n = 40
N_D200_0 vs N_D200_1	0.10	0.95	0.97	1.00
N_D70_0 vs N_D70_1	0.05	1.00	1.00	1.00
N_D70s_0 vs N_D70s_1	0.12	0.98	1.00	1.00
C_I70_0 vs C_I70_1	0.00	0.98	1.00	1.00
C_I70_0 vs C_I70_2	0.00	0.96	1.00	1.00
C_I70_1 vs C_I70_2	0.00	1.00	1.00	1.00
Average	0.05	0.98	1.00	1.00

we use multiple JPEG images together to identify the camera device. Then it is proposed to divide the "*Testing Subset*" into several overlapping groups of n images for each group, here of course $n < N$ with N the total number of images in the "*Testing Subset*". If the value of n is set very small, we could lose the accuracy of the estimation, since our parameters estimation is based on LS algorithm which needs the large number of data (see Appendix B.1). To select the optimal number of each group, let us do the test with increasing the number n (see Table 6.2). Based on the detection power of the proposed test, therefore, it is proposed to set $n = 40$ in our test GLRT $\widehat{\delta}_1^{\star}$ and $\widehat{\delta}_2^{\star}$, which can guarantee the high detection power while decreasing the probability of missed detection.

First, let us highlight the relevance of the proposed GLRT $\widehat{\delta}_1^{\star}$. In this scenario, the camera fingerprints $(\widetilde{a}_{k,j}, \widetilde{b}_{k,j})$ are known (or estimated) as well as the linear model parameters (c_j, d_j), $j \in \{0, 1\}$ for all devices from the same camera model. Our goal is to identify the inspected images captured with device 0, corresponding here to hypothesis $\mathcal{H}_0$, or with camera device 1, corresponding to hypothesis $\mathcal{H}_1$. In [21], *Test one* is designed for identifying the inspected image from camera device 0 or 1, that deals with the same case as the proposed GLRT $\widehat{\delta}_1^{\star}$. Thus, it is reasonable that let us compare the performance of our proposed test $\widehat{\delta}_1^{\star}$ with *Test one* of [21]. As Table 6.3 illustrates, our proposed method obviously outperforms *Test one*. Based on the generalized signal-dependent noise model in [21], *Test one* can only identify two camera devices from different models. By contrast, our proposed test $\widehat{\delta}_1^{\star}$ can identify source camera device of the same model.

Next, let us study the detection performance of the test $\widehat{\delta}_2^{\star}$. In this scenario, we can not obtain the camera fingerprints $(\widetilde{a}_{k,1}, \widetilde{b}_{k,1})$ and linear model parameters (c_1, d_1), $j \in \{0, 1\}$ for device 1, which are estimated from the group of "*Testing Subset*". The goal of the test $\widehat{\delta}_2^{\star}$ is to identify the inspected JPEG images acquired by device 0 or any other one. In [21], *Test two* is designed for identifying the inspected image from camera device 0 or any other one as well. Besides, the state-of-the-art

Table 6.3: Detection performance comparison at the given FAR $\alpha_\tau = 0.01$, assuming that two devices 0 and 1 from the same camera model. This test is corresponding to the case of the GLRT (6.33).

Camera Device	Proposed Test $\widehat{\delta}_1^\star$	*Test one* of [21]
N_D200_0 vs N_D200_1	1.00	0.02
N_D70_0 vs N_D70_1	1.00	0.01
N_D70s_0 vs N_D70s_1	1.00	0.03
C_I70_0 vs C_I70_1	1.00	0.16
C_I70_0 vs C_I70_2	1.00	0.03
C_I70_1 vs C_I70_2	1.00	0.07
Average	1.00	0.05

Table 6.4: Detection performance comparison at the given FAR $\alpha_\tau = 0.01$, assuming that two devices 0 and 1 from the same camera model. This test is corresponding to the case of the GLRT (6.43).

Camera Device	Proposed Test $\widehat{\delta}_2^\star$	*Test two* of [21]	Test [48]
N_D200_0 vs N_D200_1	1.00	0.01	1.00
N_D70_0 vs N_D70_1	1.00	0.01	1.00
N_D70s_0 vs N_D70s_1	1.00	0.01	1.00
C_I70_0 vs C_I70_1	1.00	0.02	1.00
C_I70_0 vs C_I70_2	1.00	0.03	1.00
C_I70_1 vs C_I70_2	1.00	0.07	1.00
Average	1.00	0.03	1.00

detector of [48], based on the reference PRNU noise as camera fingerprints, can also deal with this scenario. Then let us compare the detection perform of those three detectors for identifying source camera device from the same model. As Table 6.4 demonstrates, the test $\widehat{\delta}_2^\star$ is approximately perfect and exhibits roughly the detection power as the state-of-the-art detector of [48], as well as largely outperforming *Test two*.

6.7 Conclusion and Discussion

This chapter studies the problem of identifying source camera device for JPEG images. In the most literature, PRNU-based detectors nearly dominate the research community of source camera identification. In this context, each camera device is characterized by its linear parametric model with parameters (c, d), which are

obtained from the estimated camera device fingerprints $(\widetilde{a}_k, \widetilde{b}_k)$ describing the proposed improved signal-dependent noise model. Then the problem of camera device identification is cast in the framework of hypothesis testing theory. Assuming that all the parameters are perfectly known, the statistical performance of the LRT is analytically established. In the practical cases, based on the estimated parameters, our designed two GLRTs perform very well at the prescribed FAR.

The main limitation of our proposed tests is that since the camera fingerprints $(\widetilde{a}_k, \widetilde{b}_k)$ are extracted based on multiple images, our designed tests can only deal with the case of testing a set of images (at least 40) together. In this scenario, if a forensic investigator aims at identifying the origin of the only one inspected image, our proposed tests will be invalid.

Chapter 7

Exposing Image Resampling by Using Linear Parametric Model

Contents

To distinguish an original image from a resampled one (including resizing, rotation and other linear transformations) is a research subfield of digital image forensics on which this chapter focuses. Generally, after image tempering such as splicing or copy-move operation, to convince the altered image, resampling operation is very essential which helps the forger create a visually perfect match. However, any manipulation unavoidably leaves traces. Thus, the research on resampling detection could serve as a auxiliary tool for digital image forensics. In 2005, authors of [68] first proposed to expose the linear correlation existing in the resampled images by using Expectation Maximization (EM) algorithm (see [72]). Meanwhile, by exploiting the second-derivative algorithm, it was proposed to uncover the periodicity of an interpolated image in [75]. Driven by these pioneer works, it was proposed to analyze the relevance between EM detector and second-derivative one, and design an equivalent accelerated and simplified detector in [73]. Instead of using a single predictor, it was proposed to use a global predictor to detect a resized image (see [74]). To the best of our knowledge, the detector based on EM algorithm in [68] is still very competitive for considering its detection power. However, model parameters estimation causes prediction errors and the iteration computation is very time-consuming, specially in the case of detecting a large database of inspected images.

Based on the periodicity of a resampled signal, we design a linear parametric model exposing the linear correlation between neighbouring samples in a one-dimensional (1-D) signal. Then, by dealing with the nuisance parameters, it is proposed to define the conditional probability of residual noise from the resampled signal by using Bayes' rule. Moreover, through Fourier spectrum, the up-sampled factor is estimated accurately. Similarly, a two-dimensional (2-D) image is described by the linear parametric model. Then, after eliminating the expectation of the inspected image, together with the Bayes' rule, the conditional probability of residual noise from the resampled image is estimated. Finally, the periodic artifact of the resampled image is exposed with symmetrical surrounding bright spots scattering in Fourier domain. Moreover, based on the probability of residual noise from images, it is proposed to design a practical Likelihood Ratio Test (LRT) for dealing with detecting a batch of resampled images.

This chapter is organized as follows. Section 7.1 briefly illustrates the principle of 1-D signal resampling. The linear parametric model is proposed for dealing with the nuisance parameters. Then we expose the artifact of 1-D resampled signal in Fourier domain. In Section 7.2, the linear correlation of a 2-D resampled image is described. Then, the similar linear model is proposed. The artifact of 2-D resampled image is demonstrated in Fourier domain. The design of a practical LRT is presented in Section 7.3. Section 7.4 presents numerical results of the proposed algorithm on natural images. The perspective of our future work is generally presented in Section 7.5. Section 7.6 finally concludes this chapter.

7.1 One-dimensional Signal Resampling and Its Artifact

7.1.1 Problem Statement of One-dimensional Signal Resampling

In general, resampling can be proceeded by three steps: up-sampling; interpolation; down-sampling (see details in [186]). Since linear interpolation is widely-adopted, for convenience and clarity, it is assumed that the interpolation algorithm studied in the present chapter is linear. In the following paragraph, a simple illustration of the resampling process is displayed (see Figure 7.1).

Without loss of generality, a 1-D discrete signal is created (see Figure 7.1a), denoted the vector $\mathbf{x} = \{x[t]\}\ t \in \{1, ..., T\}$, where T denotes the length of the vector. First, by up-sampling with the factor p, a new up-sampled vector $\mathbf{x}_u = \{x_u[t]\}$ is generated with the length $p(T-1)+1$ (see Figure 7.1b). In the case of $t = 1, 2.., T$, $x_u[p(t-1)+1] = x[t]$; otherwise, $x_u[t] = 0$. Second, by convolving the vector $\mathbf{x}_u$ with a linear low-pass filter $h[t]$, the interpolated vector $\mathbf{x}_i = \{x_i[t]\}$, $t \in \{1, ..., p(T-1)+1\}$ is obtained, where $x_i[t] = x_u[t] * h[t]$ (see Figure 7.1c). Third, by down-sampling with the factor q, a new down-sampled vector $\mathbf{x}_d = \{x_d[t]\}$ with the length $\lfloor \frac{p}{q}(T-1)+1 \rfloor$, where $x_d[t] = x_i[q(t-1)+1]$, $t = 1, 2, ..., \lfloor \frac{p}{q}(T-1)+1 \rfloor$ (see Figure 7.1d). Let us denote the resampled signal as $\mathbf{y} = \{y[t]\}$, where $\mathbf{y} \equiv \mathbf{x}_d$. It should be noted that $p, q \in \mathbb{N}^+$. Then, it is immediately proposed that the

resampled signal $\mathbf{y}$ is described by the following linear equation:

$$\mathbf{y} = \mathbf{B}_{\frac{p}{q}}\mathbf{x} \tag{7.1}$$

where $\mathbf{y} = \{y_m\}$, $m \in \{1, ..., M\}$, denoted a resampled signal, $\mathbf{x} = \{x_n\}$, $n \in \{1, ..., N\}$, denoted an original signal. A $M \times N$ linear resampling matrix $\mathbf{B}_{\frac{p}{q}}$ controls the resampling procedure, including up-sampling, interpolation, and down-sampling. For instance, the linear resampling matrix $\mathbf{B}_{\frac{4}{3}}$ (see Figure 7.1) takes the form:

$$\mathbf{B}_{\frac{4}{3}} = \begin{pmatrix} 1 & 0 & 0 & 0 & \dots \\ 0.25 & 0.75 & 0 & 0 & \dots \\ 0 & 0.5 & 0.5 & 0 & \dots \\ 0 & 0 & 0.75 & 0.25 & \dots \\ 0 & 0 & 0 & 1 & \dots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}$$

where $\mathbf{B}_{\frac{4}{3}}$ has the period 4. Since the linear resampling matrix $\mathbf{B}_{\frac{p}{q}}$ introduces the periodicity, the resampled signal $\mathbf{y}$ owns the period p. Moreover, the expectation of the periodic signal y_m should be linear combination from the neighbouring samples, which is defined by:

$$\hat{y}_m = \sum_{k=-L}^{L} \alpha_k y_{m+k} \tag{7.2}$$

where α_k denotes the weighted factor with $\alpha_0 = 0$ and $2L$ ($L \in \mathbb{N}^+$) the number of used neighbouring samples. In the present chapter the moments of (7.2) is not proved detailed, the interested reader is referred to [68]. It should be noted that the linear combination (7.2) can not hold true in each sample along the signal y_m, but only be defined in the interval of p. In the practical resampling, if an up-sampling factor p set as 1 while a down-sampling factor $q > 1$, the resampled signal does not expose the periodicity, and then its expectation can not be formulated by linear combination (7.2).

Then it is proposed to use a linear parametric model to deal with a 1-D resampled signal $\mathbf{y}$, which has been realized successfully in [55]. To this end, the vector $\mathbf{y} = \{y_1, \dots, y_M\}^T$ is defined by:

$$\mathbf{y} \sim \mathcal{N}(\boldsymbol{\mu}, \sigma^2\mathbf{I}_M) = \boldsymbol{\mu} + \mathbf{n} \tag{7.3}$$

where $\boldsymbol{\mu} = (\mu_1, \dots, \mu_M)^T$ denotes the expectation of $\mathbf{y}$ and $\mathbf{n} = (n_1, \dots, n_M)^T$ denotes residual noise following the Gaussian model with the variance $\sigma^2\mathbf{I}_M$ and $\mathbf{I}_M$ the identity matrix of size $M \times M$. Furthermore, $\boldsymbol{\mu}$ can be described with the following linear parametric model:

$$\boldsymbol{\mu} = \mathbf{A}\boldsymbol{\omega} \tag{7.4}$$

where $\mathbf{A}$ is a known full rank matrix of size $M \times N$, with $M > N$, and $\boldsymbol{\omega}$ is a $N \times 1$ vector of parameters describing the expectation of $\mathbf{y}$. Note that in the present

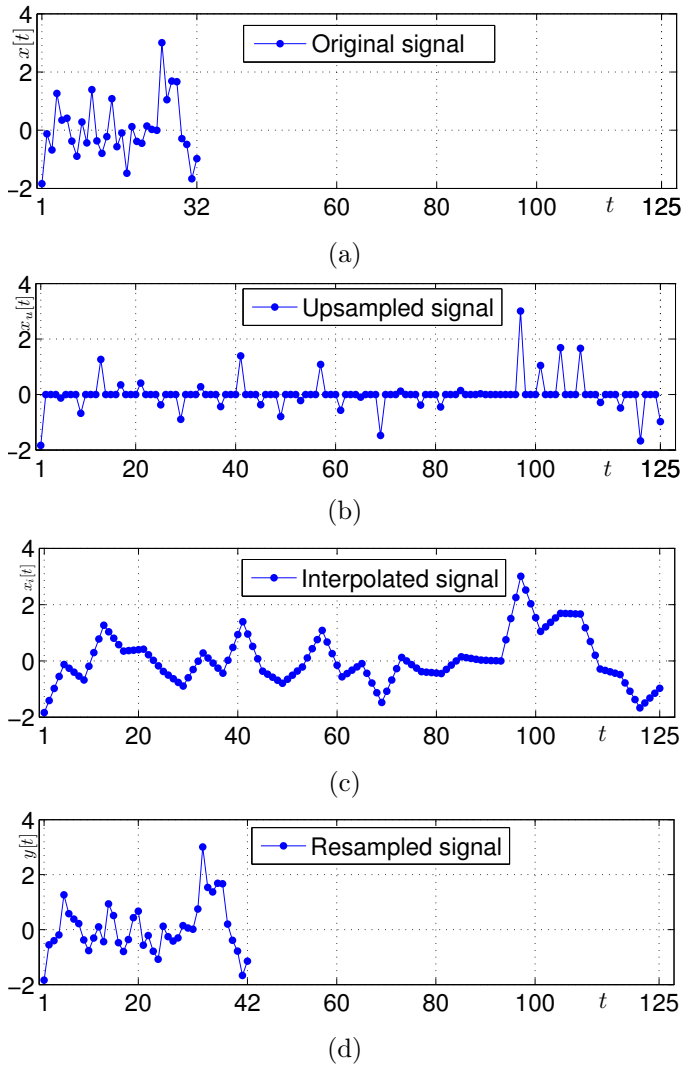


Figure 7.1: Illustration of resampling 1-D signal by a factor $\frac{p}{q} = \frac{4}{3}$: (a) 1-D original signal, (b) up-sampled signal by a factor $p = 4$, (c) interpolated signal, (d) resampled signal after down-sampling by using a factor $q = 3$.

chapter, the chosen linear parametric model is the following matrix $\mathbf{A}$:

$$\mathbf{A} = \begin{pmatrix} y_1 & \cdots & y_L & y_{L+2} & \cdots & y_{2L+1} \\ y_2 & \cdots & y_{L+1} & y_{L+3} & \cdots & y_{2L+2} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ y_m & \cdots & y_{L+m-1} & y_{L+m+1} & \cdots & y_{2L+m} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ y_{M-2L} & \cdots & y_{M-L-1} & y_{M-L+1} & \cdots & y_M \end{pmatrix},$$

where m denotes an index of vector $\mathbf{y}$ and $2L$ a number of used neighbouring samples, with $M > 3L$. It should be addressed that the expectations of the first and last L

samples are not considered in our proposed model.

The idea of using such a linear parametric model is that it allows an easy elimination of nuisance parameters. To apply this theory, let us define $C(\mathbf{A})$ the column space spanned by $\mathbf{A}$, with dim $(C(\mathbf{A}))$ = rank $(\mathbf{A}) = N$ and $C(\mathbf{A})^{\perp}$ its orthogonal complement, sometimes referred to as the "parity space", with dim $(C(\mathbf{A})) = M - N$. The projection of observation vector $\mathbf{y}$ onto the parity space is obtained by $\mathbf{n} = \mathbf{W}\mathbf{y}$ where the matrix $\mathbf{W}$ verifies, among others, the following useful properties:

$$\mathbf{W}\mathbf{A} = \mathbf{0} \text{ and } \mathbf{W}\mathbf{W}^{T} = \mathbf{I}_{M-N}. \tag{7.5}$$

Hence, the projection of observation vector $\mathbf{y}$ onto the parity space $C(\mathbf{A})^{\perp}$ yields: $\mathbf{n} = \mathbf{W}\mathbf{y} = \mathbf{W}\xi \sim \mathcal{N}(0, \sigma^2\mathbf{I}_{M-N})$. Note that the use of projection matrix $\mathbf{W}$ can be replaced by using the Maximum Likelihood Estimation (MLE). By rejecting the nuisance parameter, let us define the estimation of $\mathbf{n}$ as

$$\widehat{\mathbf{n}} = \mathbf{y} - \widehat{\mathbf{y}} = \mathbf{P}_{\mathbf{A}}^{\perp}\mathbf{y} \text{ with } \mathbf{P}_{\mathbf{A}}^{\perp} = \mathbf{I}_M - \mathbf{A}(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T. \tag{7.6}$$

In fact, a straightforward calculation, using the properties (7.5), shows that :

$$\left\| \mathbf{P}_{\mathbf{A}}^{\perp}\mathbf{y} \right\|_2^2 = \mathbf{y}^{\mathrm{T}}\mathbf{W}^{\mathrm{T}}\mathbf{W}\mathbf{W}^{\mathrm{T}}\mathbf{W}\mathbf{y} = \left\| \mathbf{W}\mathbf{y} \right\|_2^2.$$

For clarity, in the present chapter the matrix $\mathbf{W}$ is used in all calculus while matrix $\mathbf{P}_{\mathbf{A}}^{\perp}$ is used for illustrations and figures, see Figure 7.2, to keep the same number of observations.

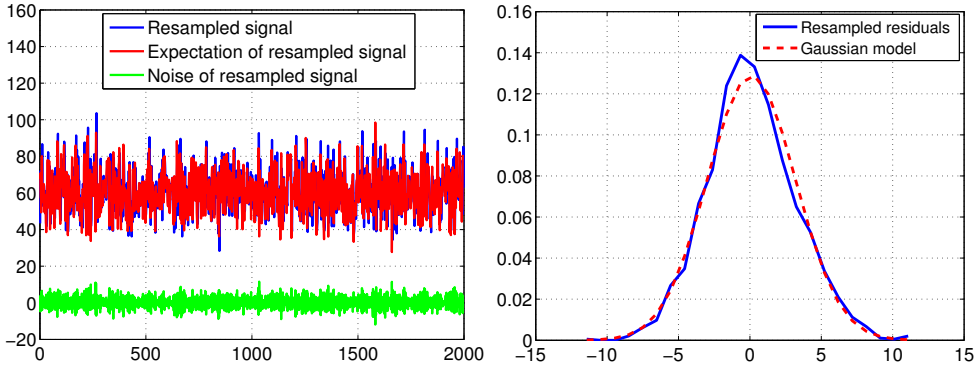
To verify the effectiveness of our proposed model, the simulated experiments are given. A group of random variables of 1000 sample is generated. Then, by using a resampling factor $\frac{p}{q} = \frac{2}{1}$, the resampled signal is obtained (see Figure 7.2a). Based on our proposed linear parametric model, the residual noise of resampled signal is acquired, which can be fitted by using the Gaussian distribution (see Figure 7.2b). Obviously, the resampled residual noise fits the Gaussian model well, which verifies our proposed assumption that residual noise extracted from the resampled signal approximately follows the Gaussian distribution.

7.1.2 Exposing One-dimensional Signal Resampling

By using the linear parametric model, the residual noise is obtained, which is denoted $\mathbf{n} = \{n_i\}$ $i \in \{1, ..., I\}$, following the Gaussian distribution (see Figure 7.2b). For simplifying the calculation, in our present chapter, it is assumed that the residual noise from non-resampled/original signal is uniformly distributed. Thus, the problem of detecting between the original and resampled signal can be formulated by the following two hypotheses:

$$\begin{cases} \mathcal{H}_0 = \{n_i \sim \mathcal{U}(n_{min}, n_{max})), \forall i = (1, ..., I)\}, \\ \mathcal{H}_1 = \{n_i \sim \mathcal{N}(0, \sigma^2), \forall i = (1, ..., I)\}, \end{cases} \tag{7.7}$$

where under hypothesis $\mathcal{H}_0$, the residual noise of the inspected signal follows the Uniform distribution; under hypothesis $\mathcal{H}_1$, the residual noise is normally distributed



(a) Illustration of resampled signal, its estimated expectation and residual noise

(b) Empirical results from the resampled residual and theoretical results from its Gaussian model fitting.

Figure 7.2: Illustration of a resampled image using a factor $\frac{p}{q} = \frac{2}{1}$.

with zero mean and σ^2 variance. It is addressed that n_{min} denotes the minimum value of the noise and n_{max} the maximum value, assuming that $n_{min} \neq n_{max}$.

Then, by using Bayes' rule, let us immediately denote the conditional probability of each noise sample from the resampled signal by:

$$\Pr\{n_i \in \mathcal{H}_1 \mid n_i\} = \frac{\Pr\{n_i \mid n_i \in \mathcal{H}_1\} \Pr\{n_i \in \mathcal{H}_1\}}{\sum_{k=0}^{1} \Pr\{n_i \mid n_i \in \mathcal{H}_k\} \Pr\{n_i \in \mathcal{H}_k\}}, \tag{7.8}$$

where it is assumed that the priors $\Pr\{n_i \in \mathcal{H}_0\}$ and $\Pr\{n_i \in \mathcal{H}_1\}$ are both equal to $\frac{1}{2}$. Then (7.7) is rewritten as follows:

$$\begin{cases} \Pr\{n_i \mid n_i \in \mathcal{H}_0\} = \frac{1}{n_{max} - n_{min}} \\ \Pr\{n_i \mid n_i \in \mathcal{H}_1\} = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{-n_i^2}{2\sigma^2}\right). \end{cases} \tag{7.9}$$

where the variance σ^2 can be estimated by using MLE as follows:

$$\widehat{\sigma^2} = \frac{1}{I-1} \sum_{i=1}^{I} (n_i - \frac{1}{I} \sum_{i=1}^{I} n_i)^2. \tag{7.10}$$

Similar to the method proposed in [75], it is proposed to use the frequential representation of the probability of resampled residuals, denoted $\mathbf{N}$ and formally defined as follows:

$$\mathbf{N} = |\mathbf{h} * |f[\Pr\{\mathbf{n} \in \mathcal{H}_1 \mid \mathbf{n}\}]| \,|, \tag{7.11}$$

where a vector $\mathbf{h} = [1/4, 1/4, -1, 1/4, 1/4]$ denotes a 1-D high-pass filter, $f[\cdot]$ represents the calculation of 1-D Discrete Fourier Transform (DFT). It should be addressed that the high-pass filter aims at removing the lower frequential components for better displaying the peak value, which is the most distinctive artifact of the resampled signal, in Fourier domain (see Figure 7.3b and 7.3c).

