



Statistical Modeling and Detection for Digital Image Forensics

Thanh Hai Thai

► To cite this version:

Thanh Hai Thai. Statistical Modeling and Detection for Digital Image Forensics. Statistics [math.ST]. Université de Technologie de Troyes, 2014. English. NNT: . tel-01072541

HAL Id: tel-01072541

<https://theses.hal.science/tel-01072541>

Submitted on 8 Oct 2014

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

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



Distributed under a Creative Commons Attribution - NonCommercial 4.0 International License

PHD THESIS

to obtain the degree of

DOCTOR of UNIVERSITY
of TECHNOLOGY of TROYES

Speciality: SYSTEMS OPTIMIZATION AND DEPENDABILITY

presented and defended by

Thanh Hai THAI

August 28th 2014

**Statistical Modeling and Detection for Digital
Image Forensics**

COMMITTEE:

Mrs. Jessica FRIDRICH	Professor	Reviewer
M. Patrick BAS	Chargé de recherche CNRS	Reviewer
M. William PUECH	Professeur des universités	Examinator
M. Igor NIKIFOROV	Professeur des universités	Examinator
M. Rémi COGRANNE	Maître de conférences	Examinator
M. Florent RETRAINT	Enseignant-Chercheur	Supervisor

To my Mom, my Dad, and my little brother,

To Thu Thao, my fiancée,

for their unlimited support, encouragement, and love.

Acknowledgments

This work has been carried out within the Laboratory of Systems Modeling and Dependability (LM2S) at the University of Technology of Troyes (UTT). It is funded in part by the strategic program COLUMBO.

This work has been accomplished under the supervision of M. Florent RETRAINT. I would like to express my deepest gratitude to him for his highly professional guidance and incessant support. He has accompanied with me from my master's internship and encouraged me to discover this field. I highly value the friendly yet professional environment created during my three-year doctoral. The high confidence he has given to me is possibly the greatest reward of my endeavors.

I am greatly indebted to my PhD co-advisor, M. Rémi COGRANNE who has assisted me with high efficiency and availability. It was my honor and pleasure to work with him. His personal and professional help during my doctoral is invaluable.

I would like to express my special thanks to Mrs. Jessica FRIDRICH and M. Patrick BAS for accepting to review my PhD thesis. I would also like to thank M. Igor NIKOFOROV and M. William PUECH for agreeing to examine this thesis. Valuable remarks provided by the respectful experts in this field like them would improve the thesis's quality.

I would like to thank the members of LM2S for the friendly and adorable environment which they have offered me since I joined the LM2S team.

Résumé

Le XXIème siècle étant le siècle du passage au tout numérique, les médias numériques jouent maintenant un rôle de plus en plus important dans la vie de tous les jours. De la même manière, les logiciels sophistiqués de retouche d'images se sont démocratisés et permettent aujourd'hui 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 numériques. Deux problèmes importants sont abordés : l'identification de l'origine d'une image et la détection d'informations cachées dans une image. 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 twenty-first century witnesses the digital revolution that allows digital media to become ubiquitous. They play a more and more important role in our everyday life. Similarly, sophisticated image editing software has been more accessible, resulting in the fact that falsified images are appearing with a growing frequency and sophistication. The credibility and trustworthiness of digital images have been eroded. To restore the trust to digital images, the field of digital image forensics was born. This thesis is part of the field of digital image forensics. Two important problems are addressed: image origin identification and hidden data detection. These problems are cast into the framework of hypothesis testing theory. The approach proposes to design a statistical test that allows us to guarantee a prescribed false alarm probability. In order to achieve a high detection performance, it is proposed to exploit statistical properties of natural images by modeling the main steps of image processing pipeline of a digital camera. The methodology throughout this manuscript consists of studying an optimal test given by the Likelihood Ratio Test in the ideal context where all model parameters are known in advance. When the model parameters are unknown, a method is proposed for parameter estimation in order to design a Generalized Likelihood Ratio Test whose statistical performances are analytically established. Numerical experiments on simulated and real images highlight the relevance of the proposed approach.

Table of Contents

1	General Introduction	1
1.1	General Context and Problem Description	1
1.2	Outline of the Thesis	3
1.3	Publications and Authors' Contribution	6
I	Overview on Digital Image Forensics and Statistical Image Modeling	9
2	Overview on Digital Image Forensics	11
2.1	Introduction	11
2.2	Image Processing Pipeline of a Digital Camera	12
2.2.1	RAW Image Formation	13
2.2.2	Post-Acquisition Processing	15
2.2.3	Image Compression	17
2.3	Passive Image Origin Identification	19
2.3.1	Lens Aberration	23
2.3.2	Sensor Imperfections	23
2.3.3	CFA Pattern and Interpolation	25
2.3.4	Image Compression	26
2.4	Passive Image Forgery Detection	27
2.5	Steganography and Steganalysis in Digital Images	29
2.5.1	LSB Replacement Paradigm and Jsteg Algorithm	32
2.5.2	Steganalysis of LSB Replacement in Spatial Domain	33
2.5.3	Steganalysis of Jsteg Algorithm	38
2.6	Conclusion	39
3	Overview on Statistical Modeling of Natural Images	41
3.1	Introduction	41
3.2	Spatial-Domain Image Model	41
3.2.1	Poisson-Gaussian and Heteroscedastic Noise Model	42
3.2.2	Non-Linear Signal-Dependent Noise Model	44
3.3	DCT Coefficient Model	45
3.3.1	First-Order Statistics of DCT Coefficients	45
3.3.2	Higher-Order Statistics of DCT Coefficients	46
3.4	Conclusion	46

II Statistical Modeling and Estimation for Natural Images from RAW Format to JPEG Format 47

4 Statistical Image Modeling and Estimation of Model Parameters	49
4.1 Introduction	50
4.2 Statistical Modeling of RAW Images	50
4.2.1 Heteroscedastic Noise Model	50
4.2.2 Estimation of Parameters (a, b) in the Heteroscedastic Noise Model	52
4.3 Statistical Modeling of TIFF Images	57
4.3.1 Generalized Noise Model	57
4.3.2 Estimation of Parameters $(\tilde{a}, \tilde{b})$ in the Generalized Noise Model	59
4.3.3 Application to Image Denoising	61
4.3.4 Numerical Experiments	62
4.4 Statistical Modeling in DCT Domain	65
4.4.1 Statistical Model of Quantized DCT Coefficients	65
4.4.2 Estimation of Parameters (α, β) from Unquantized DCT Coefficients	69
4.4.3 Estimation of Parameters (α, β) from Quantized DCT Coefficients	70
4.4.4 Numerical Experiments	71
4.5 Conclusion	73