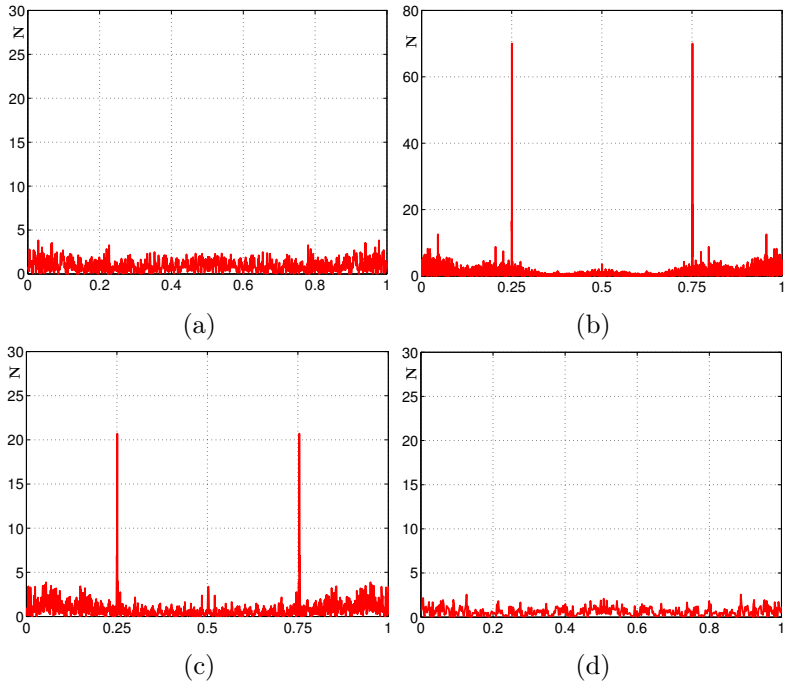


Figure 7.3: Probability comparison of residual noise between original and resampled signal in Fourier domain, (a) original signal, (b) resampled signal by a factor $\frac{p}{q} = \frac{4}{1}$, (c) resampled signal by a factor $\frac{p}{q} = \frac{4}{3}$, (d) resampled signal by a factor $\frac{p}{q} = \frac{1}{2}$.

To verify the effectiveness of our proposed algorithm, it is proposed to compare the original signal with the resampled one by using different resampling factors. As Figure 7.3 illustrates, the resampled signal (see Figure 7.3b and 7.3c) exposes the obvious peaks in the frequency domain, but the original signal (see Figure 7.3a) has nothing. Moreover, surprisingly, the resampled period is accurately estimated by:

$$\hat{T} = \frac{1}{f_p} \tag{7.12}$$

where $\hat{T}$ denotes the estimated period or interpolated rate of the resampled signal, which equals to the up-sampled factor p. f_p is the normalized frequency, corresponding to the position of the first peak in Fourier domain. By counting the number of the local peaks with a threshold τ_1, it is proposed to design a test of detecting a resampled signal.

It is noted that our proposed algorithm can not differentiate between two resampled signals by using the same up-sampled factor. For instance, the peaks associated with a resampled factor $\frac{4}{1}$ (see Figure 7.3b) appears similar to those by a resampled factor $\frac{4}{3}$ (see Figure 7.3c). Since our target is to detect whether the inspected signal is resampled, this limitation is not crucial. In addition, in the case of the resampled signal by an up-sampled factor $p = 1$, where it has no obvious peaks (see Figure 7.3d), it can not be distinguished with the original signal. In this case,

each sample of the resampled signal exposes no periodic linear correlation with its neighbouring ones. Thus, our proposed algorithm is invalid.

7.2 Two-dimensional Signal Resampling and Its Artifact

In this section, inspired by extracting the artifact of the 1-D resampled signal (7.11), we can extend a 1-D signal to a 2-D image straightforward. Compared with a 1-D resampled signal, a 2-D resampled image has more complex linear correlation. A 1-D signal is interpolated by one direction, but a 2-D image is interpolated by the horizontal and vertical direction simultaneously. When a 2-D image is resampled only by the horizontal or vertical direction, it will cause the image visual distortion, which is not discussed in this context.

7.2.1 Problem Statement of Two-dimensional Signal Resampling

Similarly, a procedure of resampling a 2-D image has three steps: up-sampling by the horizontal and vertical direction; interpolation by two direction; down-sampling by two direction. Then similar to the linear combination of 1-D signal, each pixel intensity of the resampled 2-D image correlates to its surrounding ones. Next, let us denote an original grey-level image as $\mathbf{X} = \{x_{i,j}\}$ $i \in \{1, ..., I\}$, $j \in \{1, ..., J\}$, where I and J denote the height and the width of the matrix $\mathbf{X}$. The case of color image with three channels can be obtained straightforward.

Let us denote the linear correlation between each pixel sample and its neighbouring ones by the following:

$$x_{i,j} = \sum_{p=-L}^{L} \sum_{q=-L}^{L} \alpha_{p,q} x_{i+p,j+q} \tag{7.13}$$

where $\alpha_{p,q}$ denotes a weighted factor with the case of $\alpha_{0,0} = 0$, $(2L+1)^2 - 1$ ($L \in \mathbb{N}^+$) the number of used neighbouring samples, $x_{i,j}$ $i \in \{2, ..., I-1\}$, $j \in \{2, ..., J-1\}$ denotes a pixel intensity. Enlarging L the range of neighbouring indeed improves the accuracy of describing the linear correlation, but at the cost of high computation. For simplicity, $L = 1$ is a good tradeoff between complexity and accuracy. Then let us rewrite a resampled image $\mathbf{X} = \{\mathbf{x}_j\}$, the vector $\mathbf{x}_j = \{x_{2,j}, ..., x_{I-1,j}\}^T$, where $j \in \{2, ..., J-1\}$. Immediately, it is proposed to use a linear parametric model to deal with the 2-D image. To this end, the vector $\mathbf{x}_j$ is defined:

$$\mathbf{x}_j \sim \mathcal{N}(\boldsymbol{\mu}_j, \sigma_j^2 \mathbf{I}_{I-2}) = \boldsymbol{\mu}_j + \mathbf{n}_j \tag{7.14}$$

where $\boldsymbol{\mu}_j = (\mu_2, \ldots, \mu_{I-1})^T$ denotes the expectation, each Gaussian-distributed column vector $\mathbf{n}_j = (n_2, \ldots, n_{I-1})^T$ takes variance $\sigma_j^2 \mathbf{I}_{I-2}$ and $\mathbf{I}_{I-2}$ the identity matrix of size $(I-2) \times (I-2)$. Furthermore, $\boldsymbol{\mu}_j$ can be described with the following linear parametric model:

$$\boldsymbol{\mu}_j = \mathbf{A}_j \boldsymbol{\alpha}_j \tag{7.15}$$

where $\mathbf{A}_j$ is a known full rank matrix of size $(I-2)\times 8$, with $I > 10$, and $\boldsymbol{\alpha}_j = \{\alpha_{p,q}\}$ where index $(p,q) \in \{-1,0,1\}$ excluding (0,0) is a 8×1 vector describing the expectation of $\mathbf{x}_j$. Then, similar to (7.6), let us define the estimation of $\mathbf{n}_j$ as

$$\hat{\mathbf{n}}_j = \mathbf{x}_j - \hat{\mathbf{x}}_j = \mathbf{x}_j - (\mathbf{A}_j(\mathbf{A}_j^T\mathbf{A}_j)^{-1}\mathbf{A}_j^T)\mathbf{x}_j. \tag{7.16}$$

Note that in the present chapter, the chosen linear parametric dealing with the 2-D image is the following matrix $\mathbf{A}_j$:

$$\mathbf{A}_j = [\mathbf{x}_j^{(-1,-1)}, \mathbf{x}_j^{(-1,0)}, \mathbf{x}_j^{(-1,1)}, \mathbf{x}_j^{(0,-1)}, \mathbf{x}_j^{(0,1)}, \mathbf{x}_j^{(1,-1)}, \mathbf{x}_j^{(1,0)}, \mathbf{x}_j^{(1,1)}] \tag{7.17}$$

where a matrix $\mathbf{X}^{(p,q)} = \{\mathbf{x}_j^{(p,q)}\}$, each column vector $\mathbf{x}_j^{(p,q)} = \{x_{2+p,j+q}, ..., x_{I+p-1,j+q}\}^T$, where $j \in \{2, ..., J-1\}$ and index $(p,q) \in \{-1,0,1\}$ not including (0,0).

7.2.2 Exposing Two-dimensional Signal Resampling

By using the linear parametric model, the residual noise of the resampled image is obtained, which is denoted $\mathbf{N} = \{\mathbf{n}_2, ..., \mathbf{n}_{J-1}\}$, each residual column vector $\mathbf{n}_j = \{n_{i,j}\}$, $i \in \{2, ..., I-1\}$ following the Gaussian distribution. Thus, the problem of detecting between the non-resampled and resampled image can be formulated by the following two hypothesis:

$$\begin{cases} \mathcal{H}_0 = \{n_{i,j} \sim \mathcal{U}(-255, 255)), \forall i = (2, ..., I-1)\}, \\ \mathcal{H}_1 = \{n_{i,j} \sim \mathcal{N}(0, \sigma_j^2), \forall i = (2, ..., I-1)\}, \end{cases} \tag{7.18}$$

where $j \in \{2, ..., J-1\}$, under hypothesis $\mathcal{H}_0$, the residual noise of the inspected image follows the Uniform distribution; under hypothesis $\mathcal{H}_1$, the residual noise is normally distributed with zero mean and σ_j^2 variance. Then, by using Bayes' rule, let us immediately denote the conditional probability of each residual sample:

$$\Pr\{n_{i,j} \in \mathcal{H}_1 \,|\, n_{i,j}\} = \frac{\Pr\{n_{i,j} \,|\, n_{i,j} \in \mathcal{H}_1\} \Pr\{n_{i,j} \in \mathcal{H}_1\}}{\sum_{k=0}^{1} \Pr\{n_{i,j} \,|\, n_{i,j} \in \mathcal{H}_k\} \Pr\{n_{i,j} \in \mathcal{H}_k\}}, \tag{7.19}$$

where it is assumed that the priors $\Pr\{n_{i,j} \in \mathcal{H}_0\}$ and $\Pr\{n_{i,j} \in \mathcal{H}_1\}$ are both equal to $\frac{1}{2}$. Then (7.18) is rewritten as follows:

$$\begin{cases} \Pr\{n_{i,j} \mid n_{i,j} \in \mathcal{H}_0\} = \frac{1}{510}, \\ \Pr\{n_{i,j} \mid n_{i,j} \in \mathcal{H}_1\} = \frac{1}{\sqrt{2\pi\widehat{\sigma}_j^2}} \exp\left(\frac{-n_{i,j}^2}{2\widehat{\sigma}_j^2}\right), \end{cases} \tag{7.20}$$

where the variance $\widehat{\sigma}_j^2$ can be estimated by using MLE as follows:

$$\widehat{\sigma}_j^2 = \frac{1}{I-3} \sum_{i=2}^{I-1} (n_{i,j} - \frac{1}{I-2} \sum_{i=2}^{I-1} n_{i,j})^2. \tag{7.21}$$

Then it is proposed to use the frequential representation of the probability of residual noise, denoted $\mathbf{N}^{2D}$ and formally defined as follows:

$$\mathbf{N}^{2D} = |\mathbf{H} * |\mathcal{F}[\mathbf{P}_{map}]|\,|, \tag{7.22}$$

where $\mathbf{H}$ denotes a 2-D high-pass filter, $\mathcal{F}[\cdot]$ represents the calculation of 2-D DFT. $\mathbf{P}_{map} = \{\Pr\{n_{i,j} \in \mathcal{H}_1 \,|\, n_{i,j}\}\}$, $i \in \{2, ..., I-1\}$, $j \in \{2, ..., J-1\}$ denotes the so-called probability map (see [68]). Similar to peaks of a 1-D resampled signal, a 2-D resampled image exposes bright spots surrounding the central point (similar to peaks in the 1-D signal) distributed symmetrically in Fourier domain, which is the critical resampling artifact of a resampled image. Based on the assumed distribution (7.18), together with hypothesis theory, we can establish a practical Likelihood Ratio Test (LRT) of detecting a resampled image. The specific discussion will be extended in the following section.

7.3 Design of the Practical LRT for Resampling Detection

Although the periodic resampling artifacts can be detected by searching for the peaks in Fourier domain, it can not be used for detecting a batch of resampled images. Prior to our proposed algorithm in this chapter, some state-of-the-art methods constraint their idea focused on the peaks in Fourier domain. Namely, all the threshold are set based on the peaks (see [68] [73]). In this section, it is proposed to solve the problem of resampling detection by designing a practical LRT.

For solving the statistical detection problem such as (7.18), it follows from the Neyman-Pearson lemma [20, Theorem 3.2.1] that the LRT is optimal in the sense described below. For definition, let

$$\mathcal{K}_\alpha = \left\{ \delta : \sup_{\sigma_j} \mathbb{P}_{\mathcal{H}_0}[\delta(\mathbf{V}) = \mathcal{H}_1] \leq \alpha \right\} \tag{7.23}$$

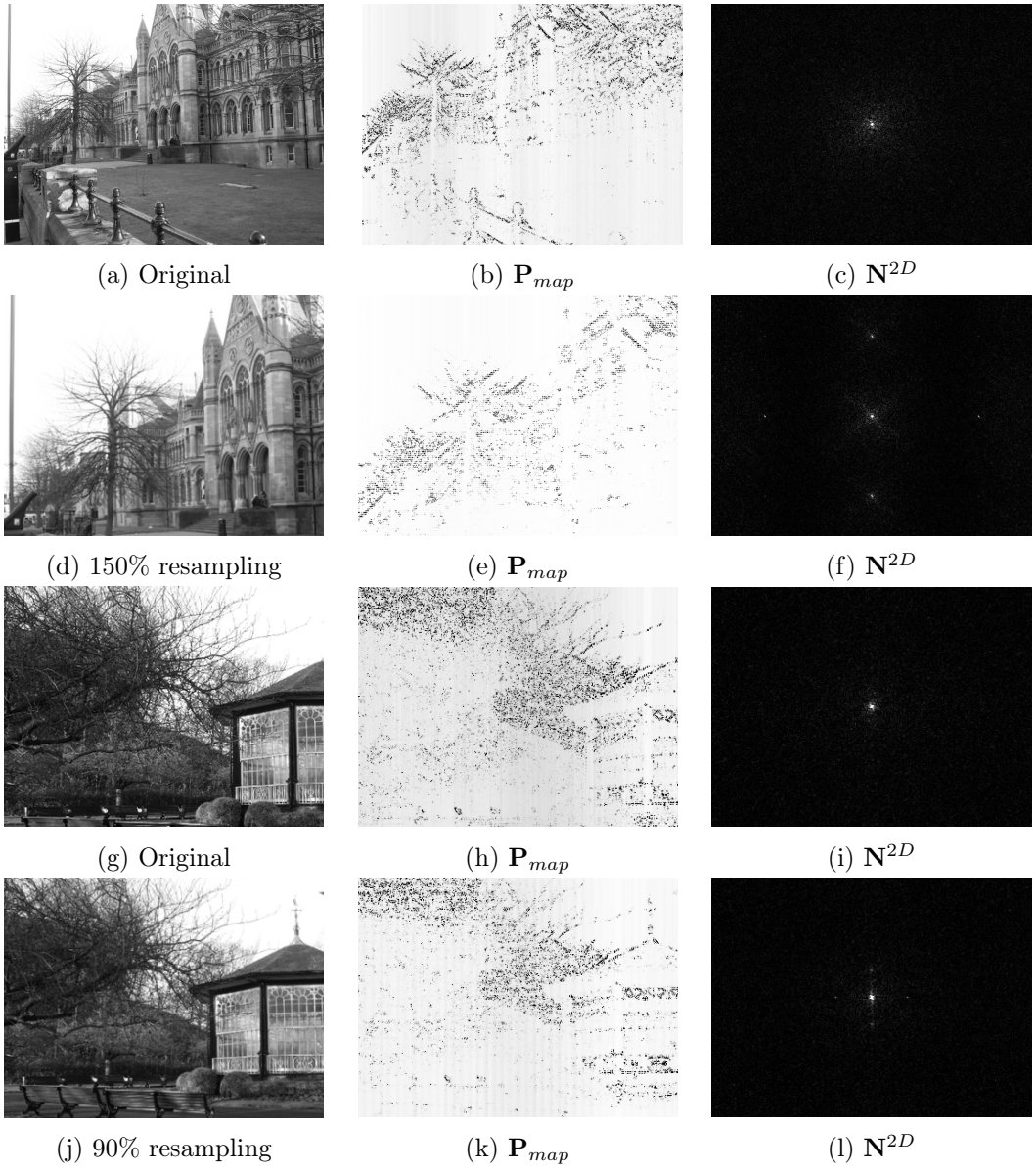
be the class of tests, solving problem (7.18), with an upper-bounded False Alarm Rate (FAR) α. Here $\mathbb{P}_{\mathcal{H}_j}[\cdot]$ is the probability under $\mathcal{H}_j, j \in \{0, 1\}$, and the supremum over model parameter σ_j can be understood as whatever the distribution parameters might be, in order to ensure that the FAR α can not be exceeded. Among all tests in $\mathcal{K}_\alpha$ the LRT is the most powerful test, it maximizes the detection power:

$$\beta_\delta = \mathbb{P}_{\mathcal{H}_1}[\delta(\mathbf{V}) = \mathcal{H}_1]. \tag{7.24}$$

In the practical test, parameter σ_j is unknown which can be successfully estimated by using MLE (see Equation (7.21)). Afterwards, we can establish our proposed practical LRT for detecting resampled images. Based on the probability density function (pdf) of hypothesis $\mathcal{H}_0$ and $\mathcal{H}_1$ (7.20), the Likelihood Ratio (LR) value for the i-th pixel-wise on j-th column vector is given by:

$$\Lambda_{i,j}(n_{i,j}) = \frac{\Pr\{n_{i,j} \mid n_{i,j} \in \mathcal{H}_1\}}{\Pr\{n_{i,j} \mid n_{i,j} \in \mathcal{H}_0\}} = \frac{510}{\widehat{\sigma}_j \sqrt{2\pi}} \exp\left(\frac{-n_{i,j}^2}{2\widehat{\sigma}_j^2} \right). \tag{7.25}$$

Then, it follows from the statistical independence of vectors $n_{i,j}$ that the practical



(a) Original (b) $\mathbf{P}_{map}$ (c) $\mathbf{N}^{2D}$

(d) 150% resampling (e) $\mathbf{P}_{map}$ (f) $\mathbf{N}^{2D}$

(g) Original (h) $\mathbf{P}_{map}$ (i) $\mathbf{N}^{2D}$

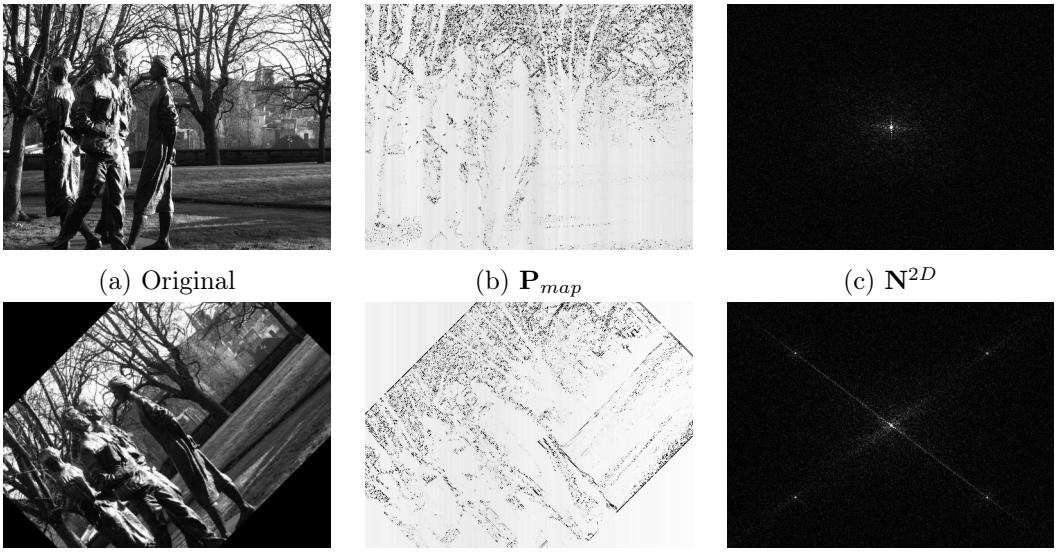
(j) 90% resampling (k) $\mathbf{P}_{map}$ (l) $\mathbf{N}^{2D}$

Figure 7.4: Left-most column: comparison between an original image and its corresponding resampled ones (*bilinear* interpolation); middle column: its corresponding $\mathbf{P}_{map}$'s; right-most column: periodic resampling artifacts in Fourier domain.

LRT for all the pixel intensities is given by:

$$\delta(\mathbf{V}) = \begin{cases} \mathcal{H}_0 \text{ if } \Lambda(\mathbf{V}) = \sum_{j=2}^{J-1} \Lambda(V_j) \leq \tau_\alpha \\ \mathcal{H}_1 \text{ if } \Lambda(\mathbf{V}) = \sum_{j=2}^{J-1} \Lambda(V_j) > \tau_\alpha \end{cases} \tag{7.26}$$

$$\text{with } \Lambda(V_j) = \sum_{i=2}^{I-1} \log(\Lambda_{i,j}) \tag{7.27}$$



(a) Original (b) $\mathbf{P}_{map}$ (c) $\mathbf{N}^{2D}$

(d) 45° (e) $\mathbf{P}_{map}$ (f) $\mathbf{N}^{2D}$

Figure 7.5: Left-most column: comparison between an original image and its corresponding resampled one (*bilinear* interpolation); middle column: its corresponding $\mathbf{P}_{map}$'s; right-most column: periodic resampling artifacts in Fourier domain.

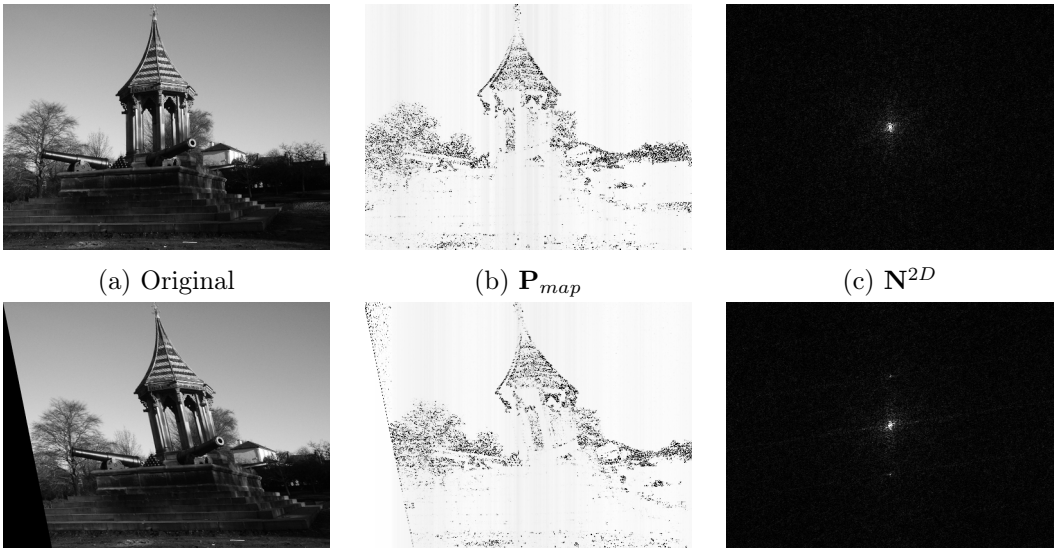
where the decision threshold τ_α is the solution of equation $\mathbb{P}_0[\Lambda(\mathbf{V}) > \tau_\alpha] = \alpha$ to guarantee that $\delta(\mathbf{V}) \in \mathcal{K}_\alpha$.

7.4 Numerical Experiments

7.4.1 Test Performance on A Single Resampled Image

To verify the sharpness and effectiveness of our proposed methodology by using a linear parametric model, numerical experiments are conducted in this section. Prior to our experiments, it is proposed to randomly choose some images from the image database [187], in which all 512×384 RAW images are transformed to TIFF format or JPEG format with low-compressed rate.

Figure 7.4 illustrates the results from resampled images by using *bilinear* interpolation. Compared with the single bright spot from an original image (see Figure 7.4c), the up-sampled image, whose resampled factors equal 150%, uncover a series of bright spots symmetrically surrounding the central point in Fourier domain (see Figure 7.4f). It is observed that with increasing the up-sampled factor, the distance between two spots horizontally or vertically is enlarged. This interesting phenomenon is very similar to the pattern of the distance between symmetric peaks in Fourier domain of the 1-D resampled signal. The results of down-sampled images are displayed in Figure 7.4l. Compared to the results of up-sampled images, the bright spots surrounding the central point are not very obvious. It should be attributed to that down-sampled images loss more information of linear correlation



(a) Original (b) $\mathbf{P}_{map}$ (c) $\mathbf{N}^{2D}$

(d) 20% horizontal shearing (e) $\mathbf{P}_{map}$ (f) $\mathbf{N}^{2D}$

Figure 7.6: Left-most column: comparison between an original image and its corresponding resampled one (*bicubic* interpolation); middle column: its corresponding $\mathbf{P}_{map}$'s; right-most column: periodic resampling artifacts in Fourier domain.

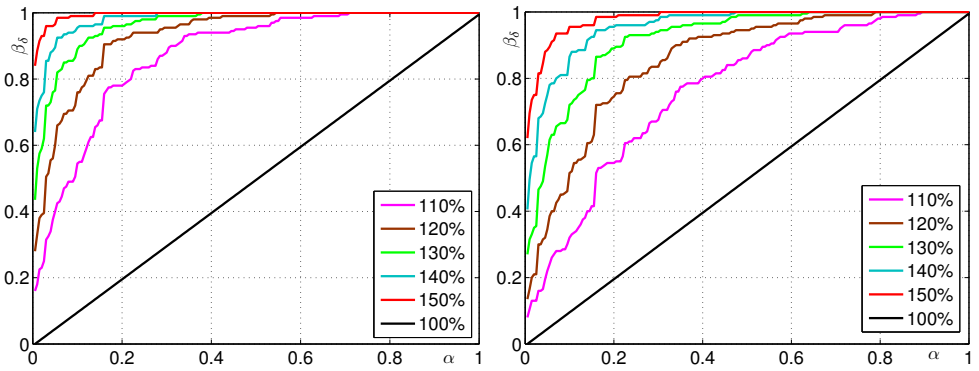
than those of up-sampled images. Thus it is more difficult to detect.

Next, it is proposed to compare the rotated images with the original image. Similarly, it can be still found that surrounding bright spots are scattered in Fourier domain (see Figure 7.5). Finally, Figure 7.6 gives the experimental results from different affine transformation by using *bicubic* interpolation. It is noted that *bicubic* interpolation generates less linear combination than *bilinear* interpolation. Besides, for better visual display of the surrounding bright spots, we can modify the high-pass filter $\mathbf{H}$ which is not mainly discussed in this chapter.

7.4.2 Test Performance on A Batch of Resampled Images

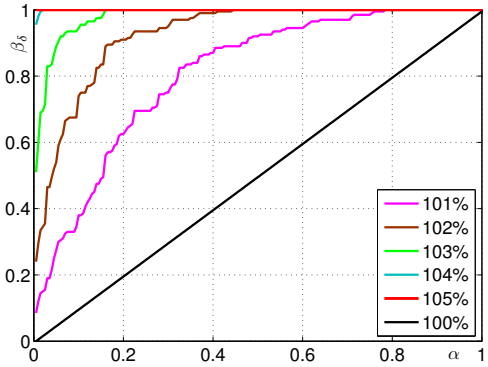
In the practical detection, it is no doubt that a plenty of images should be investigated. Prior to our proposed test in this chapter, some state-of-the-art detectors applied the algorithms based on the peak values [73] [68]. By using the peak value, it might be effective in judging a single inspected image. However, when processing numerical images, the detector of [68] fails to analyze the performance of its proposed algorithm. Although the detector of [73] present the detection performance, the statistical property still remains unknown.

For a large scale verification, it is proposed to give the experimental performance of our detector for processing a batch of inspected images. Then, we establish an image database of 200 uncompressed gray-level images, which is randomly chosen from Bossbase database [155]. Prior to using it, it is proposed to down-sample original images. In the following experiments, three practical interpolation methods,



(a) Interpolation by using *Bilinear* algorithm.

(b) Interpolation by using *Bicubic* algorithm.

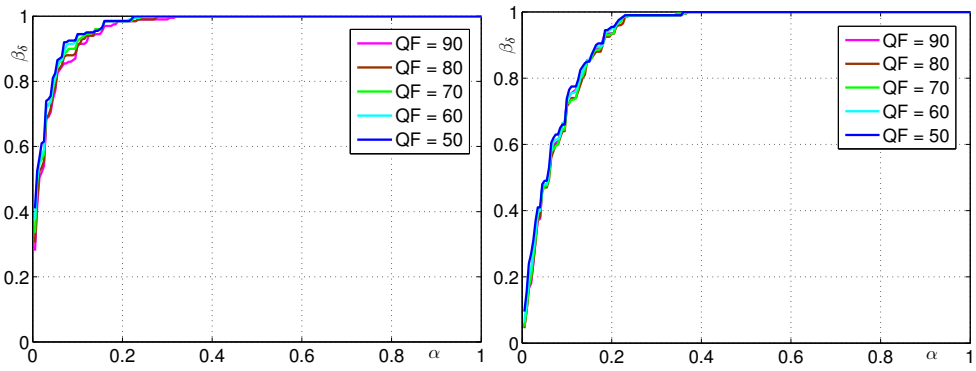


(c) Interpolation by using *Nearest-neighbor* algorithm.

Figure 7.7: ROC curves comparison by using up-sampled images where the legend denotes the interpolation factor. Note that 100% denotes non-resampled images.

that are *bilinear*, *bicubic*, *nearest-neighbor*, are used for verifying the effectiveness of our proposed algorithm. As the inspected images are up-sampled (see Figure 7.7) , by using *bilinear* or *nearest-neighbor*, our detector performs its ability to detect very tiny resampled images. When *nearest-neighbor* is applied for interpolating, even larger than 105% resampled images can be easily detected at the cost of very small FAR. Compared with *bilinear* or *nearest-neighbor*, since *bicubic* partly destroys the correlation existing in the resampled image, the performance of the detector is degraded. Besides, with increasing the interpolation factor, the performance of our detector is upgraded largely. Therefore the larger the interpolation factor is, the more stronger correlation exits between pixel intensities.

Apart from detecting resampled uncompressed images, our proposed detector is capable of detecting resampled compressed images. It is proposed to use two image databases. In the first database, we use the uncompressed 200 images in TIFF format. Then after resampling original uncompressed images by using the interpolation factor 150%, all the images are saved as JPEG format with Quality Factor (QF) spanning from 50 to 90 (see Figure 7.8a). By observing the ROC curves, JPEG com-



(a) Detection performance of JPEG format image.

(b) Detection performance of double-compressed JPEG format image.

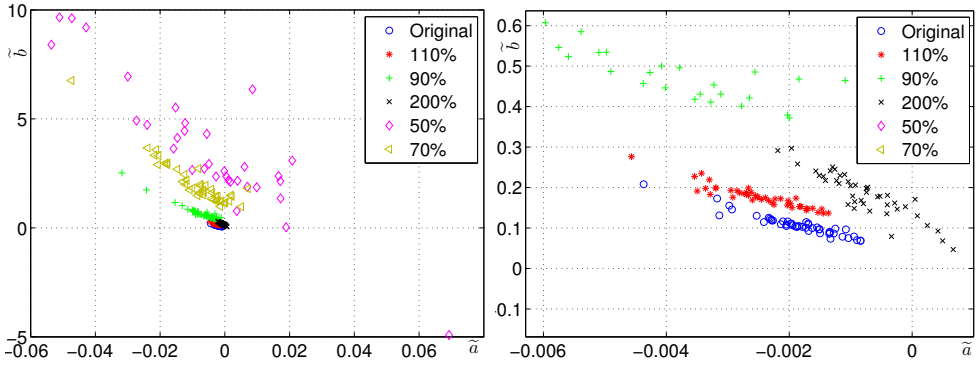
Figure 7.8: ROC curves comparison by using resampled compressed images where the legend denotes the corresponding QF.

pression does not impact our detection power. Then in the second database, we use the 200 JPEG format images with QF 70, and after resampling original compressed images by using the interpolation factor 150%, all the images are saved as JPEG format again with QF still spanning from 50 to 90 (see Figure 7.8b). Compared with results from Figure 7.8a, double-compressed JPEG images (see Figure 7.8b) indeed impact our detection accuracy, nevertheless the detection power is still better than the state-of-the-art algorithms, which have no ability to detect the JPEG format images, especially in the case of low QF's.

In practice, the periodic blocking artifact of the JPEG format image coincides with the periodic trace of the resampled image. Therefore, when a batch of resampled JPEG format images are inspected, it is very difficulty to authenticate whether the periodic pattern is from the blocking artifact or the resampling artifact. Since the artifact is taken as the benchmark to detect a batch of images, the state-of-the-art algorithms could not solve this tough problem. In our proposed algorithm, the practical LRT does not use the resampling artifact directly, but the ratio of the probability of residual noise, which effectively deals with the problem caused by JPEG compression.

7.5 Discussion of Our Future Work

Although our proposed test based on the Gaussian model performs relevantly well, especially in the case of detecting up-sampled images, it possibly can not detect the image with down-sampling operation. It should be explained that since the down-sampled image loses plenty of correlated information existing in the neighborhood pixels, our detector based on the correlation unavoidably becomes less sensitive to the down-resampled image, especially in the case of down-sampling sharply. Thus, the artifact of resampled signal will be vanish (see Figure 7.3d for one-dimensional signal) or become a little indistinguishable (see Figure 7.4l). Besides, it is not very



(a) Illustration of fingerprints from 50 resampled images.

(b) Zoom-in version of (a).

Figure 7.9: Fingerprints from JPEG format images with different resampling factors, which are 50%, 70%, 90%, 110%, 200%. It should be noted that "Original" denotes non-resampled images.

accurate to model the pixel distribution of the original image without resampling operation as the Uniform distribution, which directly impacts the performance of our proposed test. Thus, it is very important to find the most accurate model fitting the pixel distribution of the non-resampled image in order to improve the detection accuracy of the test.

Inspired by our prior research in [183], it is proposed to model pixel distribution from a non-resampled/original JPEG format image by using the generalized signal-dependent noise model (see details in Section 6.2 of Chapter 6). Based on this statistical model, it is proposed to extract fingerprint from JPEG format images acquired by the same camera model, which can be used as intrinsic artifact to detect image resampling. First, it is assumed that resampling operation would unavoidably result in the change of camera fingerprint. Then, let us utilize 50 images from *Nikon D70* in JPEG format to verify our assumption. Figure 7.9 illustrates the change of camera fingerprints before and after resampling operation. Obviously, the camera fingerprints change their original positions due to post-camera operation such as resampling.

Therefore, based on the distinguishable property of camera fingerprints from non-resampled and resampled images, let us propose two methodologies to solve the problem of designing a test for resampling detection. In the first methodology, it is proposed to use the supervised algorithm. Camera fingerprints can be used as features in the training stage. In the testing stage, we extract the features from the inspected images and detect them using the trained classifier. Since the features are very distinguishable, especially in the case of resampling with small factors such as 50% or 70% (see Figure 7.9a), it is can be expected that the designed supervised detector could achieve high detection performance. However, the statistical performance of the trained classifier can only be known from the testing set, which can not be statistically established. In fact, all machine learning methods hardly

solve this open problem [27]. In the second methodology, it is proposed to design an unsupervised detector. Since the pixel distributions from a non-resampled and a resampled JPEG image have been both established, based on the pdf of pixels, it is very easy to design a LRT with knowing all the concerning model parameters. Moreover, the statistical performance of this optimal LRT can be served as an upper bound for detection power. In the practical detection, when all the model parameters are unknown, which can be successfully estimated by using the MLE algorithm (see details in Section 3 of [183]), we can design a GLRT to detect whether or not an inspected image has been resampled.

7.6 Conclusion

This chapter studies the linear correlation of 1-D and 2-D resampled signals. By using a linear parametric model, together with the Bayes' rule, we propose the faster algorithm of exposing the linear artifact. Numerical experiments verify that our proposed algorithm is capable of distinguishing the original signal and resampled one. Meanwhile, the up-sampled factor of a resampled 1-D signal can be estimated accurately. Without considering parameters estimation and iteration computation, compared with the state-of-art EM detector [68], our proposed algorithm improves the efficiency of computation. Similarly, the artifact from a resampled image can also be detected successfully. In the case of detecting a batch of images, based on the pdfs of pixels from a non-resampled and resampled image respectively, it is proposed to design the practical LRT detector for authenticating a resampled image.

In this chapter, the Uniform distribution describing the non-resampled 1-D signal or pixels of the non-resampled image is not very accurate. It can be replaced by the more accurate generalized model, which has been studied in [183]. Thus, the future work will be extended based on our discussion in Section 7.5.

Chapter 8

Conclusions and Perspectives

Contents

8.1 Conclusions

With the development of digital imaging technology, image tempering has become ubiquitous with the help of some low-cost image editing tools. To reestablish the trust of digital images, it is urgent to improve and extend the current research in the field of digital image forensics. In this thesis, we mainly address three important problems: image origin identification, detection of hidden information and image resampling detection.

In Chapter 1, we first generalize the research background of digital image forensics and draw a general outline of this thesis. Finally, we list all the contributions of the present PhD thesis that gave birth to publication, which shows the quality and the originality of the work presented in this thesis.

In Chapter 2, we first introduce the image acquisition pipeline of a digital still camera. Then we propose a general classification of all digital image falsifications. Based on the tempering techniques description a general classification of digital forensic tools, both active and passive, is proposed. Finally, this chapter also provides a brief overview on a specific digital image modification, that aims at preserving the content while hiding a secret message, and its detection, that is steganography and steganalysis.

All the original forensic detectors presented in this thesis are based on a statistical model of digital images. However, it has been chosen not to present all those models altogether but instead to present each statistical model of natural image pixels along with its application for digital image forensics. In Fourier domain, the Gaussian model is proposed to deal with the residual noise caused by CFA interpolation. Based on this model, an original forensic detector is designed

for distinguishing between PIM images and CG images in Chapter 3. In DCT domain, the Laplacian model is utilized for describing the statistical distribution of DCT coefficients. Based on this model we propose a steganalysis statistical test to detect secret message embedded in DCT coefficients with JSteg algorithm in Chapter 4. The main originality of this statistical test is to challenge the very common assumption that DCT coefficients from a subband are i.i.d. In the spatial domain, we first propose an enhanced Poissonian-Gaussian noise model, describing pixels from a RAW format image, based on which a statistical test for identifying source camera device is proposed in Chapter 5. Then, Chapter 6 presents the proposed improved signal-dependent noise model describing the pixels from a JPEG format image, which is exploited within a statistical test for solving the problem of source camera device identification from JPEG images. Finally it is proposed, in Chapter 7, to use a simpler Gaussian model describing the distribution of residual noise from the resampled image and to use it to design a statistical test for resampling detection. Hence, this thesis has studied all the three main problems of image forensics, as stating from the beginning:

1. image origin identification (in Chapter 3, 5 and 6) ;
2. detection of hidden information (in Chapter 4) ;
3. image resampling detection (in Chapter 7), which can be used as a first step for forgery detection.

Based on each of those different statistical models of natural images, each forensic detector is formally designed within the framework of hypothesis testing theory. First of all, it is assumed that all the model parameters are perfectly known. This theoretical setting is used to design the optimal LRT, which can maximize the detection power for a given false alarm probability, and to establish analytically its statistical performance. Besides, the statistical performance of this LRT can also be used as a theoretical upper bound of any test based on this model, and especially, the one proposed in practical settings. In the operational context, when all model parameters are not known, it is proposed to use the estimated parameters in the LRT to design practical detector. When the estimation is carried out with the method of Maximum Likelihood, such a practical test corresponds to the well celebrated GLRT (see Chapters 5 and 6). When the model parameters are estimated by using any other algorithm, the practical detector can be designed (see Chapter 4).

8.2 Perspectives

In this thesis, it is proposed to exploit accurate statistical noise models for solving the problem of source camera identification. The statistical model can also be utilized for image tempering detection such as splicing or copy-move forgery. In the tests for source camera identification, bi-criteria hypotheses test can be extended to the multi-hypotheses approach. Besides, the designed steganalyzer for JSteg algorithm

is based on the Laplacian model, it is reasonable that we can utilize the more accurate model for describing the DCT coefficients. In the following subsections, let us give a more detailed perspective of our future research for each of those directions.

8.2.1 Image Tempering Detection

Image splicing forgery (see Figure 2.5) is one of most ubiquitous tempering techniques. Splicing is a manipulation technique that duplicates one or more objects from one or more images and copies it or them onto another original host image. Thus, the pixel's distribution from the copied object probably acquired by another digital camera is inconsistent with the pixel's distribution from the background of the original host image which accounts for the vast majority of the new composite image pixels. Based on our proposed image statistical models, namely the Poissonian-Gaussian noise model for RAW images and its enhanced version, the generalized signal-dependent noise model for JPEG image as well as its improved version, it is reasonable to think that we could establish a statistical detector for image splicing detection. In fact, in the forensic community of splicing detection, most literature focus on designing a test for JPEG images. To the best of our knowledge, very few researchers explored the problem using RAW images. Therefore, the design of such a splicing detector for RAW images, based on the presented statistical models of image noise, will open a new way of image splicing detection. Recently, a pioneer work has been proposed for detecting the splicing falsification of a RAW image in [188].

Copy-move forgery is another popular tempering technique, which denotes that one object in an image is first copied, then pasted within the same image but, of course, at a different position (see Figure 2.4). Inspired by the above-proposed future application of image noise statistical model for splicing detection, let us extend the detection of splicing to copy-move forgery. Since the pixel's distribution of the copied part is definitely same as the distribution of its original part in the manipulated image, the similarity between two parts can help us establish the test based on a predictor of the correlation proposed in [47].

8.2.2 Image Origin Identification

In the test for identifying source camera device based on RAW images, see Chapter 5, the detector can only identify two camera devices, namely under hypothesis $\mathcal{H}_0$ or $\mathcal{H}_1$. In other words, if the inspected image is acquired neither from device 0 nor 1, the designed detector becomes invalid. Similarly, the first designed GLRT for identifying source camera based on JPEG images of Chapter 6 shares the same limitation. Even though the second GLRT can deal with the case in which the model parameters are known only for hypothesis $\mathcal{H}_0$, it is still limited to a specific device. In other words, if this image is not acquired by device 0, the forensic investigator has no idea of its origin. Hence, it would be interesting to study the extension of the proposed tests in a multi-hypotheses scenario, for testing several devices, see

for instance [142] for a practical application of multi-hypotheses testing. The multi-hypotheses testing framework can also be used to reject both hypotheses $\mathcal{H}_0$ and $\mathcal{H}_1$, see [189] for methodological details.

8.2.3 Steganalysis for JSteg Algorithm

In Chapter 4, the proposed JSteg steganalyzer is based on the Laplacian distribution model for the DCT coefficients even though it does not rely on the assumption that all those coefficients, in each subband, are i. i. d. However, the Laplacian model for describing the statistical distribution of DCT coefficients is probably less accurate than other models such as the Generalized Gaussian [150] and the Generalized Gamma [151]. Therefore, it is urgent to find the more accurate model which can lead us to design a better steganalyzer with higher detection performance. Recently, a more accurate model of DCT coefficients has been proposed to detect JSteg steganography [136]. Nevertheless, all these more accurate models assume that DCT coefficients within a subband are i. i. d. , which does not always hold true. Hence, it is reasonable to study the design of a JSteg detection method that leverages the accuracy of recent models of DCT coefficients and a methodology in which coefficients from a subband are not assumed to be i. i. d.

Appendix A

Appendix of Chapter 4

A.1 Quantized Laplacian PMF

Let X be a Laplacian random variable with expectation μ and variance b. Its pdf is thus, see (4.13):

$$f_{\mu,b}(x) = \frac{1}{2b}\exp\left(-\frac{|x-\mu|}{b}\right),$$

and a straightforward calculation shows that its cdf is given by:

$$F_{\mu,b}(x) = \frac{1}{2}+\frac{1}{2}\,\mathrm{sign}(x-\mu)\left(1-\exp\left(-\frac{|x-\mu|}{b}\right)\right), \tag{A.1}$$

$$= \begin{cases} \frac{1}{2}\exp\left(\frac{x-\mu}{b}\right) & \text{if } x < \mu, \\ 1-\frac{1}{2}\exp\left(-\frac{x-\mu}{b}\right) & \text{if } x \geq \mu. \end{cases} \tag{A.2}$$

Now consider the result from quantization of this random variable $Y = \lfloor X/\Delta \rfloor$, it is immediate to establish the pmf of this random variable. Let us first consider the case $\Delta(k+1/2) < \mu$ (due to the symmetry of Laplacian pdf, the case $\Delta(k-1/2) > \mu$ is treated similarly).
The pmf of Y is given by:

$$\begin{aligned} \mathbb{P}[Y=k] &= \mathbb{P}[\Delta(k-\frac{1}{2}) \leq X < \Delta(k+\frac{1}{2})], \\ &= \frac{1}{2}\exp\left(\frac{\Delta(k+\frac{1}{2})-\mu}{b}\right) \\ &\quad -\frac{1}{2}\exp\left(\frac{\Delta(k-\frac{1}{2})-\mu}{b}\right), \\ &= \frac{1}{2}\exp\left(\frac{\Delta k-\mu}{b}\right)\exp\left(\frac{\Delta}{2b}\right) \\ &\quad -\frac{1}{2}\exp\left(\frac{\Delta k-\mu}{b}\right)\exp\left(\frac{-\Delta}{2b}\right), \\ &= \exp\left(\frac{\Delta k-\mu}{b}\right)\sinh\left(\frac{\Delta}{2b}\right), \end{aligned}$$

Applying similar calculations for case $\Delta(k-1/2) > \mu$, one gets:

$$\mathbb{P}[Y=k] = \exp\left(-\frac{|\Delta k-\mu|}{b}\right)\sinh\left(\frac{\Delta}{2b}\right), \tag{A.3}$$

which corresponds to the pmf given in Eq. (4.14). The case $\Delta(k-1/2) < \mu < \Delta(k+1/2)$ is treated similarly.

A.2 Log-Likelihood Ratio Calculation

By putting the expression of quantized Laplacian pmf (4.14) into the expression of the LR (4.7), it is immediate to write:

$$\Lambda^{\mathrm{lr}}(u_{k,l}) = \log\left(1 - \frac{R}{2} + \frac{R}{2}\frac{\exp\left(-\frac{|\Delta\bar{k}-\mu|}{b}\right)\sinh\left(\frac{\Delta}{2b}\right)}{\exp\left(-\frac{|\Delta k-\mu|}{b}\right)\sinh\left(\frac{\Delta}{2b}\right)}\right).$$

Let us study the term:

$$\begin{aligned}
&\frac{\exp\left(-\frac{|\Delta\bar{k}-\mu|}{b}\right)\sinh\left(\frac{\Delta}{2b}\right)}{\exp\left(-\frac{|\Delta k-\mu|}{b}\right)\sinh\left(\frac{\Delta}{2b}\right)} = \frac{\exp\left(-\frac{|\Delta\bar{k}-\mu|}{b}\right)}{\exp\left(-\frac{|\Delta k-\mu|}{b}\right)},\\
=&\frac{\exp\left(-\frac{|\Delta k+\Delta(\bar{k}-k)-\mu|}{b}\right)}{\exp\left(-\frac{|\Delta k-\mu|}{b}\right)},\\
=&\frac{\exp\left(-\frac{|\Delta k-\mu|}{b}\right)\exp\left(\frac{\Delta\,\mathrm{sign}(\Delta k-\mu)(k-\bar{k})}{b}\right)}{\exp\left(-\frac{|\Delta k-\mu|}{b}\right)},\\
=&\exp\left(\frac{\Delta\,\mathrm{sign}(\Delta k-\mu)(k-\bar{k})}{b}\right).
\end{aligned} \tag{A.4}$$

From this Eq. (A.4), it is immediate to establish the expression (4.15):

$$\log\left(1 - \frac{R}{2} + \frac{R}{2}\exp\left(\frac{\Delta\,\mathrm{sign}(\Delta k-\mu)(k-\bar{k})}{b}\right)\right).$$

By using a Taylor expansion, $\Lambda^{\mathrm{lr}}(u_{k,l})$ can be approximated by:

$$\begin{aligned}
&\log\left(1-\frac{R}{2}+\frac{R}{2}\left(1+\frac{\Delta\,\mathrm{sign}(\Delta k-\mu)(k-\bar{k})}{b}\right)\right)\\
\approx&\log\left(1+\left(\frac{R\Delta\,\mathrm{sign}(\Delta k-\mu)(k-\bar{k})}{2b}\right)\right),\\
\approx&\frac{R\Delta}{2b}(k-\bar{k})\,\mathrm{sign}(\Delta k-\mu).
\end{aligned}$$

A.3 LR Based on the Gaussian Model (WS detector)

Let X be a Gaussian random variable with expectation μ and variance σ^2. Define the quantized Gaussian random variable as follows $Y = \lfloor X/\Delta \rfloor$, its pmf is given by

$P_{\mu,\sigma} = \{p_{\mu,\sigma}[k]\}_{k=-\infty}^{\infty}$ with:

$$p_{\mu,\sigma}[k] = \mathbb{P}[Y = k] = \int_{\Delta(k-1/2)}^{\Delta(k+1/2)} \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) dx.$$

Assuming that the quantization step Δ is "small enough" compared to the variance $\Delta << \sigma$, it holds true that [111, 130]:

$$p_{\mu,\sigma}[k] \approx \frac{\Delta}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(\Delta k - \mu)^2}{2\sigma^2}\right), \tag{A.5}$$

and

$$p_{\mu,\sigma}[k] + p_{\mu,\sigma}[\bar{k}] \approx \frac{2\Delta}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(\Delta\frac{(k+\bar{k})}{2} - \mu)^2}{2\sigma^2}\right). \tag{A.6}$$

Let us rewrite the LR for the detection of JSteg (5.13) as follows

$$\begin{aligned}\Lambda^{\mathrm{lr}}(u_{k,l}) &= \log\left(1 - \frac{R}{2} + \frac{R}{2}\frac{p_{\mu,\sigma}[\bar{k}]}{p_{\mu,\sigma}[k]}\right), \\ &= \log\left(1 - R + \frac{R}{2}\frac{p_{\mu,\sigma}[\bar{k}] + p_{\mu,\sigma}[k]}{p_{\mu,\sigma}[k]}\right).\end{aligned} \tag{A.7}$$