III Camera Model Identification in Hypothesis Testing Framework 75

5 Camera Model Identification Based on the Heteroscedastic Noise Model	77
5.1 Introduction	77
5.2 Camera Fingerprint	78
5.3 Optimal Detector for Camera Model Identification Problem	80
5.3.1 Hypothesis Testing Formulation	80
5.3.2 LRT for Two Simple Hypotheses	80
5.4 GLRT with Unknown Image Parameters	83
5.5 GLRT with Unknown Image and Camera Parameters	86
5.6 Numerical Experiments	89
5.6.1 Detection Performance on Simulated Database	89
5.6.2 Detection Performance on Two <i>Nikon D70</i> and <i>Nikon D200</i> Camera Models	90
5.6.3 Detection Performance on a Large Image Database	91
5.7 Conclusion	94
5.8 Appendix	95

5.8.1	Expectation and Variance of the GLR $\bar{\Lambda}_{\text{het}}(z_{k,i}^{\text{wapp}})$ under Hypothesis $\mathcal{H}_j$	95
5.8.2	Expectation and Variance of the GLR $\tilde{\Lambda}_{\text{het}}(z_{k,i}^{\text{wapp}})$ under Hypothesis $\mathcal{H}_j$	96
6	Camera Model Identification Based on the Generalized Noise Model	99
6.1	Introduction	99
6.2	Camera Fingerprint	100
6.3	Optimal Detector for Camera Model Identification Problem	102
6.3.1	Hypothesis Testing Formulation	102
6.3.2	LRT for Two Simple Hypotheses	103
6.4	Practical Context: GLRT	105
6.4.1	GLRT with Unknown Image Parameters	105
6.4.2	GLRT with Unknown Image and Camera Parameters	106
6.5	Numerical Experiments	108
6.5.1	Detection Performance on Simulated Database	108
6.5.2	Detection Performance on Two <i>Nikon D70</i> and <i>Nikon D200</i> Camera Models	109
6.5.3	Detection Performance on a Large Image Database	110
6.6	Conclusion	111
6.7	Appendix	112
7	Camera Model Identification Based on DCT Coefficient Statistics	115
7.1	Introduction	116
7.2	Camera Fingerprint	116
7.2.1	Design of Camera Fingerprint	116
7.2.2	Extraction of Camera Fingerprint	117
7.2.3	Property of Camera Fingerprint	119
7.3	Optimal Detector for Camera Model Identification Problem	120
7.3.1	Hypothesis Testing Formulation	120
7.3.2	LRT for Two Simple Hypotheses	121
7.4	Practical Context: GLRT	123
7.4.1	GLRT with Unknown Parameters α_k	123
7.4.2	GLRT with Unknown Parameters $(\alpha_k, \tilde{c}_{k,1}, \tilde{d}_{k,1})$	124
7.5	Numerical Experiments	126
7.5.1	Detection Performance on Simulated Database	127
7.5.2	Detection Performance on Two <i>Canon Ixus 70</i> and <i>Nikon D200</i> Camera Models	128
7.5.3	Detection Performance on a Large Image Database	129
7.6	Conclusion	132
7.7	Appendix	132
7.7.1	Relation between the Parameters $(\tilde{a}, \tilde{b}, \gamma)$ and $(\alpha_{u,v}, \beta_{u,v})$	132
7.7.2	Laplace's Approximation of DCT Coefficient Model	133

7.7.3	Expectation and Variance of the LR $\Lambda_{\text{dct}}(I_{k,i})$ under Hypothesis $\mathcal{H}_j$	134
7.7.4	Asymptotic Expectation and Variance of the GLR $\tilde{\Lambda}_{\text{dct}}(I_{k,i})$ under Hypothesis $\mathcal{H}_j$	135
IV	Statistical Detection of Hidden Data in Natural Images	137
8	Statistical Detection of Data Embedded in Least Significant Bits of Clipped Images	139
8.1	Introduction	140
8.2	Cover-Image Model	141
8.2.1	Non-Clipped Image Model	141
8.2.2	Clipped Image Model	142
8.3	GLRT for Non-Clipped Images	142
8.3.1	Impact of LSB Replacement: Stego-Image Model	142
8.3.2	Hypothesis Testing Formulation	143
8.3.3	ML Estimation of Image Parameters	144
8.3.4	Design of GLRT	145
8.4	GLRT for Clipped Images	148
8.4.1	ML Estimation of Image Parameters	148
8.4.2	Design of GLRT	148
8.5	Numerical Experiments	151
8.5.1	Detection Performance on Simulated Database	151
8.5.2	Detection Performance on Real Image Database	153
8.6	Conclusion	155
8.7	Appendix	155
8.7.1	Denoising Filter for Non-Clipped RAW Images Corrupted by Signal-Dependent Noise	155
8.7.2	Statistical distribution of the GLR $\hat{\Lambda}_{\text{ncl}}(\mathbf{Z})$ under hypothesis $\mathcal{H}_0$	157
8.7.3	Statistical distribution of the GLR $\hat{\Lambda}_{\text{ncl}}(\mathbf{Z})$ under hypothesis $\mathcal{H}_1$	158
8.7.4	ML Estimation of Parameters in Truncated Gaussian Data	159
8.7.5	Statistical distribution of the GLR $\hat{\Lambda}_{\text{cl}}(\mathbf{Z})$	160
9	Steganalysis of Jsteg Algorithm Based on a Novel Statistical Model of Quantized DCT Coefficients	163
9.1	Introduction	163
9.2	Optimal Detector for Steganalysis of Jsteg Algorithm	164
9.2.1	Hypothesis Testing Formulation	164
9.2.2	LRT for Two Simple Hypotheses	165
9.3	Quantitative Steganalysis of Jsteg Algorithm	168
9.3.1	ML Estimation of Embedding Rate	168

9.3.2	Revisiting WS estimator	169
9.4	Numerical Experiments	169
9.4.1	Detection Performance of the proposed LRT	169
9.4.2	Accuracy of the Proposed Estimator	172
9.5	Conclusion	173
10	Conclusions and Perspectives	175
10.1	Conclusions	175
10.2	Perspectives	178
10.2.1	Perspectives to Digital Forensics	178
10.2.2	Perspectives to Statistical Image Modeling	180
10.2.3	Perspectives to Statistical Hypothesis Testing Theory Applied for Digital Forensics	180
A	Statistical Hypothesis Testing Theory	181
A.1	Introduction	181
A.2	Basic Concepts	182
A.3	Test between Two Simple Hypotheses	183
A.4	Test between Two Composite Hypotheses	187
A.5	Conclusion	191
	Bibliography	193