Using the expressions (A.5) and (A.6) let us study the following ratio:

$$\begin{aligned}\frac{p_{\mu,\sigma}[\bar{k}] + p_{\mu,\sigma}[k]}{p_{\mu,\sigma}[k]} &= 2\frac{\exp\left(-\frac{\Delta\frac{(k+\bar{k})}{2}-\mu)^2}{2\sigma^2}\right)}{\exp\left(-\frac{(\Delta k-\mu)^2}{2\sigma^2}\right)}, \\ &= 2\frac{\exp\left(-\frac{\left(\Delta k-\mu+\Delta/2(\bar{k}-k)\right)^2}{2\sigma^2}\right)}{\exp\left(-\frac{(\Delta k-\mu)^2}{2\sigma^2}\right)}, \\ &= 2\frac{\exp\left(-\frac{(\Delta k-\mu)^2}{2\sigma^2}\right)\exp\left(\frac{\Delta(\Delta k-\mu)(k-\bar{k})}{2\sigma^2}\right)\exp\left(-\frac{\Delta^2}{8\sigma^2}\right)}{\exp\left(-\frac{(\Delta k-\mu)^2}{2\sigma^2}\right)}, \\ &= 2\exp\left(\frac{\Delta(\Delta k-\mu)(k-\bar{k})}{2\sigma^2}\right)\exp\left(-\frac{\Delta^2}{8\sigma^2}\right).\end{aligned} \tag{A.8}$$

Putting the expression (A.8) into the expression of the log-LR (A.7) immediately gives:

$$\Lambda^{\mathrm{lr}}(u_{k,l}) = \log\left(1 + R\left(\exp\left(\frac{\Delta(\Delta k-\mu)(k-\bar{k})}{2\sigma^2}\right)\exp\left(-\frac{\Delta^2}{8\sigma^2}\right) - 1\right)\right) \tag{A.9}$$

from which a Taylor expansion around $\Delta/\sigma = 0$, this results from the assumption that $\Delta << \sigma$, and finally gives the well-known expression of the WS:

$$\Lambda^{\mathrm{lr}}(u_{k,l}) \approx \frac{R\Delta}{2\sigma^2}(k - \bar{k})(\Delta k - \mu) \tag{A.10}$$

Appendix B

Appendix of Chapter 6

B.1 Block Fingerprints $(\widetilde{a}_k, \widetilde{b}_k)$ Estimation

In this appendix, it is proposed to detail the estimation of $(\widetilde{a}_k, \widetilde{b}_k)$ by using Least Square algorithm (LS). Then let us transform Equation (6.8) to:

$$p_{k,n} = q_{k,n}\widetilde{a}_k + \widetilde{b}_k \tag{B.1}$$

where $p_{k,n} = (\sigma^2_{k,n} - \frac{\Delta}{12})\frac{\gamma}{\mu^{2-2\gamma}_{k,n}}$ where γ can be previously estimated using the algorithm proposed in [183] and $q_{k,n} = \mu_{k,n}$, the block index $k \in \{1, \cdots, K\}$, the image index $n \in \{1, \cdots, N\}$ with N denoting the number of JPEG images. Moreover, one can express Equation (B.1) in the form of vector like:

$$\begin{aligned} \mathbf{P} &= \mathbf{Q} \cdot \begin{pmatrix} \widetilde{a}_k \\ \widetilde{b}_k \end{pmatrix} \\ \text{where} \quad \mathbf{P} &= \begin{pmatrix} p_{k,1} \\ \vdots \\ p_{k,N} \end{pmatrix}, \mathbf{Q} = \begin{pmatrix} q_{k,1} & 1 \\ \vdots & \vdots \\ q_{k,N} & 1 \end{pmatrix}. \end{aligned} \tag{B.2}$$

Then, we can estimate the camera device fingerprints by:

$$\begin{pmatrix} \widehat{\widetilde{a}}_k \\ \widehat{\widetilde{b}}_k \end{pmatrix} = (\mathbf{Q}^{\mathrm{T}}\mathbf{Q})^{-1}\mathbf{Q}^{\mathrm{T}}\mathbf{P} \tag{B.3}$$

where $\mathbf{Q}^{\mathrm{T}}$ denotes the transpose of $\mathbf{Q}$.

B.2 Expectation and Variance of LR $\Lambda^{\mathrm{lr}}(\mathbf{e}_k)$

In this appendix, let us discuss the expectation $E_{\mathcal{H}_j}(\Lambda^{\mathrm{lr}}(\mathbf{e}_k))$ and the variance $V_{\mathcal{H}_j}(\Lambda^{\mathrm{lr}}(\mathbf{e}_k))$ of the LR $\Lambda^{\mathrm{lr}}(\mathbf{e}_k)$ under hypothesis $\mathcal{H}_j, j = \{0, 1\}$. First, under hypothesis $\mathcal{H}_0$ with camera device 0 fingerprints $(\widetilde{a}_{k,0}, \widetilde{b}_{k,0})$, we can immediately obtain the following Gaussian distribution:

$$\begin{cases} e_{k,0} = \widetilde{a}_{k,0} - (c_0\widetilde{b}_{k,0} + d_0) \sim \mathcal{N}(0, \sigma^2_0), \\ e_{k,1} = \widetilde{a}_{k,0} - (c_1\widetilde{b}_{k,0} + d_1) \sim \mathcal{N}(\mu_{0,1}, \sigma^2_{0,1}). \end{cases} \tag{B.4}$$

By putting the expression of the distribution from $e_{k,0}$ and $e_{k,1}$ respectively (B.4) into the expression LR $\Lambda^{\text{lr}}(\mathbf{e}_k)$ (6.18), the expectation $E_{\mathcal{H}_0}(\Lambda^{\text{lr}}(\mathbf{e}_k))$ and the variance $V_{\mathcal{H}_0}(\Lambda^{\text{lr}}(\mathbf{e}_k))$ can be calculated by:

$$\begin{aligned} E_{\mathcal{H}_0}(\Lambda^{\text{lr}}(\mathbf{e}_k)) &= \log(\frac{\sigma_0}{\sigma_1}) \\ &\quad + \frac{1}{2}\left(E_{\mathcal{H}_0}\left(\frac{e_{k,0}^2}{\sigma_0^2}\right) - E_{\mathcal{H}_0}\left(\frac{e_{k,1}^2}{\sigma_1^2}\right)\right) \\ &= \log(\frac{\sigma_0}{\sigma_1}) + \frac{1}{2}\left(1 - \frac{\sigma_{0,1}^2+\mu_{0,1}^2}{\sigma_1^2}\right), \end{aligned} \tag{B.5}$$

$$\begin{aligned} V_{\mathcal{H}_0}(\Lambda^{\text{lr}}(\mathbf{e}_k)) &= \frac{1}{4}\left(V_{\mathcal{H}_0}\left(\frac{e_{k,0}^2}{\sigma_0^2}\right) + V_{\mathcal{H}_0}\left(\frac{e_{k,1}^2}{\sigma_1^2}\right)\right) \\ &= \frac{1}{4}\left(2 + \frac{V_{\mathcal{H}_0}(e_{k,1}^2)}{\sigma_1^4}\right) \\ &= \frac{1}{2}\left(1 + \frac{\sigma_{0,1}^4 - 2\mu_{0,1}^2\sigma_{0,1}^2}{\sigma_1^4}\right). \end{aligned} \tag{B.6}$$

Similarly, under hypothesis $\mathcal{H}_1$ with camera device 1 fingerprints $(\widetilde{a}_{k,1}, \widetilde{b}_{k,1})$, we can immediately obtain the following Gaussian distribution:

$$\begin{cases} e_{k,0} = \widetilde{a}_{k,1} - (c_0\widetilde{b}_{k,1} + d_0) \sim \mathcal{N}(\mu_{1,0}, \sigma_{1,0}^2), \\ e_{k,1} = \widetilde{a}_{k,1} - (c_1\widetilde{b}_{k,1} + d_1) \sim \mathcal{N}(0, \sigma_1^2). \end{cases} \tag{B.7}$$

The expectation $E_{\mathcal{H}_1}(\Lambda^{\text{lr}}(\mathbf{e}_k))$ and the variance $V_{\mathcal{H}_1}(\Lambda^{\text{lr}}(\mathbf{e}_k))$ can be expressed by:

$$E_{\mathcal{H}_1}(\Lambda^{\text{lr}}(\mathbf{e}_k)) = \log(\frac{\sigma_0}{\sigma_1}) + \frac{1}{2}\left(\frac{\sigma_{1,0}^2+\mu_{1,0}^2}{\sigma_0^2} - 1\right). \tag{B.8}$$

$$\begin{aligned} V_{\mathcal{H}_1}(\Lambda^{\text{lr}}(\mathbf{e}_k)) &= \frac{1}{4}\left(2 + \frac{V_{\mathcal{H}_1}(e_{k,0}^2)}{\sigma_0^4}\right) \\ &= \frac{1}{2}\left(\frac{\sigma_{1,0}^4 - 2\mu_{1,0}^2\sigma_{1,0}^2}{\sigma_0^4} + 1\right). \end{aligned} \tag{B.9}$$

B.3 Expectation and Variance of GLR $\widehat{\Lambda}_1^{\text{lr}}(\mathbf{e}_k)$

In this practical GLRT (6.33), we never know the statistical parameter of the errors $\mathbf{e}_k$, that is variance value. Thus, let us assume that the ML estimated variance $\widehat{\sigma}_j^2$ can be obtained by using Equation (6.28). In addition, under hypothesis $\mathcal{H}_j$ of the GLRT $\widehat{\delta}_1^{\text{lr}}$, camera parameters $(\widetilde{a}_{k,j}, \widetilde{b}_{k,j})$ and linear model parameters (c_j, d_j) are perfectly known. Therefore, for simplicity, the expectation and variance of $\widehat{\Lambda}_1^{\text{lr}}(\mathbf{e}_k)$

can be obtained by putting $(\widetilde{a}_{k,j}, \widetilde{b}_{k,j})$ and $\widehat{\sigma}_j^2$ into Equations (B.5), (B.6), (B.8), (B.9), which are expressed by:

$$E_{\mathcal{H}_0}(\widehat{\Lambda}_1^{\mathrm{lr}}(\mathbf{e}_k)) = \log(\frac{\widehat{\sigma}_0}{\widehat{\sigma}_1}) + \frac{1}{2}\left(1 - \frac{\widehat{\sigma}_{0,1}^2 + \widehat{\mu}_{0,1}^2}{\widehat{\sigma}_1^2}\right), \tag{B.10}$$

$$V_{\mathcal{H}_0}(\widehat{\Lambda}_1^{\mathrm{lr}}(\mathbf{e}_k)) = \frac{1}{4}\left(2 + \frac{V_{\mathcal{H}_0}(e_{k,1}^2)}{\widehat{\sigma}_1^4}\right), \tag{B.11}$$

where under hypothesis $\mathcal{H}_0$, $\widehat{\mu}_{0,1} = \frac{1}{K}\sum_{k=1}^{K}(\widetilde{a}_{k,0} - c_1\widetilde{b}_{k,0} - d_1)$ and $\widehat{\sigma}_{0,1}^2 = \frac{1}{K-1}\sum_{k=1}^{K}(\widetilde{a}_{k,0} - c_1\widetilde{b}_{k,0} - d_1 - \widehat{\mu}_{0,1})^2$ and $V_{\mathcal{H}_0}(e_{k,1}^2) = \frac{1}{K-1}\sum_{k=1}^{K}((\widetilde{a}_{k,0} - c_1\widetilde{b}_{k,0} - d_1)^2 - \frac{1}{K}\sum_{k=1}^{K}(\widetilde{a}_{k,0} - c_1\widetilde{b}_{k,0} - d_1)^2)^2$;

$$E_{\mathcal{H}_1}(\widehat{\Lambda}_1^{\mathrm{lr}}(\mathbf{e}_k)) = \log(\frac{\widehat{\sigma}_0}{\widehat{\sigma}_1}) + \frac{1}{2}\left(\frac{\widehat{\sigma}_{1,0}^2 + \widehat{\mu}_{1,0}^2}{\widehat{\sigma}_0^2} - 1\right), \tag{B.12}$$

$$V_{\mathcal{H}_1}(\widehat{\Lambda}_1^{\mathrm{lr}}(\mathbf{e}_k)) = \frac{1}{4}\left(\frac{V_{\mathcal{H}_1}(e_{k,0}^2)}{\widehat{\sigma}_0^4} + 2\right), \tag{B.13}$$

where under hypothesis $\mathcal{H}_1$, $\widehat{\mu}_{1,0} = \frac{1}{K}\sum_{k=1}^{K}(\widetilde{a}_{k,1} - c_0\widetilde{b}_{k,1} - d_0)$ and $\widehat{\sigma}_{1,0}^2 = \frac{1}{K-1}\sum_{k=1}^{K}(\widetilde{a}_{k,1} - c_0\widetilde{b}_{k,1} - d_0 - \widehat{\mu}_{1,0})^2$ and $V_{\mathcal{H}_1}(e_{k,0}^2) = \frac{1}{K-1}\sum_{k=1}^{K}((\widetilde{a}_{k,1} - c_0\widetilde{b}_{k,1} - d_0)^2 - \frac{1}{K}\sum_{k=1}^{K}(\widetilde{a}_{k,1} - c_0\widetilde{b}_{k,1} - d_0)^2)^2$.

B.4 Expectation and Variance of GLR $\widehat{\Lambda}_2^{\mathrm{lr}}(\mathbf{e}_k)$

In the case of GLRT (6.43), the unknown parameters (c_1, d_1) and $(\widetilde{a}_{k,1}, \widetilde{b}_{k,1})$ can be estimated by using several tested images. Let us estimate parameters (c_1, d_1) by using LS algorithm, see B.1, and estimate $(\widetilde{a}_{k,1}, \widetilde{b}_{k,1})$ by using the proposed algorithm of block fingerprints estimation in subsection 6.2.2. Then, for clarity, it is proposed to replace the unknown parameters by their corresponding estimates and put them into the calculations of expectation and variance of GLR $\widehat{\Lambda}_1^{\mathrm{lr}}(\mathbf{e}_k)$, see Equation (B.10)$\sim$(B.13). Finally, it is very easy to obtain the expectation and variance of GLR $\widehat{\Lambda}_2^{\mathrm{lr}}(\mathbf{e}_k)$.

Appendix C

Appendix of French Abstract: Détection statistique pour la criminalistique des images numériques

Contents

C.1 Introduction

Dans les dernières décennies, nous avons observé l'évolution remarquable des technologies informatiques et technologies d'imagerie numérique qui permet l'omniprésence des images numériques dans le monde. En raison d'une grande disponibilité d'outils de retouche d'image à faible coût, les images numériques peuvent être retouchées ou modifiées facilement. La modification des images numériques est devenue une réalité incontournable. Ces modifications peuvent être relativement anodines (retoucher l'apparence d'une personne pour lui enlever des imperfections cutanées) ou dérangeantes (faire disparaître les défauts d'un objet en vente en ligne). Malheureusement, les modifications peuvent conduire à des situations plus graves, par exemple lorsque les images falsifiées qui ont été utilisées comme preuve dans une salle d'audience pourraient induire en erreur le jugement et conduire à une peine d'emprisonnement pour les innocents ou la liberté pour le coupable. Autrement dit, cette évolution technologique peut fournir des moyens pour des fins malveillantes,

ce qui peut entraîner des conséquences importantes au niveau politique, économique et social.

Afin de rétablir la crédibilité des images numériques, le domaine de la criminalistique numérique des images (digital image forensics en anglais) s'est développé. En raison de l'importance de la sécurité de l'information dans de nombreux domaines, la criminalistique numérique des images a attiré une grande attention des chercheurs scientifiques ainsi que des forces de sécurité et les services de renseignement. Dans ce domaine, on identifie deux problèmes clés: l'identification de l'origine de l'image et la détection de la falsification dans une image. L'identification de l'origine de l'image vise à vérifier si une image donnée est acquise par un appareil photographique spécifique (c-a-d une instance), ou déterminer le modèle ou la marque d'appareil utilisé. La détection de la falsification vise à détecter tout acte de manipulation effectué dans une image comme l'épissage ou l'amélioration du contraste.

De manière générale, il existe deux approches pour résoudre ces problèmes. L'approche active, c'est-à-dire le tatouage, consiste à générer extrinsèquement des mesures de sécurité et les ajouter à l'image lors de la formation d'image dans l'appareil photographique. En raison de nombreuses contraintes, l'application de l'approche active est très limitée aux environnements spécifiques qui nécessitent un haut niveau de sécurité tels que les appareils photographiques prennant des photos des scènes de crime. De nombreux appareils photographiques sur le marché ne sont pas équipés de ces technologies. Par conséquent, il est souhaitable d'établir un cadre nouveau pour résoudre les problèmes dans leur ensemble. Dans la dernière décennie, l'approche passive a été de plus en plus étudiée. L'approche passive n'impose aucune contrainte, ni nécessite aucune information préalable, y compris l'image originale, dans son mode de fonctionnement. Les analystes criminalistiques disposent seulement de l'image suspecte and ils doivent explorer le plus possible d'informations utiles à partir de cette image. L'approche passive peut être largement appliquée à des millions d'images qui circulent quotidiennement sur les réseaux de communication.

L'insertion du message secret dans une image numérique est également un type de manipulation. Cela concerne le domaine de la stéganographie consistant à dissimuler des messages secrets entre deux parties de façon sécurisée, sans éveiller les soupçons de l'adversaire. Cependant, le concept de stéganographie a été abusé par des éléments anti-sociaux, criminels ou terroristes, ce qui a affecté gravement la sécurité intérieure et la défense nationale. Il est urgent de construire des outils fiables pour détecter des informations cachées et/ou révéler leur contenu. Cette tâche est appelée stéganalyse, qui est une des tâches importantes des analystes criminalistiques.

Les travaux présentés dans ce manuscrit concernent deux problèmes importants dans le domaine de la criminalistique numérique des images : l'identification de l'origine de l'image et la détection d'informations cachées dans une image. Contrairement à la plupart des méthodes antérieures construites dans le cadre de la classification supervisée, les travaux s'inscrivent dans le cadre de la théorie de la décision statistique, permettant de respecter une contrainte sur la probabilité de fausse

alarme. En outre, afin d'atteindre une performance de détection élevée en pratique, il est proposé d'exploiter des propriétés des images naturelles en modélisant les principales étapes de la chaîne d'acquisition d'un appareil photographique en établissant un modèle d'image pertinent.

ce chapitre est organisé comme suit :

- La section C.2 étudie la discrimination entre les images numériques naturelles et les images de synthèse. La méthode proposée exploite les traces d'interpolation du filtre CFA présentes dans les images naturelles. En utilisant la théorie des tests d'hypothèses, la méthode proposée permet de distinguer les images naturelles des images de synthèse en garantissant un taux de fausse alarme prescrit. Les résultats expérimentaux montrent l'efficacité de l'approche proposée et la grande robustesse de la méthode par rapport aux techniques cherchant à contrer cette discrimination.

- La section C.3 explore la détection statistique de la stéganographie des images JPEG. L'approche est basée sur un modèle statistique de coefficients DCT. Le problème de la détection de l'information cachée est étudié dans le cadre de la théorie des tests d'hypothèses. Dans un contexte idéal où tous les paramètres du modèle sont parfaitement connus, le test du rapport de vraissemblance est présenté et ses performances sont théoriquement établies. Pour une utilisation pratique où les paramètres du modèle sont inconnus, en exploitant une méthode de sélection ad hoc des coefficients DCT, un détecteur basé sur l'estimation de ces paramètres est proposé. La perte de puissance du détecteur, par rapport au test optimal est faible. Les résultats numériques montrent la pertinence de l'approche proposée.

- La section C.4 traite du problème d'identification du dispositif d'imagerie source à partir d'images naturelles au format RAW. Pour cela, il est proposé d'exploiter un modèle statistique du bruit. Les paramètres du modèle sont considérées comme des empreintes digitales uniques et par conséquent utilisés pour identifier le dispositif d'acquisition. Encore une fois, le problème est traité dans le cadre de la théorie des tests d'hypothèses. Le test du rapport de vraissemblance est présenté et sa performance est théoriquement établie. Afin de traiter les paramètres de nuisance inconnus, un test pratique sur la base de l'estimation de ces paramètres est conçu. Les résultats numériques montrant la pertinence de l'approche proposée sont présentés pour conclure cette section.

- La section C.5 aborde le problème de l'identification de dispositif d'acquisition pour des images naturelles compressées selon la norme JPEG. Cette approche est également basée sur un modèle statistique du bruit. On montre qu'en raison de la réponse non linéaire de pixels les paramètres du modèle de bruit sont liés linéairement. Cette relation est utilisée pour identifier le d caméra source. Puis, la même approche est proposé, le problème est coulé dans le

cadre de la théorie des tests d'hypothèses. Le TLR est présentée et sa performance est théoriquement établi. Puis, lorsque les paramètres de nuisance sont inconnus, deux tests pratiques sont proposées sur la base de l'estimation de ces paramètres. Les résultats numériques sur données simulées et des images naturelles réels sont présentés pour soutenir la pertinence de l'approche proposée.

- La section C.6 explore la caractéristique périodique du de rééchantillonnage d'un signal. Cette caractéristique peut être modélisé à l'aide de modèle paramétrique linéaire. Après avoir traité ce paramètre de nuisance, il est proposé d'utiliser la probabilité conditionnelle des bruits résiduels comme empreinte digitale unique du signal rééchantillonné. Ensuite, il est proposé d'étudier l'artefact de rééchantillonnage d'une image. Des expériences numériques montrent la pertinence de l'algorithme proposé avec une faible complexité et une grande efficacité.

- La section C.7 conclut le chapitre par une brève synthèse des travaux présentés et des perspectives ouverts.

C.2 Authentification des images naturelles

Le problème de la distinction d'une image photographique et d'une image générée par ordinateur est l'un des axes de recherche de la criminalistique de l'image numérique.

Pour distingeur ces deux types d'images, la plupart des méthodes proposent d'exploiter des statistiques d'ordre élevé et/ou des caractéristiques physiques en utilisant des méthodes d'apprentissage supervisées. Bien que ces méthodes peuvent atteindre une précision de détection élevée, l'apprentissage statistique supervisé prend du temps. En outre, plusieurs problèmes tels que l'inadéquation de jeu de tests et l'établissement de la performance de détection restent ouverts.

Les images naturelles sont obtenues à partir d'un processus d'imagerie complexe, alors que les images de synthèse sont générées par un logiciel graphique, et non par un appareil photographique. La figure C.1 illustre le système d'acquisition d'une image à partir d'un appareil photographique. Les photons qui rayonnent à partir d'un objet rentrent dans le système optique. Ensuite, le CFA filtre le spectre de la lumière de telle sorte que chacun des pixels représente un canal de couleur (rouge, vert ou bleu); les deux canaux de couleur manquants sont complétées par interpolation. Enfin, une image naturelle est générée après plusieurs post-traitements de l'image tels que la balance des blancs et la correction gamma.

La figure C.2 montre le filtre CFA plus largement appelé filtre de *Bayer*. Basé sur les caractéristiques du filtre de Bayer, les images naturelles et les images de synthèse peuvent être différenciées par la présence ou non d'un pic dans le domaine fréquenciel. Pour un grand nombre d'images naturelle, le pic n'est pas présent. Par conséquent, cela permet pas de les distinguer des images de synthèse. Ainsi, une amélioration de la méthode proposée dans [54] est nécessaire.

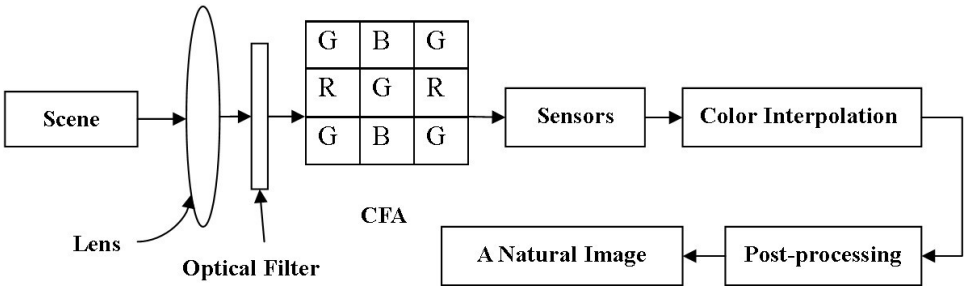


Figure C.1: Illustration du système d'acquisition d'une image numérique.

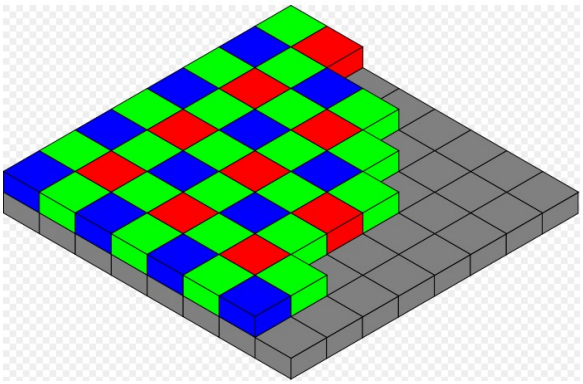


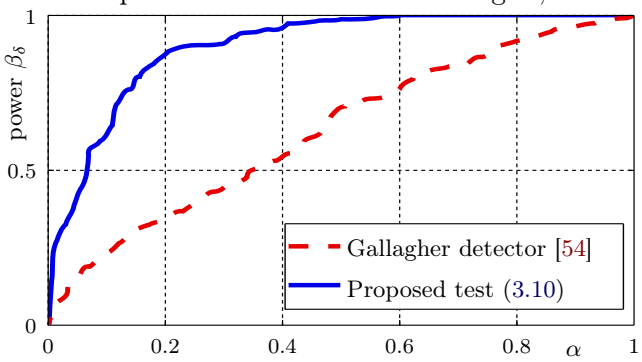
Figure C.2: Filtre de Bayer

L'algortithme proposé permet d'améliorer le procédé de [54] par deux moyens. Tout d'abord, il est proposé d'utiliser la variance dans le domaine fréquentiel sous l'hypothèse selon laquelle les post-traitements réduisent la variance. D'autre part, un modèle paramétrique linéaire est utilisé pour traiter les paramètres de nuisance et un test d'hypothèse est établi. Les résultats expérimentaux montrent l'efficacité et la robustesse de l'algorithme par rapport à la méthode proposée dans [54].