List of Figures

1.1	Example of falsification.	2
1.2	Structure of the work presented in this thesis.	4
2.1	Image processing pipeline of a digital camera.	13
2.2	Sample color filter arrays.	14
2.3	JPEG compression chain.	17
2.4	Typical steganographic system.	29
2.5	Operations of LSB replacement (top) and Jsteg (bottom).	31
2.6	Diagram of transition probabilities between trace subsets under LSB replacement.	34
4.1	Scatter-plot of pixels' expectation and variance from a natural RAW image with ISO 200 captured by <i>Nikon D70</i> and <i>Nikon D200</i> cameras. The image is segmented into homogeneous segments. In each segment, the expectation and variance are calculated and the parameters (a, b) are estimated as proposed in Section 4.2.2. The dash line is drawn using the estimated parameters (a, b) . Only the red channel is used in this experiment.	51
4.2	Scatter-plot of pixels' mean and variance from JPEG images with ISO 200 issued from <i>Nikon D70</i> and <i>Nikon D200</i> cameras. The red channel is used in this experiment. The image is segmented into homogeneous segments to estimate local means and variances. The generalized noise model is used to fit to the data.	58
4.3	Estimated parameters $(\tilde{a}, \tilde{b})$ on JPEG images issued from different camera models in Dresden image database.	63
4.4	Comparison between the proposed method and Farid's for estimation of gamma factor on JPEG images issued from <i>Nikon D200</i> camera model.	64
4.5	Comparison between the Laplacian, $G\Gamma$ and proposed model of DCT coefficients.	67
4.6	Comparison between the quantized Laplacian, quantized $G\Gamma$ and proposed model for quantized AC coefficient.	68
4.7	Averaged χ^2 test statistics of GG, $G\Gamma$ and proposed model for 63 quantized AC coefficients	71
5.1	Estimated camera parameters (a, b) on 20 RAW images of different camera model with ISO 200 and different camera settings.	79
5.2	Estimated camera parameters (a, b) of different devices per camera model with ISO 200 and different camera settings.	79

5.3	The detection performance of the GLRT $\bar{\delta}_{\text{het}}^*$ with 50 pixels selected randomly from simulated images.	85
5.4	The detection performance of the GLRT $\tilde{\delta}_{\text{het}}^*$ with 50 pixels selected randomly on simulated images.	88
5.5	The detection performance of the test $\bar{\delta}_{\text{het}}^*$ with 200 pixels selected randomly on simulated images for $a_0 = 0.0115$ and different parameters a_1	89
5.6	The detection performance of the test $\bar{\delta}_{\text{het}}^*$ and $\tilde{\delta}_{\text{het}}^*$ on simulated images for different numbers of pixels.	89
5.7	The detection performance of the GLRTs $\bar{\delta}_{\text{het}}^*$ and $\tilde{\delta}_{\text{het}}^*$ on the Dresden database for different numbers of pixels.	91
5.8	Comparison between the theoretical false alarm probability (FAP) and the empirical FAP, from real images of Dresden database, plotted as a function of decision threshold τ	92
6.1	Empirical distribution of noise residuals $\tilde{z}_{k,i}^{\text{res}}$ in a segment compared with theoretical Gaussian distribution.	101
6.2	Estimated parameters $(\tilde{a}, \tilde{b})$ on JPEG images issued from <i>Canon Ixus 70</i> camera with different camera settings.	102
6.3	Estimated parameters $(\tilde{a}, \tilde{b})$ on JPEG images issued from different devices of <i>Canon Ixus 70</i> model.	103
6.4	Detection performance of the proposed tests for 50 and 100 pixels extracted randomly from simulated JPEG images with quality factor 100.	108
6.5	Detection performance of the GLRT $\tilde{\delta}_{\text{gen}}^*$ for 100 pixels extracted randomly from simulated JPEG images with different quality factors.	109
6.6	Detection performance of the GLRT $\bar{\delta}_{\text{gen}}^*$ and $\tilde{\delta}_{\text{gen}}^*$ for 50 and 100 pixels extracted randomly from JPEG images of <i>Nikon D70</i> and <i>Nikon D200</i> cameras.	110
6.7	Comparison between the theoretical false alarm probability (FAP) and the empirical FAP, plotted as a function of decision threshold τ	111
7.1	Estimated parameters (α, β) at frequency $(0, 1)$ and $(8, 8)$ of uniform images generated using $\tilde{a} = 0.1$, $\tilde{b} = 2$, $\gamma = 2.2$	117
7.2	Estimated parameters (α, β) at frequency $(8, 8)$ of natural JPEG images issued from <i>Canon Ixus 70</i> and <i>Nikon D200</i> camera models.	119
7.3	Estimated parameters $(\tilde{c}, \tilde{d})$ at frequency $(8, 8)$ of natural JPEG images issued from different camera models in Dresden database.	120
7.4	Detection performance of proposed tests on simulated vectors with 1024 coefficients.	124
7.5	Detection performance of proposed tests on simulated vectors with 4096 coefficients.	125

7.6	Detection performance of proposed GLRTs for 1024 coefficients at frequency (8, 8) extracted randomly from simulated images with different quality factors.	126
7.7	Detection performance of proposed tests for different number of coefficients at frequency (8, 8) of natural JPEG images taken by <i>Canon Ixus 70</i> and <i>Nikon D200</i> camera models.	127
7.8	Detection performance of the GLRT $\tilde{\delta}_{\text{det}}^*$ for 4096 coefficients at different frequencies of natural JPEG images taken by <i>Canon Ixus 70</i> and <i>Nikon D200</i> camera models.	128
7.9	Comparison between the theoretical false alarm probability (FAP) and the empirical FAP, plotted as a function of decision threshold τ , of the proposed tests at the frequency (8, 8) of natural images.	129
8.1	Detection performance on non-clipped simulated images for embedding rate $R = 0.05$	147
8.2	Detection performance on clipped simulated images for embedding rate $R = 0.05$	149
8.3	Detection performance on real clipped images for embedding rate $R = 0.2$	151
8.4	Detection performance on real clipped images for embedding rate $R = 0.4$	152
8.5	Detection performance on real clipped images for embedding rate $R = 0.6$	152
8.6	Detection performance on 12-bit images taken by Canon 400D with ISO 100 from BOSS database for embedding rate $R = 0.05$	153
8.7	Detection performance on 5000 images from BOSS database for embedding rate $R = 0.05$	153
8.8	Empirical false-alarm probability from real images of BOSS database plotted as a function of decision threshold, compared with theoretical FAP.	154
9.1	Detection performance of the test δ_{jst}^* based on the proposed model with embedding rate $R = 0.05$ on the simulated images and real images.	167
9.2	Detection performance of the test δ_{jst}^* based on the quantized Laplacian, quantized GG, quantized GF, and proposed model on the BOSS-Base with embedding rate $R = 0.05$	170
9.3	Detection performance of the test δ_{jst}^* based on the quantized Laplacian, quantized GG, quantized GF, and proposed model on the subset of 1000 images from the BOSSBase with embedding rate $R = 0.05$	171
9.4	Comparison between the proposed test δ_{jst}^* , ZMH-Sym detector, ZP detector, WS detector and quantized Laplacian-based test.	172
9.5	Mean absolute error for all estimators.	172
9.6	Mean absolute error for proposed ML estimator, standard WS estimator and improved WS estimator.	173

9.7 Comparison between the proposed test δ_{jst}^* , standard WS detector and improved WS detector.	174
--	-----

List of Tables

4.1	Parameter estimation on synthetic images	63
4.2	PSNR of the extended LLMMSE filter	64
4.3	χ^2 test statistics of Laplacian, GG, GF, and proposed model for the first 9 quantized coefficients of 3 testing standard images.	71
5.1	Camera Model Used in Experiments (the symbol * indicates our own camera)	92
5.2	Detection performance of the proposed detector.	93
5.3	Detection performance of PRNU-based detector for ISO 200.	93
6.1	Camera Model Used in Experiments	112
6.2	Performance of proposed detector	112
6.3	Performance of SVM-based detector (the symbol * represents values smaller than 2%)	113
6.4	Performance of PRNU-based detector	113
7.1	Camera Model Used in Experiments	129
7.2	Detection performance of proposed detector $\tilde{\delta}_{\text{dct}}^*$ (the symbol * represents values smaller than 2%)	130
7.3	Detection performance of SVM-based detector	130
7.4	Detection performance of PRNU-based detector	131
7.5	Detection performance of proposed detector $\bar{\delta}_{\text{dct}}^*$	131
7.6	Detection performance of proposed detector $\tilde{\delta}_{\text{dct}}^*$ on 4 camera models of BOSS database	131
10.1	List of proposed statistical tests.	177

List of Abbreviations

Acronym	What (it) Stands For
AC	Alternating Current.
AUMP	Asymptotically Uniformly Most Powerful.
AWGN	Additive White Gaussian Noise.
cdf	cumulative distribution function.
CCD	Charge-Coupled Device.
CFA	Color Filter Array.
CLT	Central Limit Theorem.
CMOS	Complementary Metal-Oxide Semiconductor.
CRF	Camera Response Function.
DC	Direct Current.
DCT	Discrete Cosine Transform.
DSC	Digital Still Camera.
DSLR	Digital Single Lens Reflex.
EXIF	Exchangeable Image File.
GG	Generalized Gaussian.
GLR	Generalized Likelihood Ratio.
GLRT	Generalized Likelihood Ratio Test.
GOF	goodness-of-fit.
GF	Generalized Gamma.
IDCT	Inverse Discrete Cosine Transform.
i.i.d	independent and identically distributed.
JPEG	Join Photographic Expert Group.
KS	Kolmogorov-Smirnov.
LSB	Least Significant Bit.
LR	Likelihood Ratio.
LRT	Likelihood Ratio Test.
LS	Least Squares.
mAE	Median Absolute Error.
MAE	Mean Absolute Error.
MGF	Moment-Generating Function
ML	Maximum Likelihood.
MM	Method of Moments.
MP	Most Powerful.
MSE	Mean Squared Error.
NP	Neyman-Pearson.
PCE	Peak to Correlation Energy.
pdf	probability density function.
PRNU	Photo-Response Non-Uniformity.
RLE	Run-Length Encoding.

ROC	Receiver Operating Characteristic.
R/T	Rounding and Truncation.
SPA	Sample Pair Analysis.
SPN	Sensor Pattern Noise.
SVM	Support Vector Machine.
TIFF	Tagged Image File Format.
UMP	Uniformly Most Powerful.
WLS	Weighted Least Squares.
WS	Weighted Stego-image.
ZMH	Zero Message Hypothesis.
ZP	Zhang and Ping.

Glossary of Notations

Notation	Definition
α_0	False alarm probability.
$\beta(\delta)$	Power of the test δ .
χ_K^2	Chi square distribution with K degree of freedom.
γ	Gamma factor.
δ	Statistical test.
η	Noise.
κ	Quantization step in the spatial domain.
μ	Expectation.
ν	Bit-depth.
σ	Standard deviation.
τ	Decision threshold.
ξ	Number of collected electrons, which is modeled by Poisson distribution.
φ	2-D normalized wavelet scaling function.
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.
Cov	Covariance.
$\mathbb{E}$	Mathematical expectation.
$\mathbb{P}[E]$	Probability that an event E occurs.
$\mathbb{R}$	Set of real numbers.
Var	Variance.
$\mathbb{Z}$	Set of integer numbers.
$\mathcal{B}(n, p)$	Binomial distribution where n is the number of experiments and p is the success probability of each experiment.
$\mathcal{D}$	Denoising filter.
$\mathcal{G}(\alpha, \beta)$	Gamma distribution with shape parameter α and scale parameter β .
$\mathcal{H}_0, \mathcal{H}_1$	Null hypothesis and alternative hypothesis.
$\mathcal{I}$	Set of pixel indices.
$\mathcal{K}_{\alpha_0}$	Class of tests whose false alarm probability is upper-bounded by α_0 .
$\mathcal{L}$	Log-likelihood function.
$\mathcal{N}(\mu, \sigma^2)$	Gaussian distribution with mean μ and variance σ^2 .

$\mathcal{P}(\lambda)$	Poisson distribution with mean λ and variance λ .
$\mathcal{Q}_\Delta$	Quantization with step Δ .
$\mathcal{S}$	Source of digital images.
$\mathcal{U}[a, b]$	Uniform distribution on the interval $[a, b]$.
$\mathcal{Z}$	Set of possible pixel values.
$\mathcal{Z}^N$	Image space.

C	Cover-image that is used for data hiding.
D	Quantized DCT coefficients.
F	Fisher information.
H_{DM}	Linear filter for demosaicing.
I	Unquantized DCT coefficients.
Id_n	Identity matrix of size $n \times n$.
K	PRNU.
M	Secret message to be embedded.
P_{CFA}	CFA pattern.
S	Stego-image that contains hidden data.
Z	Natural image.

(a, b)	Parameters of heteroscedastic noise model.
$(\tilde{a}, \tilde{b})$	Parameters of generalized noise model.
c	Color channel.
$(\tilde{c}, \tilde{d})$	Parameters characterizing the relation between α and β , which are the parameters of DCT coefficient model.
f_{CRF}	Camera response function.
$f_X(x)$	Probability density function of a random variable X .
g_{WB}	Gain for white-balancing.
$p_X(x)$	Probability mass function of a random variable X .
r	Change rate, $r = \frac{R}{2}$.
B	Boundary of the dynamic range, $B = 2^\nu - 1$.
L	Secret message length.
N	Number of pixels in the image Z , $N = N_r \times N_c$ if Z is a grayscale image and $N = N_r \times N_c \times 3$ if Z is a full-color image.
N_{blk}	Number of blocks.
N_c	Number of columns.
N_r	Number of rows.
N_s	Number of sources.
P_θ	Probability distribution characterized by a parameter vector θ .
$Q_{R,\theta}$	Probability distribution after embedding at rate R .
R	Embedding rate.