Pour vérifier les résultats théoriquement établis, une simulation de Monte-Carlo est effectuée. Il est proposé d'utiliser une base de données d'image contenant 300 images naturelles (avec 150 images de Nikon D70 et 150 images de Canon 10D) et 300 images générées par oridnateur téléchargées à partir de www.pandromeda.com. Toutes ces 600 images sont enregistrées au format JPEG avec un facteur de qualité de 85.

Afin de souligner l'amélioration du test proposé par rapport à la méthode de Gallagher proposé dans [54], la figure C.3 présente les performances de détection de ces détecteurs sous la forme de courbes ROC. On peut noter que, par exemple, pour $\alpha = 0,2$, la puissance du test proposé est au-dessus de $0,85$ alors qu'elle est inférieure à $0,35$ en utilisant la méthode de Gallagher. En outre, lorsque les images détectées avec une faible qualité d'image, le test proposé conserve une performance de détection élevé. Cela est souligné dans la figure C.3, qui présente les courbes ROC pour des images non compressées et des images compressées avec un facteur

Comparison de la performance des 2 méthodologies, Courbes ROC.



Performance de détection pour differents facteur de qualité d'images JPEG.

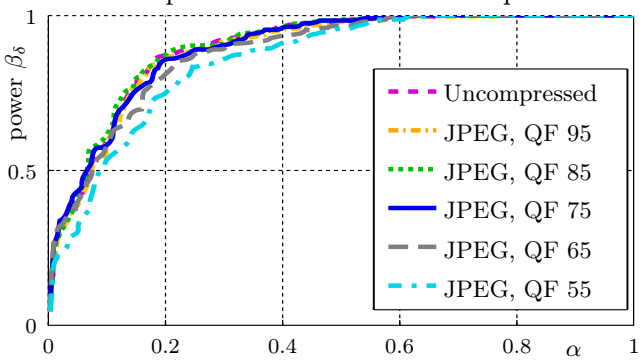


Figure C.3: Illustration de la performance du test.

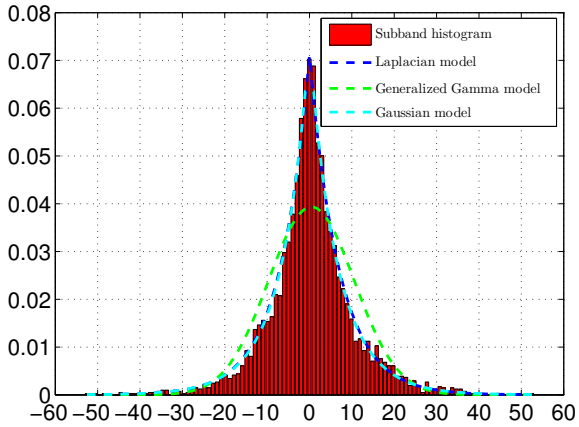
de qualité JPEG allant de 55 à 95.

C.3 Stéganalyse d'images JPEG

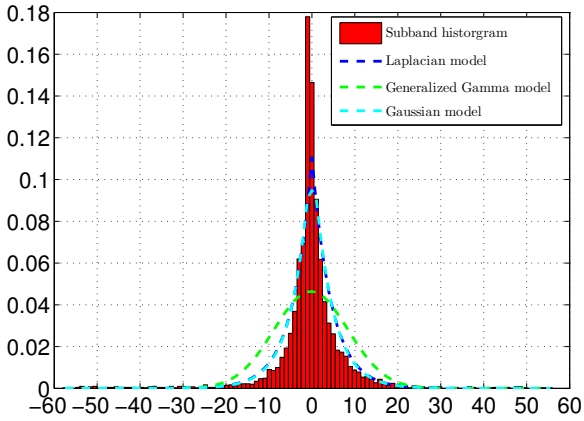
La stéganographie et la stéganalyse ont reçu de plus en plus d'attention au cours des deux dernières décennies. La stéganographie est l'art et la science de cacher des messages secrets dans des médias numériques. A l'inverse, la stéganalyse cherche à détecter les informations cachées dans les médias.

Il est à noter que de nombreuses approches ont été proposées pour la détection des données cachées dans les coefficients DCT des images JPEG, pour ne citer que quelques-uns, la détection structurelle [116], le détecteur WS [123], et les détecteurs universels ou aveugles [64,138]. Toutefois, l'établissement des propriétés statistiques de ces détecteurs reste un travail difficile qui n'a pas encore été étudié. En outre, les détecteurs les plus précis basés sur l'apprentissage statistique sont sensibles au contenu des images [155]: la phase de d'apprentissage doit être effectuée avec précaution.

Dans ce contexte, le détecteur proposé dans [135] est une alternative intéressante; mais il est basé sur l'hypothèse que les coefficients DCT sont indépendants et identiquement distribués et, ont une espérance nulle. Dans la pratique, ce modèle n'est pas indépendant du contenu de l'image, ce qui donne de bons résultats



(a) *Baboon*



(b) *Sky*

Figure C.4: Illustration of the quantized DCT coefficient subband (2,1), (a) a high-texture image: *Baboon*, (b) a low-texture image: *Sky*.

que dans le cas d'images de haute texture (voir la figure C.4), mais il est plus ou moins réaliste dans le cas d'image de faible texture(voir la figure C.4). A l'inverse, il est proposé un modèle statistique supposant que chaque coefficient DCT a une espérance et une variance différentes. L'utilisation de ce modèle permet de concevoir le test du rapport de vraisemblance le plus puissant (LRT) lorsque les paramètres de la distribution (espérance et variance) sont connus. Ensuite, dans le cas pratique où l'on ne connait pas les paramètres, on utilise leurs estimations. En prenant en compte les paramètres estimés , il est montré que la perte de puissance par rapport à la détection optimale est faible.

Par conséquent, les contributions de ce section sont les suivantes:

1. Tout d'abord, un nouveau modèle des coefficients DCT est proposé; son originalité majeure réside dans le fait ce modèle ne suppose pas que tous les coefficients sont identiquement distribués.

2. Deuxièmement, en supposant que tous les paramètres sont connus, ce modèle

statistique des coefficients DCT est utilisé pour concevoir le test optimal pour détecter des données cachées dans les images JPEG.

3. En outre, en supposant que tous les paramètres ne sont pas connus, une approche simple est proposée pour estimer l'espérance de chaque coefficient en utilisant une estimation de l'espérance de pixels dans le domaine spatial; la variance est également estimée localement.

4. Le détecteur conçu est amélioré en utilisant une sélection des coefficeints DCT proposée dans [156, 157] qui sélectionne uniquement un sous-ensemble des coefficients DCT dans lesquels l'incorporation est plus probable.

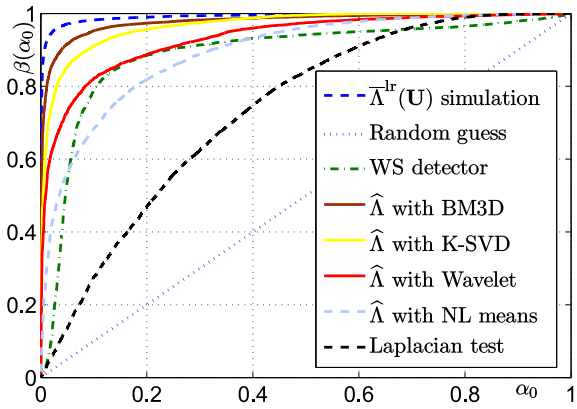
5. Les résultats numériques montrent la netteté des résultats théoriquement établis et la bonne performance du test statistique proposé. Une comparaison avec le test statistique basé sur la distribution Laplacienne (voir [135]), montre la pertinence de la méthode proposée. En outre, par rapport au détecteur WS [123], les résultats expérimentaux montrent l'efficacité du détecteur proposé.

Pour vérifier la pertinence de la méthode, il est proposé de comparer le test statistique avec deux autres détecteurs. Le premier concurrent choisi est le test statistique proposé dans [135]. En effet, ce test est également basé sur le modèle Laplacien, mais ne prend pas en compte les paramètres de nuisance; ce modèle considère également que les coefficients DCT sont identiquement distribués. La comparaison avec ce test est significative car elle nous permet de mesurer à quel point les performances de détection sont améliorées en éliminant l'hypothèse selon laquelle les coefficients DCT sont identidiquement distribués. Le deuxième concurrent choisi est le détecteur de WS [123] en raison de sa similitude avec le test statistique proposé.

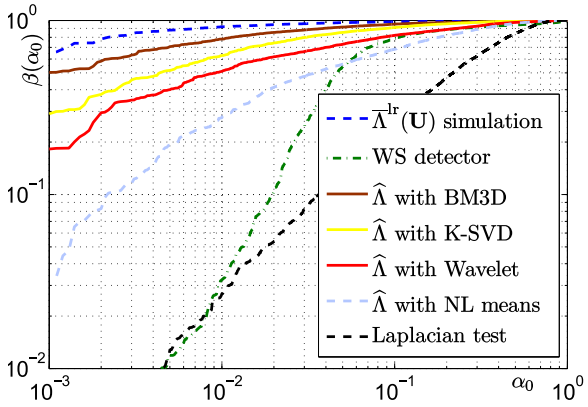
Pour une vérification à grande échelle, il est proposé d'utiliser la base de données BOSS, composée de 10 000 images en niveaux de gris de taille 512×512 pixels, utilisé avec un taux d'insertion $R = 0,05$. Quatre méthodes de débruitage ont été testés pour estimer l'espérance de chaque coefficient DCT, à savoir le K-SVD, le BM3D, les NL-moyens et les algorithmes de débruitage par ondelettes.

La figure C.5 montre les performances de détection obtenus sur la base de données BOSS compressée avec un facteur de qualité de 70, et les compare avec l'état de l'art. Les performances de détection sont représentés par les courbes ROC où la puissance de détection est tracée en fonction de la probabilité de fausse alarme. La figure C.5 souligne notamment que le test de [135] ne fonctionne pas bien alors que la méthode proposée, qui prend en compte les paramètres de nuisance nous permet d'améliorer largement la performance. De même, le détecteur WS réalise une bonne performance globale de détection. Cependant, il peut être indiqué sur la figure C.5, qui présente les mêmes résultats en utilisant une échelle logarithmique, que pour les faibles probabilités de fausses alarmes, la performance du WS diminue de manière significative. A l'inverse, le test statistique proposé remplit toujours bien son rôle.

Parmi les quatre algorithmes de débruitage qui ont été testés, le BM3D réalise la meilleure performance. Cependant, il peut être observé sur la figure C.5 que les



(a) Comparison of detection performance for BOSS database with quality factor 70 (linear scale).



(b) Comparison of detection performance for BOSS database with quality factor 70 (logarithmic scale).

Figure C.5: Comparison of detection performance for BOSS database with quality factor 70.

performances obtenues avec le K-SVD et en utilisant les méthodes de débruitage par ondelettes sont également très bon .

C.4 Identification du modèle d'appareil à partir d'une image RAW

La philosophie commune dans l'approche passive pour l'identification de l'origine de l'image est de s'appuyer sur des empreintes digitales engendrées par le processus d'acquisition photographique. En général, les empreintes sont des paramètres qui sont invariants par rapport au contenu de l'image et qui restent constants pour la même source (e.g. appareil individuel, modèle, ou marque). Chaque étape dans la chaîne d'acquisition de l'image peut fournir ces empreintes, par exemple l'aberration,

bruit de structure du capteur, structure de CFA et algorithmes de dématriçage, et JPEG compression. Malgré les méthodes proposées dans la littérature fonctionnent efficacement, leur performances statistiques ne sont jamais établies analytiquement et les méthodes sont évaluées seulement empiriquement sur une large base d'images. D'ailleurs, dans le context opérationnel, il est important également de respecter une contrainte sur la probabilité de fausse alarme, ce que les méthodes existantes ne peuvent pas permettre.

Au regard de l'identification des modèles d'appareil photographique, il est admis que l'ensemble des images $\mathcal{Z}^N$ peut être partitionné en un nombre fini de sous-espaces disjoints dans lequel les images de même caractéristiques issues de même modèle d'appareil photographique sont regroupées. Ainsi, le problème d'identification des modèles d'appareil photographique à partir d'une image $\mathbf{Z}$ peut être décrit comme un test d'hypothèse binaire

$$\begin{cases} \mathcal{H}_0 : & \mathbf{Z} \text{ est acquise par le modèle } \mathcal{S}_0 \\ \mathcal{H}_1 : & \mathbf{Z} \text{ est acquise par le modèle } \mathcal{S}_1 \text{ qui est différent de } \mathcal{S}_0, \end{cases} \tag{C.1}$$

où $\mathcal{S}_0$ et $\mathcal{S}_1$ sont deux modèles d'appareil photographique. Supposons que le modèle $\mathcal{S}_0$ est disponible en avance, il est souhaitable de construire un détecteur qui vérifie si l'image en question $\mathbf{Z}$ est capturée par le modèle $\mathcal{S}_0$.

Dans le cadre de la théorie de décision statistique, il est supposé que les observations $\mathbf{Z}$ dans chaque sous-ensemble $\mathcal{S}_j$, $j \in \{0, 1\}$, sont caractérisées par une loi de probabilité $P_{\boldsymbol{\theta}_j}$ où $\boldsymbol{\theta}_j$ est un paramètre vectoriel connu ou inconnu. Ainsi, le problème d'identification (C.1) peut être ré-écrit comme suit

$$\begin{cases} \mathcal{H}_0 : & \mathbf{Z} \sim P_{\boldsymbol{\theta}_0} \\ \mathcal{H}_1 : & \mathbf{Z} \sim P_{\boldsymbol{\theta}_1}. \end{cases} \tag{C.2}$$

Autrement dit, l'origine de l'image en question $\mathbf{Z}$ peut être identifiée via différentes propriétés statistiques dans les images naturelles acquises par différents modèles d'appareil photographique. Comme montré dans le problème (C.2), la pertinence de la distribution $P_{\boldsymbol{\theta}}$ qui est utilisée afin de modéliser une image naturelle est d'importance cruciale. Généralement, le paramètre $\boldsymbol{\theta}$ est composé d'un paramètre informatif $\boldsymbol{\lambda}$ et un paramètre de nuisance $\boldsymbol{\eta}$. Ce dernier est lié au contenu de l'image et n'a aucun intérêt pour l'identification des modèles d'appareil photographique. Par ailleurs, le paramètre informatif $\boldsymbol{\lambda}$ est invariant par rapport au contenu de l'image et discriminant pour différent modèles d'appareil photographique. Dans les travaux présentés, le paramètre informatif $\boldsymbol{\lambda}$ est utilisé comme unique empreinte sur laquelle les tests d'hypothèses statistiques s'appuient.

Le problème (C.2) révèle trois difficultés principales. Premièrement, du point de vue théorique, le test le plus puissant (PP) dans le context idéal où les paramètres $\boldsymbol{\theta}_0$ et $\boldsymbol{\theta}_1$ sont connus en avance, donné par le test du rapport de vraisemblance (RV), n'est jamais étudié. Ses performances statistiques ne sont jamais établies analytiquement et restent inconnues. La deuxième difficulté est la présence des paramètres de nuisance inconnus qui intervient dans la définition de la distribution $P_{\boldsymbol{\theta}}$ mais

ne présentent pas d'intérêt dans la prise de décision. Finalement, les empreints $\boldsymbol{\lambda}$ sont inconnus dans la pratique, ce qui implique que les hypothèses $\mathcal{H}_j$ deviennent composites. La construction d'un test optimal dans ce context reste discutable.

Tout d'abord, les travaux présentés dans ce manuscrit consistent à étudier le test RV lorsque les paramètres sont entièrement connus. Les propriétés statistiques de ce test sont ensuite analytiquement établies et servent de borne supérieure de la puissance qui peut être attendue de tout autre test. Afin de surmonter la difficulté des paramètres inconnus, il est proposé d'élaborer un test du rapport de vraisemblance généralisé (RVG) en estimant les paramètres inconnus par l'approche de maximum de vraisemblance (MV). Dans ce cas, deux tests statistiques sont proposés. Le premier test RVG considère le problème d'identification entre deux modèles d'appareils photographiques connus, i.e. les paramètres $\boldsymbol{\lambda}_0$ et $\boldsymbol{\lambda}_1$ sont connus tandis que les paramètres de nuisance $\boldsymbol{\eta}$ sont inconnus. Ce test peut être considéré comme un test d'hypothèse fermé car il s'agit de décider si une image provient d'un modèle d'appareil $\mathcal{S}_0$ connu ou bien d'un autre modèle $\mathcal{S}_1$ connu. Le second test aborde le problème d'identification lorsque seulement un modèle d'appareil photographique est connu, c'est à dire lorsque l'on souhaite déterminer si une image provient d'un modèle d'appareil $\mathcal{S}_0$ ou non. Comme le modèle d'appareil photographique $\mathcal{S}_0$ est supposé disponible, l'empreinte $\boldsymbol{\lambda}_0$ est connue en avance. D'ailleurs, l'empreinte $\boldsymbol{\lambda}_1$ et les paramètres de nuisance $\boldsymbol{\eta}$ sont inconnus. Ce second test est considéré comme un test d'hypothèse ouvert car l'image en question $\mathbf{Z}$ est autorisée d'être acquis par un modèle d'appareil photographique inconnu. Il est notable que la performance de ces tests d'hypothèses statistiques dépendrait de trois conditions suivantes :

- La pertinence du modèle d'image statistique.
- La discrimination des empreintes utilisées.
- La précision de l'estimation des paramètres lorsque ces derniers sont inconnus en pratique.

C.5 Identification du modèle d'appareil à partir d'une image JPEG

Tout d'abord, le problème d'identification (C.2) est approché par utiliser le modèle de bruit hétéroscédastique. Il est proposé de considérer les paramètres (a, b) de ce dernier comme l'empreinte pour identifier des modèles d'appareil photographique. Le modèle d'appareil photographique $\mathcal{S}_j$ est désormais caractérisé par deux paramètres (a_j, b_j) pour un nombre d'ISO fixé, $(a_0, b_0) \neq (a_1, b_1)$. Soit l'image RAW $\mathbf{Z} = (z_1, \ldots, z_N)$, le problème (C.2) peut être ré-écrit comme suit

$$\begin{cases} \mathcal{H}_0 & = \Big\{ z_i \sim \mathcal{N}(\mu_i, \sigma^2_{i,0}),\ \forall i \in \{1, \ldots, N\} \Big\} \\ \mathcal{H}_1 & = \Big\{ z_i \sim \mathcal{N}(\mu_i, \sigma^2_{i,1}),\ \forall i \in \{1, \ldots, N\} \Big\}. \end{cases} \tag{C.3}$$

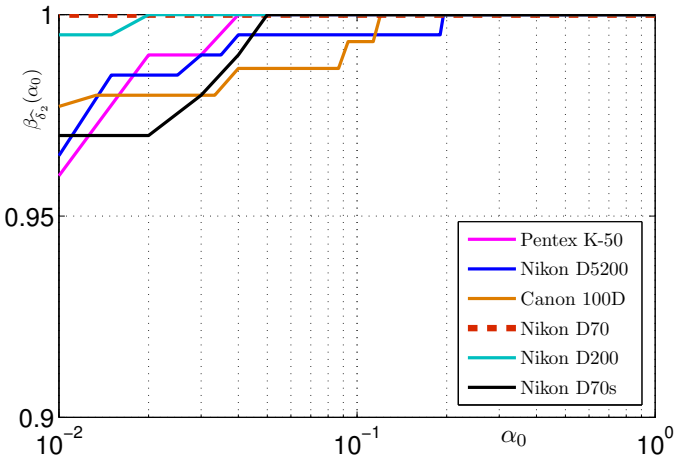


Figure C.6: Comparison of detection performance on a large scale database (logarithmic scale).

Dans le context idéal où les paramètres μ_i et (a_j, b_j) sont connus, le RV pour une observation z_i est défini par

$$\Lambda_{\text{het}}(z_i) = \frac{1}{2}h_1(\mu_i) + \frac{1}{2}h_2(\mu_i)(z_i - \mu_i)^2, \tag{C.4}$$

où

$$h_1(x) = \log(a_0 x + b_0) - \log(a_1 x + b_1) \quad \text{and} \quad h_2(x) = \frac{1}{a_0 x + b_0} - \frac{1}{a_1 x + b_1}, x \in \mathbb{R}_+.$$

Afin d'établir analytiquement les performances statistiques du test, il est nécessaire d'étudier la distribution statistique du RV $\Lambda_{\text{het}}(\mathbf{Z}) = \sum_{i=1}^{N} \Lambda_{\text{het}}(z_i)$. Cependant, il est difficile de définir la distribution exacte du $\Lambda_{\text{het}}(\mathbf{Z})$ à cause de l'hétérogénéité et la non-stationnarité des bruits d'acquisition. Une solution alternative est de reposer sur le théorème central limite (TCL) de Lindeberg [127, Theorem 11.2.5] en faveur d'un large nombre de pixels dans une image naturelle. Un avantage de ce théorème est que des variables aléatoires ne sont pas requis d'être distribué identiquement. Afin de pouvoir exploiter le TCL de Lindeberg, il est crucial de déterminer l'espérance et la variance du RV $\Lambda_{\text{het}}(z_i)$.

En partant de $z_i \sim \mathcal{N}(\mu_i, \sigma_{i,j}^2)$ sous chaque hypothesis $\mathcal{H}_j$ pour calculer l'espérance et la variance du RV $\Lambda_{\text{het}}(z_i)$ et utilisant le TCL de Lindeberg, la distribution statistique du RV $\Lambda_{\text{het}}(\mathbf{Z})$ sous l'hypothèse $\mathcal{H}_j$ est donnée par

$$\Lambda_{\text{het}}(\mathbf{Z}) \xrightarrow{d} \mathcal{N}\left(m_{\text{het}}^{(j)}, v_{\text{het}}^{(j)}\right) \tag{C.5}$$

où l'espérance $m_{\text{het}}^{(j)}$ and la variance $v_{\text{het}}^{(j)}$ sont données par

$$m_{\text{het}}^{(j)} = \sum_{i=1}^{N} \left[\frac{1}{2} h_1(\mu_i) + \frac{1}{2} h_2(\mu_i) \sigma_{i,j}^2 \right] \tag{C.6}$$

$$v_{\text{het}}^{(j)} = \sum_{i=1}^{N} \frac{1}{2} h_2^2(\mu_i) \sigma_{i,j}^4. \tag{C.7}$$

Puisqu'une image naturelle est hétérogène, il est proposé de normaliser le RV $\Lambda_{\text{het}}(\mathbf{Z})$ pour que le seuil de décision soit déterminé indépendamment du contenu de l'image. Le RV normalisé est défini par

$$\Lambda_{\text{het}}^{\star}(\mathbf{Z}) = \frac{\Lambda_{\text{het}}(\mathbf{Z}) - m_{\text{het}}^{(0)}}{\sqrt{v_{\text{het}}^{(0)}}}. \tag{C.8}$$

Par conséquence, le test RV $\delta_{\text{het}}^{\star}$ basé sur le RV normalisé $\Lambda_{\text{het}}^{\star}(\mathbf{Z})$ s'écrit comme suit

$$\delta_{\text{het}}^{\star}(\mathbf{Z}) = \begin{cases} \mathcal{H}_0 & \text{if} \quad \Lambda_{\text{het}}^{\star}(\mathbf{Z}) < \tau_{\text{het}}^{\star} \\ \mathcal{H}_1 & \text{if} \quad \Lambda_{\text{het}}^{\star}(\mathbf{Z}) \geq \tau_{\text{het}}^{\star} \end{cases} \tag{C.9}$$

où le seuil de décision $\tau_{\text{het}}^{\star}$ est la solution de l'équation $\mathbb{P}_{\mathcal{H}_0}\left[\Lambda_{\text{het}}^{\star}(\mathbf{Z}) \geq \tau_{\text{het}}^{\star}\right] = \alpha_0$. Le seuil de décision $\tau_{\text{het}}^{\star}$ et la puissance $\beta(\delta_{\text{het}}^{\star})$ sont donnés dans le théorème suivant.