General Introduction

1.1 General Context and Problem Description

Traditionally, images are considered as trustworthy since as they are known as being captured through analog acquisition devices to depict the real-world happenings. This traditional trustworthiness is built on remarkable difficulties of image content modification. Indeed, modifying the content of a film-based photo requires special skills, yet time-consuming and costly, through dark room tricks. Therefore, this modification is of limited extent.

In the past decades, we have witnessed the evolution of digital imaging technology with a dramatic improvement of digital images' quality. This improvement is not only due to advances in semiconductor fabrication technology that makes it possible to reduce the pixel size in an image sensor and thus raises the total number of pixels, but also advances in image processing technology that allows to reduce noise introduced in a camera and enhance details of the physical scene. The digital revolution largely replace their analog counterparts to enable ease of digital content creation and processing at affordable cost and in mass scale. Nowadays, digital still cameras (DSCs) are taking over a major segment of the consumer photography marketplace. Only at the very high end (large format, professional cameras with interchangeable and highly adjustable lenses) and very low end (inexpensive automated snapshot cameras) are traditional film cameras holding their own. Besides, the development of communication and networking infrastructure allows digital content to be more accessible. One of the greatest advantage of digital images acquired by DSCs is the ease of transmission over communication networks, which film cameras are difficult to enable.

Unfortunately, this path of technological evolution may provide means for malicious purposes. Digital images can be easily edited, altered or falsified because of a large availability of low-cost image editing tools. Consequently, falsified photographs are appearing with a growing frequency and sophistication. The credibility and trustworthiness of digital images have been eroded. This is more crucial when falsified images that were utilized as evidence in a courtroom could mislead the judgement and lead to either imprisonment for the innocent or freedom for the guilty. In general, the falsification of digital images may result in important consequences in terms of political, economic, and social issues.

One example of falsification that causes political issues is given in Figure 1.1. In the left corner image, President G.W. Bush and a young child are both reading from *A Patriotic Primer* by Lynne Cheney. But if we look closely, it

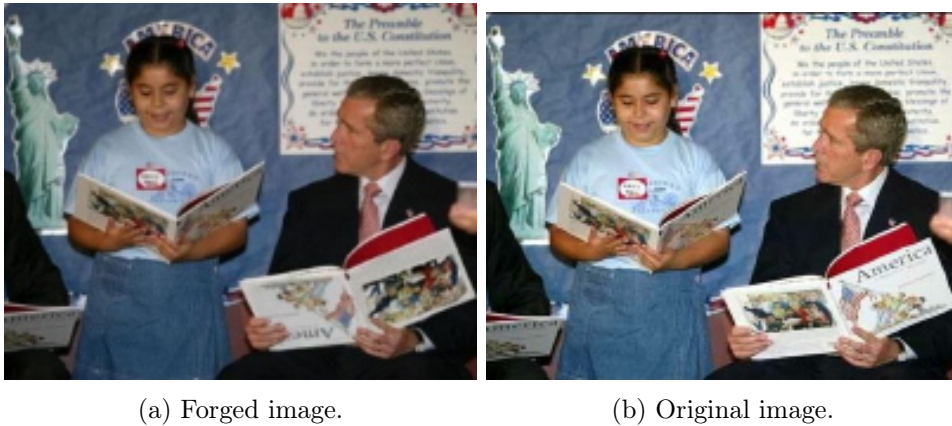


Figure 1.1: Example of falsification.

appears that President Bush is holding his book upside down. An unknown hoaxer has horizontally and vertically flipped the image on the back of the book in Bush's hands. This photo of George Bush holding a picture book the wrong way up during a visit to a school delighted some opponents of the Republican president, and helped foster his buffoonish image. But press photos from the event in 2002 revealed that Mr Bush had been holding the book correctly, i.e. hoaxers had simply used photo editing software to rotate the cover. The original version of the photo (right corner) was taken in the Summer of 2002 while Bush was visiting George Sanchez Charter School in Houston. It was distributed by the Associated Press. By comparing the forged photo and original photo, it can be noted that a dark blue spot is close to the spine of Bush's book, but this same spot in the girl's copy is near the left-hand edge of the book. This forensic clue can be considered as evidence of forgery. However in most of the cases, the forgery is not as easy to detect. The human eyes can hardly differentiate a genuine scene from a deliberately forged scene. Overall, the digital revolution has raised a number of information security challenges.

To restore the trust to digital images, the field of *digital image forensics* was born. Because of importance of information security in many domains, digital image forensics has attracted a great attention from academic researchers, law enforcement, security, and intelligence agencies. Conducting forensic analysis is a difficult mission since forensic analysts need to answer several questions before stating that digital content is authentic:

1. What is the true origin of this content? How was it generated? By whom was it taken?
2. Is the image still depicting the captured original scene? Has its content been altered in some way? How has it been processed?

The first question involves the problem of *image origin identification*. Source information of digital images represents useful forensic clues because knowing the

source device that captures the inspected image can facilitate verification or tracing of device owner as well as the camera taker. This situation is as identical as bullet scratches allowing forensic analysts to match a bullet to a particular barrel or gun and trace the gun owner.¹ Besides, knowing device model or brand information can help forensic analysts know more about characteristics of acquisition devices, which leads to a potential improvement of detecting the underlying forgeries that could be performed in the inspected image. Another issue is to determine what imaging mechanism has been used to generate the inspected image (e.g. scanners, cell-phone cameras, or computer graphics) before assuming that the inspected image is taken by a digital camera, which can significantly narrow down the search range for the next step of the investigation.

The second problem is *image content integrity*. An image has to be proven authentic and its content has not be forged before it can be used as forensic clues or as evidence in a legal context. Determining whether an image is forged, which manipulation has been performed on the image, or which region of the image has been altered are fundamental tasks.

Beside some basic manipulations such as adding, splicing, and removal, the image can be also manipulated by embedding a message into image content directly. The message remains secret such that it is only known by the sender and receiver and an adversary does not recognize its existence visually. This concept is called *steganography*, which is a discipline of the field of information hiding. However, the concept of steganography has been misused for illegal activities. Detecting existence of secret messages and revealing their content are also the tasks of forensic analysts. This task is called *steganalysis*.

The field of digital image forensics, including steganalysis, is part of an effort to counter cyber-attacks, which is nowadays one of strategy priorities for defence and national security in most countries.

1.2 Outline of the Thesis

The main goal of this thesis is to address information security challenges in the field of digital image forensics. In particular, the problems of image origin identification and hidden data detection are studied. The thesis is structured in four main parts. Apart from the first part providing an overview on the field of digital image forensics and statistical image modeling, the rest of the thesis involves many contributions. All the work presented in this thesis is illustrated in Figure 1.2.