Theorem C.1. *Dans un context idéal où tous les paramètres $(\boldsymbol{\mu}, a_j, b_j)$ sont connus en avance, le seuil de décision et la puissance du test RV $\delta_{\text{het}}^{\star}$ sont donnés par*

$$\tau_{\text{het}}^{\star} = \Phi^{-1}(1 - \alpha_0) \tag{C.10}$$

$$\beta(\delta_{\text{het}}^{\star}) = 1 - \Phi\left(\frac{m_{\text{het}}^{(0)} - m_{\text{het}}^{(1)} + \tau_{\text{het}}^{\star} \sqrt{v_{\text{het}}^{(0)}}}{\sqrt{v_{\text{het}}^{(1)}}} \right). \tag{C.11}$$

Un context plus réaliste considère que les paramètres de nuisance μ_i sont inconnus tandis que les empreintes (a_j, b_j) restent connus. Afin de construire un test RVG, il est nécessaire d'estimer les paramètres μ_i par la méthode MV. Pour ce faire, il est proposé de diviser l'image $\mathbf{Z}$ en K segments homogènes S_k de taille n_k sans chevauchement, $k \in \{1, \ldots, K\}$ (voir la technique de segmentation en [172]). Les pixels dans chaque segment S_k sont supposés indépendants et identiquement distribués. En utilisant l'estimateur MV $\hat{\mu}_k$, le RVG pour une observation $z_{k,i}^{\text{wapp}}$ est défini par

$$\overline{\Lambda}_{\text{het}}(z_{k,i}^{\text{wapp}}) = \frac{1}{2} h_1(\hat{\mu}_k) + \frac{1}{2} h_2(\hat{\mu}_k) \frac{(z_{k,i}^{\text{wapp}} - \hat{\mu}_k)^2}{\|\varphi\|_2^2}. \tag{C.12}$$

En utilisant le TCL de Lindeberg et prenant en compte des propriétés statistiques de l'estimateur MV $\hat{\mu}_k$, la distribution statistique du RVG $\overline{\Lambda}_{\text{het}}(\mathbf{Z}) = \sum_{k=1}^{K} \sum_{i=1}^{n_k} \overline{\Lambda}_{\text{het}}(z_{k,i}^{\text{wapp}})$ sous l'hypothesis $\mathcal{H}_j$ est donnée par

$$\overline{\Lambda}_{\text{het}}(\mathbf{Z}) \xrightarrow{d} \mathcal{N}\left(\overline{m}_{\text{het}}^{(j)}, \overline{v}_{\text{het}}^{(j)} \right), \tag{C.13}$$

où l'espérance $\overline{m}_{\text{het}}^{(j)}$ et la variance $\overline{v}_{\text{het}}^{(j)}$ sont données par

$$\overline{m}_{\text{het}}^{(j)} = \sum_{k=1}^{K}\sum_{i=1}^{n_k}\left[\frac{1}{2}h_1(\mu_k) + \frac{1}{2}h_2(\mu_k)\sigma_{k,j}^2\Big(1+\frac{1}{n_k}\Big)\right] \tag{C.14}$$

$$\begin{aligned}\overline{v}_{\text{het}}^{(j)} = \sum_{k=1}^{K}\sum_{i=1}^{n_k}\Bigg[&\frac{1}{4}\big(h_1'(\mu_k)\big)^2\frac{\|\varphi\|_2^2}{n_k}\sigma_{k,j}^2 + \frac{1}{2}\big(h_2(\mu_k)\big)^2\sigma_{k,j}^4\Big(1+\frac{1}{n_k}\Big)^2 \\ &+ \frac{3}{4}\big(h_2'(\mu_k)\big)^2\frac{\|\varphi\|_2^2}{n_k}\sigma_{k,j}^6\Big(1+\frac{1}{n_k}\Big)^2\Bigg].\end{aligned} \tag{C.15}$$

De façon similaire, il est proposé de normaliser le RVG $\overline{\Lambda}_{\text{het}}(\mathbf{Z})$, donné par

$$\overline{\Lambda}_{\text{het}}^{\star}(\mathbf{Z}) = \frac{\overline{\Lambda}_{\text{het}}(\mathbf{Z}) - \widehat{\overline{m}}_{\text{het}}^{(0)}}{\sqrt{\widehat{\overline{v}}_{\text{het}}^{(0)}}}, \tag{C.16}$$

où $\widehat{\overline{m}}_{\text{het}}^{(0)}$ and $\widehat{\overline{v}}_{\text{het}}^{(0)}$ sont des estimateurs de $\overline{m}_{\text{het}}^{(0)}$ and $\overline{v}_{\text{het}}^{(0)}$ en remplaçant μ_k par $\hat{\mu}_k$ en (C.14) and (C.15). Le test RVG correspondant $\overline{\delta}_{\text{het}}^{\star}$ est donné par la règle de décision suivante

$$\overline{\delta}_{\text{het}}^{\star}(\mathbf{Z}) = \begin{cases}\mathcal{H}_0 & \text{if} \quad \overline{\Lambda}_{\text{het}}^{\star}(\mathbf{Z}) < \overline{\tau}_{\text{het}}^{\star} \\ \mathcal{H}_1 & \text{if} \quad \overline{\Lambda}_{\text{het}}^{\star}(\mathbf{Z}) \geq \overline{\tau}_{\text{het}}^{\star}.\end{cases} \tag{C.17}$$

Theorem C.2. *En utilisant le théorème de Slutsky [?, theorem 11.2.11], le seuil de décision et la puissance du test RVG $\overline{\delta}_{\text{het}}^{\star}$ lors de la prise de décision entre deux modèles d'appareil photographique connus $\mathcal{S}_0$ and $\mathcal{S}_1$, i.e. les paramètres (a_0, b_0) and (a_1, b_1) sont connus, sont donnés par*

$$\overline{\tau}_{\text{het}}^{\star} = \Phi^{-1}(1-\alpha_0) \tag{C.18}$$

$$\beta(\overline{\delta}_{\text{het}}^{\star}) = 1 - \Phi\left(\frac{\overline{m}_{\text{het}}^{(0)} - \overline{m}_{\text{het}}^{(1)} + \overline{\tau}_{\text{het}}^{\star}\sqrt{\overline{v}_{\text{het}}^{(0)}}}{\sqrt{\overline{v}_{\text{het}}^{(1)}}}\right). \tag{C.19}$$

Le problème d'identification (C.3) devient plus délicat lorsque les paramètres (a_1, b_1) sont inconnus en pratique. Dans ce cas, il est proposé d'estimer ces derniers à partir de l'image en question $\mathbf{Z}$. Puisque l'estimateur MV de ces paramètres ne peut que être résolu numériquement [172], donc leur propriétés statistiques restent inconnues. Il est proposé d'estimer les paramètres (a_1, b_1) par la méthode des moindres carrés pondérés (Weighted Least Squares (WLS) en anglais). L'estimateur WLS est asymptotiquement équivalent à l'estimateur MV, donc il est également pertinent d'être utilisé pour la construction du test RVG. En utilisant l'estimateur MV $\hat{\mu}_k$ et l'estimateur WLS $(\hat{a}_1, \hat{b}_1)$, le RVG pour une observation $z_{k,i}^{\text{wapp}}$ est défini par

$$\begin{aligned}\widetilde{\Lambda}_{\text{het}}(z_{k,i}^{\text{wapp}}) = &\frac{1}{2}\Big(\log\left(a_0\hat{\mu}_k + b_0\right) - \log\left(\hat{a}_1\hat{\mu}_k + \hat{b}_1\right)\Big) \\ &+ \frac{1}{2}\Big(\frac{1}{a_0\hat{\mu}_k + b_0} - \frac{1}{\hat{a}_1\hat{\mu}_k + \hat{b}_1}\Big)\frac{(z_{k,i}^{\text{wapp}} - \hat{\mu}_k)^2}{\|\varphi\|_2^2}.\end{aligned} \tag{C.20}$$

Il est souhaitable de prendre en compte des propriétés statistiques de l'estimateur WLS $(\hat{a}_1, \hat{b}_1)$ dans l'établissement des performances statistiques du test RVG. Supposons que $\sigma^2_{a_1}$, $\sigma^2_{b_1}$, $\sigma_{a_1 b_1}$ représentent respectivement la variance d $\hat{a}_1$, la variance d $\hat{b}_1$ et la covariance entre $\hat{a}_1$ and $\hat{b}_1$. En calculant l'espérance et la variance du RVG $\widetilde{\Lambda}_{\text{het}}(z^{\text{wapp}}_{k,i})$ par la méthode de Delta [127, theorem 11.2.14]) et en utilisant le TCL de Lindeberg, un calcul immédiate montre que le RVG $\widetilde{\Lambda}_{\text{het}}(\mathbf{Z}) = \sum_{k=1}^{K}\sum_{i=1}^{n_k}\widetilde{\Lambda}_{\text{het}}(z^{\text{wapp}}_{k,i})$ suit également une distribution Gaussienne sous l'hypothèse $\mathcal{H}_j$

$$\widetilde{\Lambda}_{\text{het}}(\mathbf{Z}) \xrightarrow{d} \mathcal{N}\Big(\widetilde{m}^{(j)}_{\text{het}}, \widetilde{v}^{(j)}_{\text{het}}\Big) \tag{C.21}$$

où

$$\widetilde{m}^{(j)}_{\text{het}} = \sum_{k=1}^{K}\sum_{i=1}^{n_k}\mathbb{E}_{\mathcal{H}_j}\Big[\widetilde{\Lambda}_{\text{het}}(z^{\text{wapp}}_{k,i})\Big] \tag{C.22}$$

$$\widetilde{v}^{(j)}_{\text{het}} = \sum_{k=1}^{K}\sum_{i=1}^{n_k}\mathbb{V}\text{ar}_{\mathcal{H}_j}\Big[\widetilde{\Lambda}_{\text{het}}(z^{\text{wapp}}_{k,i})\Big]. \tag{C.23}$$

avec

$$\mathbb{E}_{\mathcal{H}_j}\Big[\widetilde{\Lambda}_{\text{het}}(z^{\text{wapp}}_{k,i})\Big] = \frac{1}{2}h_1(\mu_k) + \frac{1}{2}h_2(\mu_k)\sigma^2_{k,j}\Big(1+\frac{1}{n_{k,j}}\Big) \tag{C.24}$$

$$\begin{aligned}\mathbb{V}\text{ar}_{\mathcal{H}_j}\Big[\widetilde{\Lambda}_{\text{het}}(z^{\text{wapp}}_{k,i})\Big] &= \frac{1}{4}\Big[\frac{a_0^2}{\sigma^4_{k,0}}\frac{\|\varphi\|_2^2}{n_k}\sigma^2_{k,j} + \frac{\mathbb{V}\text{ar}_{\mathcal{H}_j}\big[\hat{a}_1\hat{\mu}_k+\hat{b}_1\big]}{\sigma^4_{k,1}}\Big] \\ &\quad + \frac{1}{2}\big(h_2(\mu_k)\big)^2\sigma^4_{k,j}\Big(1+\frac{1}{n_k}\Big)^2 \\ &\quad + \frac{3}{4}\sigma^4_{k,j}\Big(1+\frac{1}{n_k}\Big)^2\Big[\frac{a_0^2}{\sigma^8_{k,0}}\frac{\|\varphi\|_2^2}{n_k}\sigma^2_{k,j} + \frac{\mathbb{V}\text{ar}_{\mathcal{H}_j}\big[\hat{a}_1\hat{\mu}_k+\hat{b}_1\big]}{\sigma^8_{k,1}}\Big].\end{aligned} \tag{C.25}$$

et

$$\mathbb{V}\text{ar}_{\mathcal{H}_j}\big[\hat{a}_1\hat{\mu}_k+\hat{b}_1\big] = a_1^2\frac{\|\varphi\|_2^2}{n_k}\sigma^2_{k,j} + \Big(\mu_k^2 + \frac{\|\varphi\|_2^2}{n_k}\sigma^2_{k,j}\Big)\sigma^2_{a_1} + 2\mu_k\sigma_{a_1b_1} + \sigma^2_{b_1}. \tag{C.26}$$

Ainsi, le test RVG basé sur le RVG normalisé $\widetilde{\Lambda}^{\star}_{\text{het}}(\mathbf{Z})$ est donné par la règle de décision suivante

$$\widetilde{\delta}^{\star}_{\text{het}}(\mathbf{Z}) = \begin{cases}\mathcal{H}_0 & \text{if} \quad \widetilde{\Lambda}^{\star}_{\text{het}}(\mathbf{Z}) < \widetilde{\tau}^{\star}_{\text{het}} \\ \mathcal{H}_1 & \text{if} \quad \widetilde{\Lambda}^{\star}_{\text{het}}(\mathbf{Z}) \geq \widetilde{\tau}^{\star}_{\text{het}}\end{cases} \tag{C.27}$$

où

$$\widetilde{\Lambda}^{\star}_{\text{het}}(\mathbf{Z}) = \frac{\widetilde{\Lambda}_{\text{het}}(\mathbf{Z}) - \widehat{\widetilde{m}}^{(0)}_{\text{het}}}{\sqrt{\widehat{\widetilde{v}}^{(0)}_{\text{het}}}}, \tag{C.28}$$

et $\widehat{\widetilde{m}}^{(0)}_{\text{het}}$ and $\widehat{\widetilde{v}}^{(0)}_{\text{het}}$ sont obtenus par remplacer (μ_k, a_1, b_1) par $(\hat{\mu}_k, \hat{a}_1, \hat{b}_1)$ en (C.22) and (C.23).

Theorem C.3. *Selon le théorème de Slutsky, le seuil de décision et la puissance du test RVG $\widetilde{\delta}^{\star}_{\text{het}}$ pour vérifier si une image $\mathbf{Z}$ provient d'un modèle d'appareil photographique connu $\mathcal{S}_0$ sont donnés par*

$$\widetilde{\tau}^{\star}_{\text{het}} = \Phi^{-1}(1-\alpha_0) \tag{C.29}$$

$$\beta(\widetilde{\delta}^{\star}_{\text{het}}) = 1 - \Phi\left(\frac{\widetilde{m}^{(0)}_{\text{het}} - \widetilde{m}^{(1)}_{\text{het}} + \widetilde{\tau}^{\star}_{\text{het}}\sqrt{\widetilde{v}^{(0)}_{\text{het}}}}{\sqrt{\widetilde{v}^{(1)}_{\text{het}}}}\right). \tag{C.30}$$

Les résultats numériques montrent que la perte de puissance entre le test RV et les tests RVG est négligeable à partir d'un petit nombre de pixels (e.g. 500 pixels). En outre, en exploitant seulement quelques centaines de pixels des images numériques, les tests proposés obtiennent des performances quasiment parfaites pour une probabilité de fausse alarme faible, ce qui montre la force de cette approche. Plus important, il est également montré que les tests proposés permettent d'assurer le respect d'une probabilité de fausse alarme prescrite en pratique. Enfin, les résultats numériques sur une large base d'images réelles montrent également que les tests statistiques se comparent favorablement avec l'état de l'art en matière d'identification de la source d'une image numérique.

L'approche basée sur le modèle de bruit hétéroscédastique concerne deux limitations principales. La première est que cette approche est applicable seulement pour des images RAW qui ne pourraient pas être disponible en pratique. La deuxième est la dépendance en nombre d'ISO. Ces deux limitations seront adressées dans lex travaux subséquents.

C.6 Détection du rééchantillonnage d'une image numérique

Dans cette section, il est proposé d'exploiter le modèle de bruit généralisé afin d'identifier des modèles d'appareil photographique à partir d'une image JPEG, qui est le format typique auquel la plupart des appareils enregistrent. Le modèle d'appareil photographique $\mathcal{S}_j$ est désormais caractérisé par l'empreinte $(\tilde{a}_j, \tilde{b}_j, \gamma_j)$. Il est supposé que $(\tilde{a}_0, \tilde{b}_0, \gamma_0) \neq (\tilde{a}_1, \tilde{b}_1, \gamma_1)$. Soit donné une image JPEG $\widetilde{\mathbf{Z}}$, une étape préalable consiste à diviser l'image $\widetilde{\mathbf{Z}}$ en K segments homogènes sans chevauchement de taille $\tilde{n}_k$. Ainsi, le problème d'identification (C.2) consiste à décider entre deux hypothèses suivantes

$$\begin{cases} \mathcal{H}_0 = \left\{\tilde{z}_{k,i} \sim \mathcal{N}\big(\tilde{\mu}_k, \tilde{\sigma}^2_{k,0}\big), \forall k \in \{1,\dots,K\}, \forall i \in \{1,\dots,\tilde{n}_k\}\right\} \\ \mathcal{H}_1 = \left\{\tilde{z}_{k,i} \sim \mathcal{N}\big(\tilde{\mu}_k, \tilde{\sigma}^2_{k,1}\big), \forall k \in \{1,\dots,K\}, \forall i \in \{1,\dots,\tilde{n}_k\}\right\}. \end{cases} \tag{C.31}$$

Il est noté le problème (C.31) est similaire à (C.3) sauf que la variance $\tilde{\sigma}^2_{k,j}$ est désormais une fonction non-linéaire de l'espérance $\tilde{\mu}_k$, ce qui caractérise la non-stationnarité des bruits dans une image JPEG.

Dans ce cas, le RV pour une observation $\tilde{z}_{k,i}$ lorsque les paramètres $(\tilde{\mu}_k, \tilde{a}_j, \tilde{b}_j, \gamma_j)$ sont connus est défini par

$$\Lambda_{\text{gen}}(\tilde{z}_{k,i}) = \frac{1}{2}\Big(\log\left(\tilde{\sigma}^2_{k,0}\right) - \log\left(\tilde{\sigma}^2_{k,1}\right)\Big) + \frac{1}{2}\Big(\frac{1}{\tilde{\sigma}^2_{k,0}} - \frac{1}{\tilde{\sigma}^2_{k,1}}\Big)(\tilde{z}_{k,i} - \tilde{\mu}_k)^2. \tag{C.32}$$

De manière similaire, en vertu du TCL de Lindeberg, le RV $\Lambda_{\text{gen}}(\widetilde{\mathbf{Z}}) = \sum_{k=1}^{K}\sum_{i=1}^{\tilde{n}_k}\Lambda_{\text{gen}}(\tilde{z}_{k,i})$ suit la distribution Gaussienne sous l'hypothèse $\mathcal{H}_j$

$$\Lambda_{\text{gen}}(\widetilde{\mathbf{Z}}) \xrightarrow{d} \mathcal{N}\Big(m^{(j)}_{\text{gen}}, v^{(j)}_{\text{gen}}\Big) \tag{C.33}$$

où l'espérance $m^{(j)}_{\text{gen}}$ et la variance $v^{(j)}_{\text{gen}}$ sont données par

$$m^{(j)}_{\text{gen}} = \sum_{k=1}^{K}\sum_{i=1}^{\tilde{n}_k}\left[\frac{1}{2}\Big(\log\left(\tilde{\sigma}^2_{k,0}\right) - \log\left(\tilde{\sigma}^2_{k,1}\right)\Big) + \frac{1}{2}\Big(\frac{1}{\tilde{\sigma}^2_{k,0}} - \frac{1}{\tilde{\sigma}^2_{k,1}}\Big)\tilde{\sigma}^2_{k,j}\right] \tag{C.34}$$

$$v^{(j)}_{\text{gen}} = \sum_{k=1}^{K}\sum_{i=1}^{\tilde{n}_k}\frac{1}{2}\Big(\frac{1}{\tilde{\sigma}^2_{k,0}} - \frac{1}{\tilde{\sigma}^2_{k,1}}\Big)^2\tilde{\sigma}^4_{k,j}. \tag{C.35}$$

En fin, le test RV $\delta^{\star}_{\text{gen}}$ basé sur le RV normalisé $\Lambda^{\star}_{\text{gen}}(\widetilde{\mathbf{Z}}) = \frac{\Lambda_{\text{gen}}(\widetilde{\mathbf{Z}}) - m^{(0)}_{\text{gen}}}{\sqrt{v^{(0)}_{\text{gen}}}}$ est donné comme suit

$$\delta^{\star}_{\text{gen}}(\widetilde{\mathbf{Z}}) = \begin{cases}\mathcal{H}_0 & \text{if} \quad \Lambda^{\star}_{\text{gen}}(\widetilde{\mathbf{Z}}) < \tau^{\star}_{\text{gen}} \\ \mathcal{H}_1 & \text{if} \quad \Lambda^{\star}_{\text{gen}}(\widetilde{\mathbf{Z}}) \geq \tau^{\star}_{\text{gen}}.\end{cases} \tag{C.36}$$

où le seuil de décision $\tau^{\star}_{\text{gen}}$ est la solution de l'équation $\mathbb{P}_{\mathcal{H}_0}\Big[\Lambda^{\star}_{\text{gen}}(\widetilde{\mathbf{Z}}) \geq \tau^{\star}_{\text{gen}}\Big] = \alpha_0$.

Theorem C.4. *Dans le context idéal où tous les paramètres $(\tilde{\boldsymbol{\mu}}, \tilde{a}_j, \tilde{b}_j, \gamma)$ sont connus, le seuil de décision et la puissance du test RV $\delta^{\star}_{\text{gen}}$ sont donnés par*

$$\tau^{\star}_{\text{gen}} = \Phi^{-1}(1 - \alpha_0) \tag{C.37}$$

$$\beta(\delta^{\star}_{\text{gen}}) = 1 - \Phi\left(\frac{m^{(0)}_{\text{gen}} - m^{(1)}_{\text{gen}} + \tau^{\star}_{\text{gen}}\sqrt{v^{(0)}_{\text{gen}}}}{\sqrt{v^{(1)}_{\text{gen}}}}\right). \tag{C.38}$$

C.7 Conclusion

Il ressort de l'état de l'art dans le domaine de la criminalistique numérique des images que les approches issues de la théorie de la décision statistique demeurent très largement minoritaires. En outre, les méthodes actuelles exploitent seulement de façon partielle et empirique des modèles d'image naturelle, ce qui limite leur portée. Les travaux présentés dans ce manuscrit se positionnent dans l'intersection entre le domaine de la criminalistique numérique, le modélisation et l'estimation statistiques des images naturelles, et la théorie de la décision statistique.

Dans ce manuscrit, deux problèmes importants du domaine de la criminalistique numérique sont abordés : identification de l'origine d'une image et détection

d'informations cachées dans une image. L'objectif de cette thèse est de développer une procédure de décision statistique la plus fiable possible en maîtrisant simultanément des probabilités d'erreur de détection, étant des enjeux principaux dans le contexte opérationnel. Afin d'atteindre cet objectif, il s'agit de construire des tests d'hypothèses statistiques dans le cadre de la théorie de la décision statistique en reposant sur un modèle paramétrique qui caractérise des propriétés physiques et statistiques des images naturelles. La méthodologie commune tout au long de ce manuscrit consiste à étudier un test optimal donné par le test RV au sens de Neyman-Pearson dans le context idéal où tous les paramètres sont connus complètement. Pour des applications pratiques, lorsque des paramètres sont inconnus, ces derniers sont estimés et remplacés afin de construire des tests RVG dont les performances statistiques sont analytiquement établies.

Un modèle pertinent d'une image naturelle est une condition primordiale afin que les tests proposés peuvent atteindre une haute performance et garantir le respect sur la contrainte des probabilités d'erreur en pratique. Ce modèle est construit à partir d'une étude complète de la chaîne d'acquisition d'un appareil photographique, traversant du format RAW jusqu'au format JPEG. Les études statistiques des images naturelles sont effectuées à la fois dans le domaine spatial et le domaine DCT. Dans le domaine spatial, le modèle de bruit hétéroscédastique et le modèle de bruit généralisé sont proposés afin de caractériser des propriétés statistiques des images RAW et des images TIFF/JPEG, respectivement. D'un autre côté, il est proposé de modéliser la distribution statistique des coefficients DCT en fournissant une justification mathématique complète basée sur le modèle doublement stochastique. Le modèle des coefficients DCT établi n'est pas seulement plus performant que des autres modèles actuelles dans la littérature, mais est également exploitable dans de nombreux domaines.

En vue de l'identification de l'origine d'une image, une approche fiable consiste à se baser sur des empreintes physiques engendrées par le processus d'acquisition de l'image. De nombreuses empreintes sont proposées dans ce manuscrit pour l'identification des modèles d'appareil photographique à partir d'une image naturelle. Ces empreintes sont des paramètres informatifs du modèle de bruit hétéroscédastique, du modèle de bruit généralisé, et du modèle des coefficients DCT. Ces paramètres sont invariants par rapport au contenu de l'image et discriminants pour des différents modèles d'appareil photographique. En outre, la variabilité des empreintes est prise en compte dans la construction des tests statistiques, permettant une perte de puissance négligeable comparé avec des puissances théoriques et le respect sur la contrainte de la probabilité de fausse alarme, comme montré dans de nombreuses expérimentations numériques. Il est noté que de nombreuses pistes possibles sont exploitées pour la problématique de l'identification des modèles d'appareil photographique.

Pour la problématique de la détection d'informations cachées, la technique de substitution de LSB et l'algorithme de Jsteg sont considérés. Plus particulièrement, le modèle de bruit hétéroscédastique est exploité afin de construire des tests RVG pour la stéganalyse de la substitution de LSB. Un autre apport de cette approche est

la prise en compte du phénomène de clipping dont l'impact sur la stéganalyse restait inconnu. D'ailleurs, le modèle des coefficients DCT est utilisé afin de construire des tests statistiques et l'estimateur MV qui sont plus performant que des autres méthodes actuelles pour la stéganalyse de l'algorithme de Jsteg.