Part II establishes a profound statistical modeling of natural images by analyzing the image processing pipeline of a digital camera, as well as proposes efficient algorithms for estimation of model parameters from a single image. Typically, the image processing pipeline is composed of three main stages: RAW image acquisition,

¹Evidently, tracing an imaging device owner is more difficult as average users have rights to buy a camera easily in a market with millions of cameras while the use of guns is banned or controlled in many countries and a gun user has to register his identities.

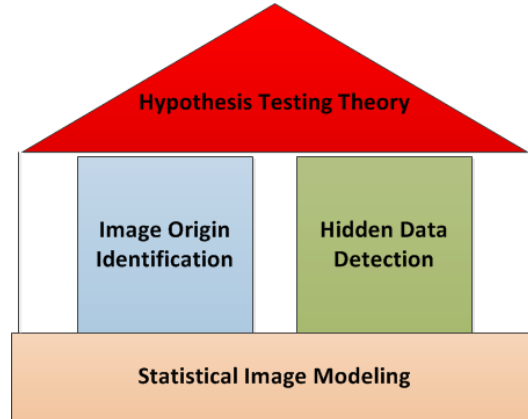


Figure 1.2: Structure of the work presented in this thesis.

post-acquisition enhancement, and JPEG compression that employs Discrete Cosine Transform (DCT). Therefore, the statistical image modeling in Part II is performed both in the spatial domain and the DCT domain. By modeling the photo-counting and read-out processes, a RAW image can be accurately characterized by the heteroscedastic noise model in which the RAW pixel is normally distributed and its variance is linearly dependent on its expectation. This model is more relevant than the Additive White Gaussian Noise (AWGN) model widely used in image processing since the latter ignores the contribution of Poisson noise in the RAW image acquisition stage. The RAW image then undergoes post-acquisition processes in order to provide a high-quality full-color image, referred to as TIFF image. Therefore, to study image statistics in a TIFF image, it is proposed to start from the heteroscedastic noise model and take into account non-linear effect of gamma correction, resulting in a generalized noise model. This latter involves a non-linear relation between pixel's expectation and variance. This generalized noise model has not been proposed yet in the literature. Overall, the study of noise statistics in the spatial domain indicates the non-stationarity of noise in a natural image, i.e. pixel's variance is dependent on the expectation rather than being constant in the whole image. Besides, pixels' expectations, namely the image content, are also heterogeneous. Apart from studying image statistics in the spatial domain, it is proposed to study DCT coefficient statistics. Modeling the distribution of DCT coefficients is not a trivial task due to heterogeneity in the natural image and complexity of image statistics. It is worth noting that most of existing models of DCT coefficients, which are only verified by conducting the goodness-of-fit test with empirical data, are given without a mathematical justification. Instead, this thesis provides a mathematical framework of modeling the statistical distribution of DCT coefficients by relying on the double stochastic model that combines the statistics of DCT coefficients in a block whose variance is constant with the variability of block variance in a natural image. The proposed model of DCT coefficients outperforms the others including the Laplacian, Generalized Gaussian, and Generalized Gamma model. Numerical

results on simulated database and real image database highlight the relevance of the proposed models and the accuracy of the proposed estimation algorithms.

The solid foundation established in Part II emphasizes several aspects of interest for application in digital image forensics. Relying on a more relevant image model and an accurate estimation of model parameters, the detector is expected to achieve better detection performance. Part III addresses the problem of image origin identification within the framework of hypothesis testing theory. More particularly, it involves designing a statistical test for camera model identification based on a parametric image model to meet the optimality bi-criteria: the warranting of a prescribed false alarm probability and the maximization of the correct detection probability. Camera model identification based on the heteroscedastic noise model, generalized noise model, and DCT coefficients is respectively presented in Chapter 5, Chapter 6, and Chapter 7. The model parameters are proposed to be exploited as unique fingerprint for camera model identification. In general, the procedure in those chapters is similar. The procedure starts from formally stating the problem of camera model identification into hypothesis testing framework. According to Neyman-Pearson lemma, the most powerful test for the decision problem is given by the Likelihood Ratio Test (LRT). The statistical performance of the LRT can be analytically established. Moreover, the LRT can meet the two required criteria of optimality. However, this test is only of theoretical interest because it is based on an assumption that all model parameters are known in advance. This assumption is hardly met in practice. To deal with the difficulty of unknown parameters, a Generalized Likelihood Ratio Test (GLRT) is proposed. The GLRT is designed by replacing unknown parameters by their Maximum Likelihood (ML) estimates in the Likelihood Ratio. Consequently, the detection performance of the GLRT strongly depends on the accuracy of employed image model and parameter estimation. It is shown in Chapter 5, 6, and 7 that the proposed GLRTs can warrant a prescribed false alarm probability while ensuring a high detection performance. Moreover, the efficiency of the proposed GLRTs is highlighted when applying on a large image database.

The problem of hidden data detection is addressed in Part IV. This problem is also formulated into hypothesis testing framework. The main idea is to rely on an accurate image model to detect small changes in statistical properties of a cover image due to message embedding. The formulation in the hypothesis testing framework allows us to design a test that can meet two above criteria of optimality. Chapter 8 addresses the steganalysis of Least Significant Bit (LSB) replacement technique in RAW images. More especially, the phenomenon of clipping is studied and taken into account in the design of the statistical test. This phenomenon is due to limited dynamic range of the imaging system. The impact of the clipping phenomenon on the detection performance of steganalysis methods has not been studied yet in the literature. The approach proposed in Chapter 8 is based on the heteroscedastic noise model instead of the AWGN model. Besides, the approach proposes to exploit the state-of-the-art denoising method to improve the estimation of pixels' expectation and variance. The detection performance of the proposed

GLRTs on non-clipped images and clipped images is studied. It is shown that the proposed GLRTs can warrant a prescribed false alarm probability and achieve a high detection performance while other detectors fail in practice, especially the Asymptotically Uniformly Most Powerful (AUMP) test. Next, Chapter 9 addresses the steganalysis of Jsteg algorithm. It should be noted that Jsteg algorithm is a variant of LSB replacement technique. Instead of embedding message bits in the spatial domain, Jsteg algorithm utilizes the LSB of quantized DCT coefficients and embeds message bits in the DCT domain. The goal of Chapter 9 is to exploit the state-of-the-art model of quantized DCT coefficients in Chapter 4 to design a LRT for the steganalysis of Jsteg algorithm. For the practical use, unknown parameters of the DCT coefficient model are replaced by their ML estimates in the Likelihood Ratio. Experiments on simulated database and real image database show a very small loss of power of the proposed test. Furthermore, the proposed test outperforms other existing detectors. Another contributions in Chapter 9 are that a Maximum Likelihood estimator for embedding rate is proposed using the proposed model of DCT coefficients as well as the improvement of the existing Weighted Stego-image estimator by modifying the technique of calculation of weights.

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.

Patents

1. **T. H. Thai**, R. Cogranne, and F. Retraint, "*Système d'identification d'un modèle d'appareil photographique associé à une image compressée au format JPEG, procédé, utilisations and applications associés*", PS/B52545/FR, 2014.
2. **T. H. Thai**, R. Cogranne, and F. Retraint, "*Système d'identification d'un modèle d'appareil photographique associé à une image compressée au format JPEG, procédé, utilisations and applications associés*", PS/B52546/FR, 2014.

Journal articles

1. **T. H. Thai**, R. Cogranne, and F. Retraint, "Camera model identification based on the heteroscedastic noise model", *IEEE Transactions on Image Processing*, vol. 23, no. 1, pp. 250-263, Jan. 2014.
2. **T. H. Thai**, F. Retraint, and R. Cogranne, "Statistical detection of data hidden in least significant bits of clipped images", *Elsevier Signal Processing*, vol. 98, pp. 263-274, May 2014.
3. **T. H. Thai**, R. Cogranne, and F. Retraint, "Statistical model of quantized DCT coefficients: application in the steganalysis of Jsteg algorithm", *IEEE Transactions on Image Processing*, vol. 23, no. 5, pp. 1980-1993, May 2014.

Journal articles under review

1. **T. H. Thai**, F. Retraint, and R. Cogranne, "Generalized signal-dependent noise model and parameter estimation for natural images", 2014.
2. **T. H. Thai**, F. Retraint, and R. Cogranne, "Camera model identification based on the generalized noise model in natural images", 2014.
3. **T. H. Thai**, R. Cogranne, and F. Retraint, "Camera model identification based on DCT coefficient statistics", 2014.

Conference papers

1. **T. H. Thai**, F. Retraint, and R. Cogranne, "Statistical model of natural images", in *IEEE International Conference on Image Processing*, pp. 2525-2528, Sep. 2012.
2. **T. H. Thai**, R. Cogranne, and F. Retraint, "Camera model identification based on hypothesis testing theory", in *European Signal Processing Conference*, pp. 1747-1751, Aug. 2012.
3. **T. H. Thai**, R. Cogranne, and F. Retraint, "Steganalysis of Jsteg algorithm based on a novel statistical model of quantized DCT coefficients", in *IEEE International Conference on Image Processing*, pp. 4427-4431, Sep. 2013.
4. R. Cogranne, **T. H. Thai**, and F. Retraint, "Asymptotically optimal detection of LSB matching data hiding", in *IEEE International Conference on Image Processing*, pp. 4437-4441, Sep. 2013.
5. **T. H. Thai**, R. Cogranne, and F. Retraint, "Optimal detector for camera model identification based on an accurate model of DCT coefficient", in *IEEE International Workshop on Multimedia Signal Processing (in press)*, Sep. 2014.
6. **T. H. Thai**, R. Cogranne, and F. Retraint, "Optimal detection of OutGuess using an accurate model of DCT coefficients", in *IEEE International Workshop on Information Forensics and Security (in press)*, Dec. 2014.

Part I

Overview on Digital Image Forensics and Statistical Image Modeling

Overview on Digital Image Forensics

Contents

2.1	Introduction	11
2.2	Image Processing Pipeline of a Digital Camera	12
2.2.1	RAW Image Formation	13
2.2.2	Post-Acquisition Processing	15
2.2.3	Image Compression	17
2.3	Passive Image Origin Identification	19
2.3.1	Lens Aberration	23
2.3.2	Sensor Imperfections	23
2.3.3	CFA Pattern and Interpolation	25
2.3.4	Image Compression	26
2.4	Passive Image Forgery Detection	27
2.5	Steganography and Steganalysis in Digital Images	29
2.5.1	LSB Replacement Paradigm and Jsteg Algorithm	32
2.5.2	Steganalysis of LSB Replacement in Spatial Domain	33
2.5.2.1	Structural Detectors	33
2.5.2.2	WS Detectors	35
2.5.2.3	Statistical Detectors	36
2.5.2.4	Universal Classifiers	37
2.5.3	Steganalysis of Jsteg Algorithm	38
2.6	Conclusion	39

2.1 Introduction

The goal of this chapter is to provide an overview on the field of digital image forensics. As described in Section 1.1, digital image forensics involve two key problems: image origin identification and image forgery detection. In general, there are two approaches to address these problems. *Active forensics* aims to authenticate image content by generating extrinsically security measures such as digital watermarks [1–6] and digital signatures [7–10] and adding them to the image file. These

security measures are referred to as *extrinsic fingerprints*. Although active forensics can provide powerful tools to secure a digital camera and restore the credibility of digital images, it is of limited extent due to many strict constraints in its protocols. In order to solve these problems in their entirety, *passive forensics* has been quickly evolved. In contrast to active forensics, passive forensics does not impose any constraint and do not require any prior information including the original reference image. Forensic analysts have only the suspect image at their disposal and must explore useful information from the image to gather forensic evidence, trace the acquisition device and detect any act of manipulation. Passive forensics works on an assumption that the image contains some internal traces left from the camera. Every stage from real-world scene acquisition to image storage can provide clues for forensic analysis. These internal traces are called *intrinsic fingerprints*. Extrinsic and intrinsic fingerprints are two forms of digital fingerprints in digital forensics, which are analogous to human fingerprints in criminal domain. Since passive forensics does not require neither any external security measures generated in the digital camera, nor any prior information, it can authenticate an image in a blind manner and can widely be applied to millions of images that circulate daily on communication networks.

This thesis mainly addresses the problem of origin identification and integrity based on passive approach. The chapter is organized as follows. Before discussing active and passive forensics, it is vital to understand deeply the creation and characteristics of digital images. Section 2.2 briefly introduces the typical image processing pipeline of a digital camera, highlighting several aspects of potential interest for applications in digital image forensics. Section 2.3 analyzes passive methods proposed for image origin identification. Section 2.4 briefly discusses passive methods for image forgery detection. Next, Section 2.5 introduces the concept of steganography, which is a type of image content manipulation, and presents prior-art methods for detecting secret data embedded in digital images. Finally, Section 2.6 concludes the chapter.

2.2 Image Processing Pipeline of a Digital Camera

This thesis only deals with DSCs and digital images acquired by them. By terminology, a *natural image* means a digital image acquired by a DSC. Other sources of digitized images such as scanners are not addressed in this thesis but a similar methodology can be easily derived.