Les travaux présentés dans ce manuscrit ouvrent la voie à l'application de cette approche statistique pour la détection de la falsification des images numériques. L'approche consiste toujours à se base les empreintes proposées en recherchant une incohérence d'une zone suspect avec d'autres zones de l'image. Tous types de falsification visant à modifier des propriétés statistiques ou des empreintes sont considérés détectables, sauf la manipulation consistant à remplacer une zone de l'image par une autre zone de la même image. Par ailleurs, la problématique de l'intégrité de l'image peut être approchée par l'exploitation du modèle des coefficients DCT proposé afin d'estimer l'historique de la compression JPEG et détecter la compression double. En outre, des approches proposées peuvent être également appliquées pour la criminalistique des vidéos, sachant que la chaîne de la formation des vidéos est plus compliquée que celle des images. De plus, puisque les méthodes de l'analyse criminalistique peuvent être accessibles par des falsificateurs, impliquant que ces derniers peuvent développer des mesures afin de dissimuler les preuves de l'origine de l'image et de la falsification. Ces mesures sont appelées contre-criminalistiques. Ainsi, il est recommandé de prendre en compte les actions des analysts criminalistiques et des falsificateurs dans un même cadre afin de développer une stratégie optimale pour l'analyse criminalistique.

Bibliography

[1] T. Gloe and R. Böhme, "The dresden image database for benchmarking digital image forensics," *Journal of Digital Forensic Practice*, vol. 3, no. 2-4, pp. 150–159, 2010. (Cited on pages xii, 94, 102, 104, 117 and 129.)

[2] R. Ramanath, W. E. Snyder, Y. Yoo, and M. S. Drew, "Color image processing pipeline," *Signal Processing Magazine, IEEE*, vol. 22, no. 1, pp. 34–43, 2005. (Cited on pages 11, 12, 88, 112 and 115.)

[3] S. Voloshynovskiy, S. Pereira, T. Pun, J. J. Eggers, and J. K. Su, "Attacks on digital watermarks: classification, estimation based attacks, and benchmarks," *Communications Magazine, IEEE*, vol. 39, no. 8, pp. 118–126, 2001. (Cited on page 21.)

[4] V. Conotter, G. Boato, H. Farid, and E. Roosevelt, "Active and passive multimedia forensics," 2011. (Cited on page 22.)

[5] M. C. Stamm, M. Wu, and K. Liu, "Information forensics: An overview of the first decade," *IEEE Access*, vol. 1, no. 1, pp. 167–200, 2013. (Cited on pages 22 and 34.)

[6] I. Cox, M. Miller, J. Bloom, J. Fridrich, and T. Kalker, *Digital watermarking and steganography*. Morgan Kaufmann, 2007. (Cited on pages 22 and 39.)

[7] X. Zhu, J. Ding, H. Dong, K. Hu, and X. Zhang, "Normalized correlation-based quantization modulation for robust watermarking," *Multimedia, IEEE Transactions on*, vol. 16, no. 7, pp. 1888–1904, 2014. (Cited on page 22.)

[8] X. Qi and X. Xin, "A singular-value-based semi-fragile watermarking scheme for image content authentication with tamper localization," *Journal of Visual Communication and Image Representation*, 2015. (Cited on page 22.)

[9] H. M. Al-Otum, "Semi-fragile watermarking for grayscale image authentication and tamper detection based on an adjusted expanded-bit multiscale quantization-based technique," *Journal of Visual Communication and Image Representation*, vol. 25, no. 5, pp. 1064–1081, 2014. (Cited on page 22.)

[10] T. H. Thai, "Statistical modeling and detection for digital image forensics," Ph.D. dissertation, Troyes, 2014. (Cited on page 24.)

[11] T. Mitchell, *Machine Learning*. McGraw Hill, 1997. (Cited on page 25.)

[12] V. N. Vapnik, *The Nature of Statistical Learning Theory*, 2nd ed. Springer Verlag, 2000. (Cited on page 25.)

[13] B. Schölkopf and A. Smola, *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond (Adaptive Computation and Machine Learning)*. MIT Press, 2001. (Cited on page 25.)

[14] S. Bayram, H. Sencar, N. Memon, and I. Avcibas, "Source camera identification based on cfa interpolation," in *Image Processing, 2005. ICIP 2005. IEEE International Conference on*, vol. 3. IEEE, 2005, pp. III–69. (Cited on pages 25, 30, 49, 88 and 112.)

[15] M. Kharrazi, H. T. Sencar, and N. Memon, "Blind source camera identification," in *Image Processing, 2004. ICIP'04. 2004 International Conference on*, vol. 1. IEEE, 2004, pp. 709–712. (Cited on pages 25, 28, 30, 88 and 112.)

[16] K. San Choi, E. Y. Lam, and K. K. Wong, "Source camera identification using footprints from lens aberration," in *Electronic Imaging 2006*. International Society for Optics and Photonics, 2006, pp. 60 690J–60 690J. (Cited on pages 25, 27, 28, 88 and 112.)

[17] K. S. Choi, E. Y. Lam, and K. K. Wong, "Automatic source camera identification using the intrinsic lens radial distortion," *Optics express*, vol. 14, no. 24, pp. 11 551–11 565, 2006. (Cited on page 25.)

[18] S. Bayram, H. T. Sencar, N. Memon, and I. Avcibas, "Improvements on source camera-model identification based on cfa interpolation," *Proc. of WG*, vol. 11, 2006. (Cited on page 25.)

[19] F. d. O. Costa, E. Silva, M. Eckmann, W. J. Scheirer, and A. Rocha, "Open set source camera attribution and device linking," *Pattern Recognition Letters*, vol. 39, pp. 92–101, 2014. (Cited on page 25.)

[20] R. J. Lehman, E., "Testing statistical hypotheses," in *Second Edition*. Springer, 2005. (Cited on pages 25, 53, 54, 121 and 144.)

[21] T. H. Thai, F. Retraint, and R. Cogranne, "Camera model identification based on the generalized noise model in natural images," *Digital Signal Processing*, vol. 48, pp. 285–297, 2016. (Cited on pages 25, 113, 116, 117, 119, 132 and 133.)

[22] T. H. Thai, R. Cogranne, and F. Retraint, "Camera model identification based on the heteroscedastic noise model," *Image Processing, IEEE Transactions on*, vol. 23, no. 1, pp. 250–263, 2014. (Cited on pages 25, 90, 91, 93, 94, 98, 99, 101, 105, 106, 107, 108, 109, 115, 116, 119 and 131.)

[23] T. H. Thai, R. Cogranne, and F. Retraint, "Optimal detector for camera model identification based on an accurate model of dct coefficients," in *Multimedia Signal Processing (MMSP), 2014 IEEE 16th International Workshop on*. IEEE, 2014, pp. 1–6. (Cited on pages 25 and 45.)

[24] T. H. Thai, R. Cogranne, and F. Retraint, "Camera model identification based on hypothesis testing theory," in *Signal Processing Conference (EUSIPCO), 2012 Proceedings of the 20th European*. IEEE, 2012, pp. 1747–1751. (Cited on page 25.)

[25] T. Qiao, F. Retraint, R. Cogranne, and T. H. Thai, "Source camera device identification based on raw images," in *Proc. of IEEE International conf. on Image Processing (ICIP)*. IEEE, 2015, pp. 3812–3816. (Cited on pages 25, 115, 116 and 131.)

[26] C.-C. Chang and C.-J. Lin, "Libsvm: A library for support vector machines," *ACM Transactions on Intelligent Systems and Technology (TIST)*, vol. 2, no. 3, p. 27, 2011. (Cited on page 25.)

[27] C. Scott, "Performance measures for neyman–pearson classification," *Information Theory, IEEE Transactions on*, vol. 53, no. 8, pp. 2852–2863, 2007. (Cited on pages 25, 89 and 151.)

[28] F. Peng, J. Shi, and M. Long, "Identifying photographic images and photorealistic computer graphics using multifractal spectrum features of prnu," in *Multimedia and Expo (ICME), 2014 IEEE International Conference on*. IEEE, 2014, pp. 1–6. (Cited on pages 26 and 30.)

[29] F. Peng and D.-l. Zhou, "Discriminating natural images and computer generated graphics based on the impact of cfa interpolation on the correlation of prnu," *Digital Investigation*, vol. 11, no. 2, pp. 111–119, 2014. (Cited on pages 26 and 30.)

[30] S. Lyu and H. Farid, "How realistic is photorealistic?" *Signal Processing, IEEE Transactions on*, vol. 53, no. 2, pp. 845–850, 2005. (Cited on pages 26 and 49.)

[31] T. Ng, S. Chang, J. Hsu, L. Xie, and M. Tsui, "Physics-motivated features for distinguishing photographic images and computer graphics," in *Proceedings of the 13th annual ACM international conference on Multimedia*. ACM, 2005, pp. 239–248. (Cited on page 26.)

[32] Y. Wang and P. Moulin, "On discrimination between photorealistic and photographic images," in *Acoustics, Speech and Signal Processing, 2006. ICASSP 2006 Proceedings. 2006 IEEE International Conference on*, vol. 2. IEEE, 2006, pp. II–II. (Cited on page 26.)

[33] W. Chen, Y. Q. Shi, and G. Xuan, "Identifying computer graphics using hsv color model and statistical moments of characteristic functions," in *Multimedia and Expo, 2007 IEEE International Conference on*. IEEE, 2007, pp. 1123–1126. (Cited on page 27.)

[34] K. Guo and R. Wang, "An effective method for identifying natural images and computer graphics," *J Comput Inform Syst*, vol. 6, no. 10, pp. 3303–8, 2010. (Cited on page 27.)

[35] Z. Li, J. Ye, and Y. Q. Shi, "Distinguishing computer graphics from photographic images using local binary patterns," in *Digital Forensics and Watermaking*. Springer, 2013, pp. 228–241. (Cited on page 27.)

[36] R. Zhang and R. Wang, "Distinguishing photorealistic computer graphics from natural images by imaging features and visual features," in *Electronics, Communications and Control (ICECC), 2011 International Conference on*. IEEE, 2011, pp. 226–229. (Cited on page 27.)

[37] R. Zhang, R.-D. Wang, and T.-T. Ng, "Distinguishing photographic images and photorealistic computer graphics using visual vocabulary on local image edges," in *Digital forensics and watermarking*. Springer, 2012, pp. 292–305. (Cited on page 27.)

[38] F. Devernay and O. D. Faugeras, "Automatic calibration and removal of distortion from scenes of structured environments," in *SPIE's 1995 International Symposium on Optical Science, Engineering, and Instrumentation*. International Society for Optics and Photonics, 1995, pp. 62–72. (Cited on page 28.)

[39] P. D. Kovesi, "Matlab and octave functions for computer vision and image processing," *Online: http://www. csse. uwa. edu. au/˜ pk/Research/MatlabFns/# match*, 2000. (Cited on page 28.)

[40] J. R. Janesick, *Scientific charge-coupled devices*. SPIE press Bellingham, 2001, vol. 117. (Cited on pages 28, 89 and 113.)

[41] J. Nakamura, *Image sensors and signal processing for digital still cameras*. Boca Raton: CRC Press, 2005. (Cited on pages 28, 64, 88, 89, 112, 113 and 115.)

[42] J. Lukas, J. Fridrich, and M. Goljan, "Digital camera identification from sensor pattern noise," *Information Forensics and Security, IEEE Transactions on*, vol. 1, no. 2, pp. 205–214, 2006. (Cited on pages 28, 29, 30, 36, 73, 81, 89, 109 and 113.)

[43] K. Kurosawa, K. Kuroki, and N. Saitoh, "Ccd fingerprint method-identification of a video camera from videotaped images," in *Image Processing, 1999. ICIP 99. Proceedings. 1999 International Conference on*, vol. 3. IEEE, 1999, pp. 537–540. (Cited on pages 28, 89 and 113.)

[44] M. K. Mihcak, I. Kozintsev, and K. Ramchandran, "Spatially adaptive statistical modeling of wavelet image coefficients and its application to denoising," in *Acoustics, Speech, and Signal Processing, 1999. Proceedings., 1999 IEEE*

International Conference on, vol. 6. IEEE, 1999, pp. 3253–3256. (Cited on page 28.)

[45] C.-T. Li, "Source camera identification using enhanced sensor pattern noise," *Information Forensics and Security, IEEE Transactions on*, vol. 5, no. 2, pp. 280–287, 2010. (Cited on pages 29, 89 and 113.)

[46] R. Li, C.-T. Li, and Y. Guan, "A reference estimator based on composite sensor pattern noise for source device identification," in *IS&T/SPIE Electronic Imaging*. International Society for Optics and Photonics, 2014, pp. 90 280O–90 280O. (Cited on page 29.)

[47] M. Chen, J. Fridrich, M. Goljan, and J. Lukás, "Determining image origin and integrity using sensor noise," *Information Forensics and Security, IEEE Transactions on*, vol. 3, no. 1, pp. 74–90, 2008. (Cited on pages 29, 30, 33, 36, 89, 109, 113 and 155.)

[48] M. Goljan, J. Fridrich, and T. Filler, "Large scale test of sensor fingerprint camera identification," in *IS&T/SPIE Electronic Imaging*. International Society for Optics and Photonics, 2009, pp. 72 540I–72 540I. (Cited on pages 29, 30, 36, 89, 90, 105, 106, 107, 108, 109, 113, 114 and 133.)

[49] X. Kang, Y. Li, Z. Qu, and J. Huang, "Enhancing source camera identification performance with a camera reference phase sensor pattern noise," *Information Forensics and Security, IEEE Transactions on*, vol. 7, no. 2, pp. 393–402, 2012. (Cited on pages 30, 89 and 113.)

[50] S. Dehnie, T. Sencar, and N. Memon, "Digital image forensics for identifying computer generated and digital camera images," in *Image Processing, 2006 IEEE International Conference on*. IEEE, 2006, pp. 2313–2316. (Cited on page 30.)

[51] A. Swaminathan, M. Wu, and K. R. Liu, "Nonintrusive component forensics of visual sensors using output images," *Information Forensics and Security, IEEE Transactions on*, vol. 2, no. 1, pp. 91–106, 2007. (Cited on pages 30, 88 and 112.)

[52] H. Cao and A. C. Kot, "Accurate detection of demosaicing regularity for digital image forensics," *Information Forensics and Security, IEEE Transactions on*, vol. 4, no. 4, pp. 899–910, 2009. (Cited on pages 31, 49, 88 and 112.)

[53] A. E. Dirik, S. Bayram, H. T. Sencar, and N. Memon, "New features to identify computer generated images," in *Image Processing, 2007. ICIP 2007. IEEE International Conference on*, vol. 4. IEEE, 2007, pp. IV–433. (Cited on page 31.)

[54] A. Gallagher and T. Chen, "Image authentication by detecting traces of demosaicing," in *Computer Vision and Pattern Recognition Workshops, 2008.*

CVPRW'08. IEEE Computer Society Conference on. IEEE, 2008, pp. 1–8. (Cited on pages 31, 50, 51, 57, 58, 59, 168, 169 and 170.)

[55] T. Qiao, R. Cogranne, and F. Retraint, "Image authentication by statistical analysis," in *Signal Processing Conference (EUSIPCO), 2013 Proceedings of the 21th European.* IEEE, EURASIP, 2013, pp. 1–5. (Cited on pages 31 and 137.)

[56] M. Mancuso and S. Battiato, "An introduction to the digital still camera technology," *ST Journal of System Research*, vol. 2, no. 2, 2001. (Cited on page 31.)

[57] K. S. Choi, E. Y. Lam, and K. K. Wong, "Source camera identification by jpeg compression statistics for image forensics," in *TENCON 2006. 2006 IEEE Region 10 Conference.* IEEE, 2006, pp. 1–4. (Cited on pages 31, 88 and 113.)

[58] G. Xu, S. Gao, Y. Q. Shi, R. Hu, and W. Su, "Camera-model identification using markovian transition probability matrix," in *Digital Watermarking.* Springer, 2009, pp. 294–307. (Cited on page 32.)

[59] H. Farid and S. Lyu, "Higher-order wavelet statistics and their application to digital forensics," in *null.* IEEE, 2003, p. 94. (Cited on page 33.)

[60] S. Bayram, I. Avcibas, B. Sankur, and N. Memon, "Image manipulation detection with binary similarity measures," in *Signal Processing Conference, 2005 13th European.* IEEE, 2005, pp. 1–4. (Cited on page 33.)

[61] A. Popescu and H. Farid, "Exposing digital forgeries by detecting duplicated image region [technical report]. 2004-515," *Hanover, Department of Computer Science, Dartmouth College. USA*, 2004. (Cited on page 33.)

[62] M. Bashar, K. Noda, N. Ohnishi, and K. Mori, "Exploring duplicated regions in natural images." *IEEE transactions on image processing: a publication of the IEEE Signal Processing Society*, no. 99, 2010. (Cited on page 33.)

[63] I. Avcibaş, M. Kharrazi, N. Memon, and B. Sankur, "Image steganalysis with binary similarity measures," *EURASIP Journal on Applied Signal Processing*, vol. 2005, pp. 2749–2757, 2005. (Cited on page 33.)

[64] S. Lyu and H. Farid, "Steganalysis using higher-order image statistics," *Information Forensics and Security, IEEE Transactions on*, vol. 1, no. 1, pp. 111–119, 2006. (Cited on pages 33, 45 and 62.)

[65] M. K. Johnson and H. Farid, "Exposing digital forgeries through chromatic aberration," in *Proceedings of the 8th workshop on Multimedia and security.* ACM, 2006, pp. 48–55. (Cited on page 33.)

[66] M. Goljan and J. Fridrich, "Estimation of lens distortion correction from single images," in *IS&T/SPIE Electronic Imaging*. International Society for Optics and Photonics, 2014, pp. 90 280N–90 280N. (Cited on page 33.)

[67] A. Swaminathan, M. Wu, and K. R. Liu, "Digital image forensics via intrinsic fingerprints," *Information Forensics and Security, IEEE Transactions on*, vol. 3, no. 1, pp. 101–117, 2008. (Cited on pages 33 and 49.)

[68] A. C. Popescu and H. Farid, "Exposing digital forgeries by detecting traces of resampling," *Signal Processing, IEEE Transactions on*, vol. 53, no. 2, pp. 758–767, 2005. (Cited on pages 33, 36, 135, 137, 144, 147 and 151.)

[69] A. E. Dirik and N. D. Memon, "Image tamper detection based on demosaicing artifacts." in *ICIP*, 2009, pp. 1497–1500. (Cited on page 33.)

[70] H. Farid, "Blind inverse gamma correction," *Image Processing, IEEE Transactions on*, vol. 10, no. 10, pp. 1428–1433, 2001. (Cited on page 33.)

[71] A. C. Popescu and H. Farid, "Statistical tools for digital forensics," in *Information Hiding*. Springer, 2005, pp. 128–147. (Cited on page 33.)

[72] A. P. Dempster, N. M. Laird, D. B. Rubin *et al.*, "Maximum likelihood from incomplete data via the em algorithm," *Journal of the Royal statistical Society*, vol. 39, no. 1, pp. 1–38, 1977. (Cited on pages 33 and 135.)

[73] M. Kirchner, "Fast and reliable resampling detection by spectral analysis of fixed linear predictor residue," in *Proceedings of the 10th ACM workshop on Multimedia and security*. ACM, 2008, pp. 11–20. (Cited on pages 33, 135, 144 and 147.)

[74] M. Kirchner, "Linear row and column predictors for the analysis of resized images," in *Proceedings of the 12th ACM workshop on Multimedia and security*. ACM, 2010, pp. 13–18. (Cited on pages 33 and 135.)

[75] A. C. Gallagher, "Detection of linear and cubic interpolation in jpeg compressed images," in *Computer and Robot Vision, 2005. Proceedings. The 2nd Canadian Conference on*. IEEE, 2005, pp. 65–72. (Cited on pages 34, 135 and 140.)

[76] B. Mahdian and S. Saic, "Blind authentication using periodic properties of interpolation," *Information Forensics and Security, IEEE Transactions on*, vol. 3, no. 3, pp. 529–538, 2008. (Cited on page 34.)

[77] W. Wei, S. Wang, X. Zhang, and Z. Tang, "Estimation of image rotation angle using interpolation-related spectral signatures with application to blind detection of image forgery," *Information Forensics and Security, IEEE Transactions on*, vol. 5, no. 3, pp. 507–517, 2010. (Cited on page 34.)

[78] X. Feng, I. J. Cox, and G. Doërr, "An energy-based method for the forensic detection of re-sampled images," in *Multimedia and Expo (ICME), 2011 IEEE International Conference on*. IEEE, 2011, pp. 1–6. (Cited on page 34.)

[79] X. Feng, I. J. Cox, and G. Doerr, "Normalized energy density-based forensic detection of resampled images," *Multimedia, IEEE Transactions on*, vol. 14, no. 3, pp. 536–545, 2012. (Cited on page 34.)

[80] S.-J. Ryu and H.-K. Lee, "Estimation of linear transformation by analyzing the periodicity of interpolation," *Pattern Recognition Letters*, vol. 36, pp. 89–99, 2014. (Cited on page 34.)

[81] M. Kirchner and J. Fridrich, "On detection of median filtering in digital images," in *IS&T/SPIE Electronic Imaging*. International Society for Optics and Photonics, 2010, pp. 754 110–754 110. (Cited on page 34.)

[82] G. Cao, Y. Zhao, R. Ni, L. Yu, and H. Tian, "Forensic detection of median filtering in digital images," in *Multimedia and Expo (ICME), 2010 IEEE International Conference on*. IEEE, 2010, pp. 89–94. (Cited on page 34.)

[83] H.-D. Yuan, "Blind forensics of median filtering in digital images," *Information Forensics and Security, IEEE Transactions on*, vol. 6, no. 4, pp. 1335–1345, 2011. (Cited on page 34.)

[84] M. C. Stamm and K. R. Liu, "Forensic detection of image manipulation using statistical intrinsic fingerprints," *Information Forensics and Security, IEEE Transactions on*, vol. 5, no. 3, pp. 492–506, 2010. (Cited on pages 34 and 36.)

[85] M. Stamm and K. R. Liu, "Blind forensics of contrast enhancement in digital images," in *Image Processing, 2008. ICIP 2008. 15th IEEE International Conference on*. IEEE, 2008, pp. 3112–3115. (Cited on pages 34 and 36.)

[86] Z. Fan and R. L. De Queiroz, "Identification of bitmap compression history: Jpeg detection and quantizer estimation," *Image Processing, IEEE Transactions on*, vol. 12, no. 2, pp. 230–235, 2003. (Cited on pages 35 and 36.)

[87] R. Neelamani, R. De Queiroz, Z. Fan, S. Dash, and R. G. Baraniuk, "Jpeg compression history estimation for color images," *Image Processing, IEEE Transactions on*, vol. 15, no. 6, pp. 1365–1378, 2006. (Cited on pages 35 and 36.)

[88] W. S. Lin, S. K. Tjoa, H. V. Zhao, and K. Liu, "Digital image source coder forensics via intrinsic fingerprints," *Information Forensics and Security, IEEE Transactions on*, vol. 4, no. 3, pp. 460–475, 2009. (Cited on pages 35 and 36.)

[89] H. Farid, "Exposing digital forgeries from jpeg ghosts," *Information Forensics and Security, IEEE Transactions on*, vol. 4, no. 1, pp. 154–160, 2009. (Cited on page 35.)

[90] J. Lukáš and J. Fridrich, "Estimation of primary quantization matrix in double compressed jpeg images," in *Proc. Digital Forensic Research Workshop*, 2003, pp. 5–8. (Cited on page 35.)

[91] T. Pevnỳ and J. Fridrich, "Detection of double-compression in jpeg images for applications in steganography," *Information Forensics and Security, IEEE Transactions on*, vol. 3, no. 2, pp. 247–258, 2008. (Cited on page 35.)

[92] F. Huang, J. Huang, and Y. Q. Shi, "Detecting double jpeg compression with the same quantization matrix," *Information Forensics and Security, IEEE Transactions on*, vol. 5, no. 4, pp. 848–856, 2010. (Cited on page 35.)

[93] Z. Lin, J. He, X. Tang, and C.-K. Tang, "Fast, automatic and fine-grained tampered jpeg image detection via dct coefficient analysis," *Pattern Recognition*, vol. 42, no. 11, pp. 2492–2501, 2009. (Cited on page 35.)

[94] T. Bianchi and A. Piva, "Image forgery localization via block-grained analysis of jpeg artifacts," *Information Forensics and Security, IEEE Transactions on*, vol. 7, no. 3, pp. 1003–1017, 2012. (Cited on page 35.)

[95] W. Li, Y. Yuan, and N. Yu, "Passive detection of doctored jpeg image via block artifact grid extraction," *Signal Processing*, vol. 89, no. 9, pp. 1821–1829, 2009. (Cited on page 35.)

[96] M. Barni, A. Costanzo, and L. Sabatini, "Identification of cut & paste tampering by means of double-jpeg detection and image segmentation," in *Circuits and Systems (ISCAS), Proceedings of 2010 IEEE International Symposium on*. IEEE, 2010, pp. 1687–1690. (Cited on page 35.)

[97] H. Farid, "Image forgery detection," *Signal Processing Magazine, IEEE*, vol. 26, no. 2, pp. 16–25, 2009. (Cited on pages 35 and 112.)

[98] M. K. Johnson and H. Farid, "Exposing digital forgeries by detecting inconsistencies in lighting," in *Proceedings of the 7th workshop on Multimedia and security*. ACM, 2005, pp. 1–10. (Cited on page 35.)

[99] M. K. Johnson and H. Farid, "Exposing digital forgeries through specular highlights on the eye," in *Information Hiding*. Springer, 2007, pp. 311–325. (Cited on page 35.)

[100] M. K. Johnson and H. Farid, "Exposing digital forgeries in complex lighting environments." (Cited on page 35.)

[101] R. Böhme and M. Kirchner, "Counter-forensics: Attacking image forensics," in *Digital Image Forensics*. Springer, 2013, pp. 327–366. (Cited on page 36.)

[102] M. Kirchner and R. Bohme, "Hiding traces of resampling in digital images," *Information Forensics and Security, IEEE Transactions on*, vol. 3, no. 4, pp. 582–592, 2008. (Cited on page 36.)

[103] T. Gloe, M. Kirchner, A. Winkler, and R. Böhme, "Can we trust digital image forensics?" in *Proceedings of the 15th international conference on Multimedia*. ACM, 2007, pp. 78–86. (Cited on pages 36, 89 and 113.)

[104] G. Cao, Y. Zhao, R. Ni, and H. Tian, "Anti-forensics of contrast enhancement in digital images," in *Proceedings of the 12th ACM Workshop on Multimedia and Security*. ACM, 2010, pp. 25–34. (Cited on page 36.)

[105] M. C. Stamm and K. R. Liu, "Anti-forensics of digital image compression," *Information Forensics and Security, IEEE Transactions on*, vol. 6, no. 3, pp. 1050–1065, 2011. (Cited on page 36.)

[106] M. Kirchner and R. Böhme, "Synthesis of color filter array pattern in digital images," in *IS&T/SPIE Electronic Imaging*. International Society for Optics and Photonics, 2009, pp. 72 540K–72 540K. (Cited on page 36.)

[107] G. J. Simmons, "Prisoners' problem and the subliminal channel," in *Proceedings of CRYPTO*, 1984, pp. 51–67. (Cited on page 37.)

[108] R. Böhme, *Advanced statistical steganalysis*. Springer, 2010. (Cited on page 39.)

[109] J. Fridrich, *Steganography in digital media: principles, algorithms, and applications*. Cambridge University Press, 2009. (Cited on page 39.)

[110] J. Fridrich and J. Kodovskỳ, "Steganalysis of lsb replacement using parity-aware features," in *Information Hiding*. Springer, 2013, pp. 31–45. (Cited on page 39.)

[111] R. Cogranne, F. Retraint, C. Zitzmann, I. Nikiforov, L. Fillatre, and P. Cornu, "Hidden information detection using decision theory and quantized samples: Methodology, difficulties and results," *Digital Signal Processing*, vol. 24, pp. 144–161, 2014. (Cited on pages 39 and 159.)

[112] J. Fridrich, M. Goljan, D. Hogea, and D. Soukal, "Quantitative steganalysis of digital images: estimating the secret message length," *Multimedia systems*, vol. 9, no. 3, pp. 288–302, 2003. (Cited on page 39.)

[113] A. D. Ker, "Quantitative evaluation of pairs and rs steganalysis," in *Electronic Imaging 2004*. International Society for Optics and Photonics, 2004, pp. 83–97. (Cited on page 39.)

[114] R. Böhme and A. D. Ker, "A two-factor error model for quantitative steganalysis," in *Electronic Imaging 2006*. International Society for Optics and Photonics, 2006, pp. 607 206–607 206. (Cited on page 39.)

[115] T. Pevny, J. Fridrich, and A. D. Ker, "From blind to quantitative steganalysis," in *IS&T/SPIE Electronic Imaging*. International Society for Optics and Photonics, 2009, pp. 72 540C–72 540C. (Cited on pages 39 and 45.)

[116] J. Kodovsky and J. Fridrich, "Quantitative structural steganalysis of jsteg," *Information Forensics and Security, IEEE Transactions on*, vol. 5, no. 4, pp. 681–693, 2010. (Cited on pages 39, 62 and 170.)

[117] N. Provos and P. Honeyman, "Detecting steganographic content on the internet," Center for Information Technology Integration, Tech. Rep., 2001. (Cited on page 39.)

[118] J. Fridrich, M. Goljan, D. Soukal, and T. Holotyak, "Forensic steganalysis: determining the stego key in spatial domain steganography," in *Electronic Imaging 2005*. International Society for Optics and Photonics, 2005, pp. 631–642. (Cited on page 39.)

[119] A. D. Ker, "A fusion of maximum likelihood and structural steganalysis," in *Information Hiding*. Springer, 2007, pp. 204–219. (Cited on page 41.)

[120] J. Fridrich and M. Goljan, "On estimation of secret message length in lsb steganography in spatial domain," in *Electronic Imaging 2004*. International Society for Optics and Photonics, 2004, pp. 23–34. (Cited on pages 41, 42, 43, 62 and 65.)

[121] O. Dabeer, K. Sullivan, U. Madhow, S. Chandrasekaran, and B. Manjunath, "Detection of hiding in the least significant bit," *Signal Processing, IEEE Transactions on*, vol. 52, no. 10, pp. 3046–3058, 2004. (Cited on pages 41, 62 and 65.)

[122] A. D. Ker and R. Böhme, "Revisiting weighted stego-image steganalysis," in *Electronic Imaging 2008*. International Society for Optics and Photonics, 2008, pp. 681 905–681 905. (Cited on pages 42, 43 and 62.)

[123] R. Böhme, "Weighted stego-image steganalysis for jpeg covers," in *Information Hiding*. Springer, 2008, pp. 178–194. (Cited on pages 43, 62, 63, 64, 82, 170 and 172.)

[124] A. D. Ker, P. Bas, R. Böhme, R. Cogranne, S. Craver, T. Filler, J. Fridrich, and T. Pevnỳ, "Moving steganography and steganalysis from the laboratory into the real world," in *Proceedings of the first ACM workshop on Information hiding and multimedia security*. ACM, 2013, pp. 45–58. (Cited on pages 43, 45, 46, 61, 62 and 66.)

[125] O. Dabeer, K. Sullivan, U. Madhow, S. Chandrasekaran, and B. Manjunath, "Detection of hiding in the least significant bit," *Signal Processing, IEEE Transactions on*, vol. 52, no. 10, pp. 3046–3058, 2004. (Cited on page 44.)

[126] W. Hoeffding, "Asymptotically optimal tests for multinomial distributions," *The Annals of Mathematical Statistics*, pp. 369–401, 1965. (Cited on page 44.)

[127] E. L. Lehmann and J. P. Romano, *Testing statistical hypotheses*. Germany: Springer, 2006. (Cited on pages 44, 66, 68, 70, 95, 96, 122, 123, 125, 176 and 179.)

[128] V. Sedighi, R. Cogranne, and J. Fridrich, "Content-adaptive steganography by minimizing statistical detectability," vol. 11, no. 2, 2016, pp. 221 – 234. (Cited on pages 44 and 84.)

[129] R. Cogranne, C. Zitzmann, L. Fillatre, F. Retraint, I. Nikiforov, and P. Cornu, "A cover image model for reliable steganalysis," in *Information Hiding*. Springer, 2011, pp. 178–192. (Cited on pages 44, 61, 70 and 123.)

[130] R. Cogranne, C. Zitzmann, L. Fillatre, F. Retraint, I. Nikiforov, and P. Cornu, "Statistical decision by using quantized observations," in *Information Theory Proceedings (ISIT), 2011 IEEE International Symposium on*. IEEE, 2011, pp. 1210–1214. (Cited on pages 44, 45, 61, 70, 77 and 159.)

[131] R. Cogranne, C. Zitzmann, L. Fillatre, I. Nikiforov, F. Retraint, and P. Cornu, "Reliable detection of hidden information based on a non-linear local model," in *Statistical Signal Processing Workshop (SSP), 2011 IEEE*. IEEE, 2011, pp. 493–496. (Cited on page 44.)

[132] C. Zitzmann, R. Cogranne, F. Retraint, I. Nikiforov, L. Fillatre, and P. Cornu, "Statistical decision methods in hidden information detection," in *Information Hiding*. Springer, 2011, pp. 163–177. (Cited on pages 44, 45, 61, 65, 70 and 77.)

[133] R. Cogranne, F. Retraint, C. Zitzmann, I. Nikiforov, L. Fillatre, and P. Cornu, "Hidden information detection using decision theory and quantized samples: Methodology, difficulties and results," *Digital Signal Processing*, vol. 24, pp. 144 – 161, 2014. (Cited on pages 44 and 97.)

[134] R. Cogranne and F. Retraint, "An asymptotically uniformly most powerful test for lsb matching detection," pp. 464–476, 2013. (Cited on pages 45, 70, 97 and 123.)

[135] C. Zitzmann, R. Cogranne, L. Fillatre, I. Nikiforov, F. Retraint, and P. Cornu, "Hidden information detection based on quantized laplacian distribution," in *Acoustics, Speech and Signal Processing (ICASSP), 2012 IEEE International Conference on*. IEEE, 2012, pp. 1793–1796. (Cited on pages 45, 62, 63, 65, 66, 70, 71, 72, 76, 77, 80, 81, 82, 123, 170 and 172.)

[136] T. Thai, R. Cogranne, and F. Retraint, "Statistical model of quantized dct coefficients: Application in the steganalysis of jsteg algorithm." *IEEE transactions on image processing: a publication of the IEEE Signal Processing Society*, vol. 23, no. 5, p. 1980, 2014. (Cited on pages 45, 66, 70, 73, 116 and 156.)

[137] J. Fridrich and J. Kodovskỳ, "Rich models for steganalysis of digital images," *Information Forensics and Security, IEEE Transactions on*, vol. 7, no. 3, pp. 868–882, 2012. (Cited on pages 45 and 83.)

[138] T. Pevny and J. Fridrich, "Multiclass detector of current steganographic methods for jpeg format," *Information Forensics and Security, IEEE Transactions on*, vol. 3, no. 4, pp. 635–650, 2008. (Cited on pages 45, 62 and 170.)

[139] T. Pevnỳ, P. Bas, and J. Fridrich, "Steganalysis by subtractive pixel adjacency matrix," *information Forensics and Security, IEEE Transactions on*, vol. 5, no. 2, pp. 215–224, 2010. (Cited on page 45.)

[140] Y. Wang and P. Moulin, "Optimized feature extraction for learning-based image steganalysis," *Information Forensics and Security, IEEE Transactions on*, vol. 2, no. 1, pp. 31–45, 2007. (Cited on page 45.)

[141] J. Kodovskỳ, J. Fridrich, and V. Holub, "Ensemble classifiers for steganalysis of digital media," *Information Forensics and Security, IEEE Transactions on*, vol. 7, no. 2, pp. 432–444, 2012. (Cited on pages 45 and 82.)

[142] R. Cogranne and J. Fridrich, "Modeling and extending the ensemble classifier for steganalysis of digital images using hypothesis testing theory," *Information Forensics and Security, IEEE Transactions on*, vol. 10, no. 12, pp. 2627–2642, Dec 2015. (Cited on pages 45, 82, 84, 97, 123 and 156.)

[143] R. Cogranne, V. Sedighi, J. Fridrich, and T. Pevny, "Is ensemble classifier needed for steganalysis in high-dimensional feature spaces?" in *Information Forensics and Security (WIFS), 2015 IEEE International Workshop on*, Nov 2015, pp. 1–6. (Cited on pages 45, 83 and 84.)

[144] T. Ng, S. Chang, J. Hsu, and M. Pepeljugoski, "Columbia photographic images and photorealistic computer graphics dataset," *Columbia University, ADVENT Technical Report*, pp. 205–2004, 2005. (Cited on page 56.)

[145] T. Zhang and X. Ping, "A fast and effective steganalytic technique against jsteg-like algorithms," in *Proceedings of the 2003 ACM symposium on Applied computing*. ACM, 2003, pp. 307–311. (Cited on page 61.)

[146] J. Fridrich, M. Goljan, and R. Du, "Reliable detection of lsb steganography in color and grayscale images," in *Proceedings of the 2001 workshop on Multimedia and security: new challenges*. ACM, 2001, pp. 27–30. (Cited on page 61.)

[147] S. Dumitrescu, X. Wu, and Z. Wang, "Detection of lsb steganography via sample pair analysis," *Signal Processing, IEEE Transactions on*, vol. 51, no. 7, pp. 1995–2007, 2003. (Cited on page 61.)

[148] R. Cogranne, C. Zitzmann, F. Retraint, I. V. Nikiforov, P. Cornu, and L. Fillatre, "A local adaptive model of natural images for almost optimal detection of hidden data," *Signal Processing*, vol. 100, pp. 169–185, 2014. (Cited on page 62.)

[149] E. Y. Lam and J. W. Goodman, "A mathematical analysis of the dct coefficient distributions for images," *Image Processing, IEEE Transactions on*, vol. 9, no. 10, pp. 1661–1666, 2000. (Cited on pages 62 and 116.)

[150] F. Muller, "Distribution shape of two-dimensional dct coefficients of natural images," *Electronics Letters*, vol. 29, no. 22, pp. 1935–1936, 1993. (Cited on pages 62 and 156.)

[151] J.-H. Chang, J. W. Shin, N. S. Kim, and S. K. Mitra, "Image probability distribution based on generalized gamma function," *Signal Processing Letters, IEEE*, vol. 12, no. 4, pp. 325–328, 2005. (Cited on pages 62 and 156.)

[152] P. Sallee, "Model-based methods for steganography and steganalysis," *International Journal of Image and graphics*, vol. 5, no. 01, pp. 167–189, 2005. (Cited on page 62.)

[153] R. Böhme and A. Westfeld, "Breaking cauchy model-based jpeg steganography with first order statistics," pp. 125–140, 2004. (Cited on page 62.)

[154] K. Lee, A. Westfeld, and S. Lee, "Category attack for lsb steganalysis of jpeg images," in *Digital Watermarking*. New York: Springer, 2006, pp. 35–48. (Cited on page 62.)

[155] P. Bas, T. Filler, and T. Pevný, "Break our steganographic system — the ins and outs of organizing boss," in *Information Hiding, 13th International Workshop*, T. Filler, Ed., 2011. (Cited on pages 62, 75, 80, 147 and 170.)

[156] T. Denemark, V. Sedighi, V. Holub, R. Cogranne, and J. Fridrich, "Selection-channel-aware rich model for steganalysis of digital images." (Cited on pages 63, 74 and 172.)

[157] W. Tang, H. Li, W. Luo, and J. Huang, "Adaptive steganalysis against wow embedding algorithm," in *Proceedings of the 2nd ACM workshop on Information hiding and multimedia security*. ACM, 2014, pp. 91–96. (Cited on pages 63, 74 and 172.)

[158] W. B. Pennebaker and J. L. Mitchell, *JPEG: Still image data compression standard*. Germany: Springer, 1993. (Cited on page 64.)

[159] D. Upham, "Jsteg steganographic algorithm," *Available on the internet http://www.filewatcher.com/m/jpeg-jsteg-v4.diff.gz.8878-0.html*, 1999. (Cited on page 65.)

[160] T. H. Thai, R. Cogranne, and F. Retraint, "Steganalysis of jsteg algorithm based on a novel statistical model of quantized dct coefficients." in *ICIP*, 2013, pp. 4427–4431. (Cited on pages 66 and 73.)

[161] T. H. Thai, F. Retraint, and R. Cogranne, "Statistical model of natural images," in *Image Processing (ICIP), 2012 19th IEEE International Conference on*. IEEE, 2012, pp. 2525–2528. (Cited on pages 73 and 116.)

[162] M. Lebrun, "An analysis and implementation of the bm3d image denoising method," *Image Processing On Line*, 2012. (Cited on pages 73, 118 and 129.)

[163] M. Lebrun and A. Leclaire, "An implementation and detailed analysis of the k-svd image denoising algorithm," *Image Processing On Line*, 2012. (Cited on page 73.)

[164] A. Buades, B. Coll, and J.-M. Morel, "A non-local algorithm for image denoising," in *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, vol. 2. IEEE, 2005, pp. 60–65. (Cited on page 73.)

[165] R. Cogranne, T. Denemark, and J. Fridrich, "Theoretical model of the FLD ensemble classifier based on hypothesis testing theory," in *Information Forensics and Security (WIFS), 2014 IEEE 6th International Workshop on*, 2014, pp. 167–172. (Cited on pages 82 and 84.)

[166] J. Kodovskỳ and J. Fridrich, "Steganalysis of jpeg images using rich models," in *IS&T/SPIE Electronic Imaging*. International Society for Optics and Photonics, 2012, pp. 83 030A–83 030A. (Cited on page 83.)

[167] B. Li, M. Wang, J. Huang, and X. Li, "A new cost function for spatial image steganography," in *Image Processing (ICIP), 2014 IEEE International Conference on*. IEEE, 2014, pp. 4206–4210. (Cited on page 84.)

[168] V. Holub, J. Fridrich, and T. Denemark, "Universal distortion design for steganography in an arbitrary domain," *EURASIP Journal on Information Security, Special Issue on Revised Selected Papers of the 1st ACM IH and MMS Workshop*, vol. 2014:1, 2014. (Cited on page 84.)

[169] Z. Deng, A. Gijsenij, and J. Zhang, "Source camera identification using auto-white balance approximation," in *Computer Vision (ICCV), 2011 IEEE International Conference on*. IEEE, 2011, pp. 57–64. (Cited on pages 88, 89 and 112.)

[170] C.-T. Li and Y. Li, "Color-decoupled photo response non-uniformity for digital image forensics," *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 22, no. 2, pp. 260–271, 2012. (Cited on pages 89 and 113.)

[171] T. Filler, J. Fridrich, and M. Goljan, "Using sensor pattern noise for camera model identification," in *Image Processing, 2008. ICIP 2008. 15th IEEE International Conference on.* IEEE, 2008, pp. 1296–1299. (Cited on pages 89 and 113.)

[172] A. Foi, M. Trimeche, V. Katkovnik, and K. Egiazarian, "Practical poissonian-gaussian noise modeling and fitting for single-image raw-data," *Image Processing, IEEE Transactions on*, vol. 17, no. 10, pp. 1737–1754, 2008. (Cited on pages 90, 91, 99, 115, 119, 177 and 178.)

[173] A. Foi, "Clipped noisy images: Heteroskedastic modeling and practical denoising," *Signal Processing*, vol. 89, no. 12, pp. 2609–2629, 2009. (Cited on page 90.)

[174] G. E. Healey and R. Kondepudy, "Radiometric ccd camera calibration and noise estimation," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 16, no. 3, pp. 267–276, 1994. (Cited on pages 90 and 91.)

[175] R. Cogranne, C. Zitzmann, F. Retraint, I. V. Nikiforov, P. Cornu, and L. Fillatre, "A local adaptive model of natural images for almost optimal detection of hidden data," *Signal Processing*, vol. 100, pp. 169 – 185, 2014. (Cited on page 97.)

[176] T. Qiao, C. Zitzmann, F. Retraint, and R. Cogranne, "Statistical detection of jsteg steganography using hypothesis testing theory," in *Image Processing (ICIP), 2014 IEEE International Conference on.* IEEE, 2014, pp. 5517–5521. (Cited on page 97.)

[177] T. Qiao, C. Ziitmann, R. Cogranne, and F. Retraint, "Detection of jsteg algorithm using hypothesis testing theory and a statistical model with nuisance parameters," in *Proceedings of the 2nd ACM workshop on Information hiding and multimedia security.* ACM, 2014, pp. 3–13. (Cited on pages 97 and 123.)

[178] T. Qiao, F. Retraint, R. Cogranne, and C. Zitzmann, "Steganalysis of jsteg algorithm using hypothesis testing theory," *EURASIP Journal on Information Security*, vol. 2015, no. 1, pp. 1–16. (Cited on page 97.)

[179] H. T. Sencar and N. Memon, *Digital image forensics: There is more to a picture than meets the eye.* Springer, 2012. (Cited on page 112.)

[180] M. Kirchner and T. Gloe, "Forensic camera model identification," in *Handbook of Digital Forensics of Multimedia Data and Devices, Chapter 9*, A. Ho and S. Li, Eds. John Wiley & Sons Ltd, 2004. (Cited on page 112.)

[181] S. Bayram, H. T. Sencar, and N. Memon, "Efficient techniques for sensor fingerprint matching in large image and video databases," pp. 754 109–754 109–8, 2010. [Online]. Available: http://dx.doi.org/10.1117/12.845737 (Cited on page 112.)

[182] M. Goljan and J. Fridrich, "Estimation of lens distortion correction from single images," pp. 90 280N–90 280N–13, 2014. [Online]. Available: http://dx.doi.org/10.1117/12.2036804 (Cited on page 112.)

[183] T. H. Thai, F. Retraint, and R. Cogranne, "Generalized signal-dependent noise model and parameter estimation for natural images," *Signal Processing*, vol. 114, pp. 164–170, 2015. (Cited on pages 113, 114, 128, 130, 150, 151 and 161.)

[184] B. Widrow, I. Kollar, and M.-C. Liu, "Statistical theory of quantization," *IEEE Transactions on Instrumentation and Measurement*, vol. 45, no. 2, pp. 353–361, 1996. (Cited on page 116.)

[185] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "Image denoising by sparse 3-d transform-domain collaborative filtering," *Image Processing, IEEE Transactions on*, vol. 16, no. 8, pp. 2080–2095, 2007. (Cited on pages 118 and 129.)

[186] A. V. Oppenheim, R. W. Schafer, J. R. Buck *et al.*, *Discrete-time signal processing*. Prentice-hall Englewood Cliffs, 1989, vol. 2. (Cited on page 136.)

[187] G. Schaefer and M. Stich, "Ucid: an uncompressed color image database," in *Electronic Imaging 2004*. International Society for Optics and Photonics, 2003, pp. 472–480. (Cited on page 146.)

[188] T. Julliand, V. Nozick, and H. Talbot, "Automated image splicing detection from noise estimation in raw images," in *Imaging for Crime Prevention and Detection (ICDP-15), 6th International Conference on*. IET, 2015, pp. 1–6. (Cited on page 155.)

[189] B. Baygun and A. Hero, "Optimal simultaneous detection and estimation under a false alarm constraint," *Information Theory, IEEE Transactions on*, vol. 41, no. 3, pp. 688–703, May 1995. (Cited on page 156.)

Tong QIAO

Doctorat : Optimisation et Sûreté des Systèmes

Année 2016

Détection statistique pour la criminalistique des images numériques

Le XXIème siècle étant le siècle du passage au tout numérique, les médias digitaux jouent un rôle de plus en plus important. Les logiciels sophistiqués de retouche d'images se sont démocratisés et permettent de diffuser facilement des images falsifiées. Ceci pose un problème sociétal puisqu'il s'agit de savoir si ce que l'on voit a été manipulé. Cette thèse s'inscrit dans le cadre de la criminalistique des images. Trois problèmes sont abordés : l'identification de l'origine d'une image, la détection d'informations cachées dans une image et la détection d'un exemple falsification : le rééchantillonnage. Ces travaux s'inscrivent dans le cadre de la théorie de la décision statistique et proposent la construction de détecteurs permettant de respecter une contrainte sur la probabilité de fausse alarme. Afin d'atteindre une performance de détection élevée, il est proposé d'exploiter les propriétés des images naturelles en modélisant les principales étapes de la chaîne d'acquisition d'un appareil photographique. La méthodologie, tout au long de ce manuscrit, consiste à étudier le détecteur optimal donné par le test du rapport de vraisemblance dans le contexte idéal où tous les paramètres du modèle sont connus. Lorsque des paramètres du modèle sont inconnus, ces derniers sont estimés afin de construire le test du rapport de vraisemblance généralisé dont les performances statistiques sont analytiquement établies. De nombreuses expérimentations sur des images simulées et réelles permettent de souligner la pertinence de l'approche proposée.

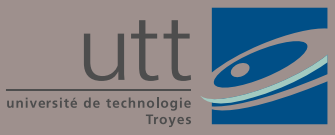
Mots clés : criminalistique - traitement d'images - modèles mathématiques - tests d'hypothèses (statistique) - estimation de paramètres.

Statistical Detection for Digital Image Forensics

The remarkable evolution of information technologies and digital imaging technology in the past decades allow digital images to be ubiquitous. The tampering of these images has become an unavoidable reality, especially in the field of cybercrime. The credibility and trustworthiness of digital images have been eroded, resulting in important consequences in terms of political, economic, and social issues. To restore the trust to digital images, the field of digital forensics was born. Three important problems are addressed in this thesis: image origin identification, detection of hidden information in a digital image and an example of tampering image detection : the resampling. The goal is to develop a statistical decision approach as reliable as possible that allows to guarantee a prescribed false alarm probability. To this end, the approach involves designing a statistical test within the framework of hypothesis testing theory based on a parametric model that characterizes physical and statistical properties of natural images. This model is developed by studying the image processing pipeline of a digital camera. As part of this work, the difficulty of the presence of unknown parameters is addressed using statistical estimation, making the application of statistical tests straightforward in practice. Numerical experiments on simulated and real images have highlighted the relevance of the proposed approach.

Keywords: forensic sciences - image processing - mathematical models - statistical hypothesis testing –parameter estimation.

Thèse réalisée en partenariat entre :



Ecole Doctorale "Sciences et Technologies"