Image processing pipeline involves several steps from light capturing to image storage performed in a digital camera [11]. After measuring light intensity at each pixel, RAW image that contains exactly information recorded by the image sensor goes through some typical post-acquisition processes, e.g. demosaicing, white-balancing and gamma correction, to render a full-color high-quality image, referred to as TIFF image. Image compression can be also performed for ease of storage and transmission. The image processing pipeline of a digital camera is shown in

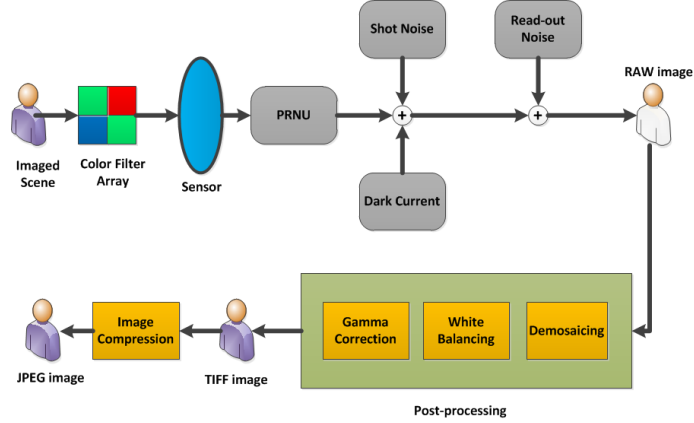


Figure 2.1: Image processing pipeline of a digital camera.

Figure 2.1. It should be noted that the sequence of operations differs from manufacturer to manufacturer but basic operations remain similar. In general, the image processing pipeline designed in a digital camera is complex, with trade-offs in the use of buffer memory, computing operations, image quality and flexibility [12]. This section only discusses some common image processing operations such as demosaicing, white balancing, gamma correction and image compression. Other processing operations, e.g. camera noise reduction and edge enhancement, are not included in this discussion.

A full-color digital image consists of three primary color components: red, green, and blue. These three color components are sufficient to represent millions of colors. Formally, the full-color image of a DSC can be represented as a three-dimensional matrix of size $N_r \times N_c \times 3$ where N_r and N_c are respectively the number of rows and columns. Let $c \in \{R, G, B\}$ denote a color channel where R, G and B stand for respectively the red, green and blue color. Typically, the output image is coded with ν bits and each pixel value is a natural integer. The set of possible pixel values is denoted by $\mathcal{Z} = \{0, 1, \dots, B\}$ with $B = 2^\nu - 1$. Therefore, an arbitrary image belongs to the finite image space $\mathcal{Z}^N$ with $N = N_r \times N_c \times 3$. In general, the image space $\mathcal{Z}^N$ is high dimensional because of a large number of pixels. To facilitate discussions, let $\mathbf{Z}$ denote an image in RAW format and $\tilde{\mathbf{Z}}$ denote an image in TIFF or JPEG format. Each color component of the image $\mathbf{Z}$ is denoted by $\mathbf{Z}^c$ and a pixel of the color channel c at the location (m, n) is denoted by $z^c(m, n)$, $1 \leq m \leq N_r$, $1 \leq n \leq N_c$.

2.2.1 RAW Image Formation

Typically, a digital camera includes an optical sub-system (e.g. lenses), an image sensor and an electronic sub-system, which can be regarded as the eye, retina, and brain in the human visual system. The optical sub-system allows to attenuate effects of infrared rays and to provide an initial optic image. The image sensor consists of a two-dimensional arrays of photodiodes (or *pixels*) fabricated on a silicon



Figure 2.2: Sample color filter arrays.

wafer. Two common types of an image sensor are Charge-Coupled Device (CCD) and Complementary Metal-Oxide Semiconductor (CMOS). Each pixel enables to convert light energy to electrical energy. The output signals of the image sensor are analog. These signals are then converted to digital signals by an analog-to-digital (A/D) converter inside the camera. The RAW image is obtained at this stage. Depending on the analog-to-digital circuit of the camera, the RAW image is recorded with 12, 14 or even 16 bits. One key advantage is that the RAW image exactly contains information recorded by the image sensor and it has not yet undergone post-acquisition operations. This offers more flexibility for further adjustments.

Although the image sensor is sensitive to light intensity, it does not differentiate light wavelength. Therefore, to record a color image, a Color Filter Array (CFA) is overlaid on the image sensor. Each pixel records a limited range of wavelength, corresponding to either red, green or blue. Some examples of CFA patterns are shown in Figure 2.2. Among available CFA patterns, the Bayer pattern is the most popular. It contains twice as many green as red or blue samples because the human eye is more sensitive to green light than both red or blue light. The higher rate of sampling for the green component enables to better capture the luminance component of light and, thus, provides better image quality. There are few digital cameras that allow to acquire a full-resolution information for all three color components (e.g. *Sigma SD9* or *Polaroid x530*). This is not only due to high production cost but also due to the requirement of a perfect alignment of three color planes together.

Let $\mathbf{Z}$ represent the RAW image recorded by the image sensor. Because of the CFA sampling, the RAW image $\mathbf{Z}$ is a single-channel image, namely that it is represented as a two-dimensional matrix of size $N_r \times N_c$. Each pixel value of the RAW image $\mathbf{Z}$ corresponds to only one color channel. For subsequent processing operations, each color component is extracted from the RAW image $\mathbf{Z}$. A pixel of each color component is given by

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

where $\mathbf{P}_{\text{CFA}}$ is the CFA pattern.

The RAW image acquisition stage is not ideal due to the degradation introduced by several noise sources. This stage involves two predominant random noise sources. The first is the Poisson-distributed noise associated with the stochastic nature of the photo-counting process (namely shot noise) and dark current generated by the thermal energy in the absence of light. Dark current is also referred to as Fixed Pattern Noise (FPN). While shot noise results from the quantum nature of light and it can not be eliminated, dark current can be subtracted from the image [13].

The second random noise sources account for all remaining electronic noises involved in the acquisition chain, e.g. read-out noise, which can be modeled by a Gaussian distribution with zero-mean. Apart from random noises, there is also a multiplicative noise associated with the sensor pattern. This noise accounts for differences of pixels response to the incident light due to imperfections during the manufacturing process and inhomogeneity of silicon wafers. Therefore, this noise is referred to as Photo-Response Non-Uniformity (PRNU). The PRNU, which is typically small compared with the signal, is a deterministic component that is present in every image. FPN and PRNU are two main components of Sensor Pattern Noise (SPN). The PRNU is unique for each sensor, thus it can be further used for forensic analysis.

2.2.2 Post-Acquisition Processing

Although the use of the CFA allows to reduce the cost of the camera, this requires to estimate the missing color values at each pixel location in order to render a full-color image. It means that all the zero values in the sub-images need to be interpolated. This estimation process is commonly referred to as *CFA demosaicing* or *CFA interpolation* [14]. Technically, demosaicing algorithms estimate a missing pixel value by using its neighborhood information. The performance of CFA demosaicing affects greatly the image quality. Demosaicing algorithms can be generally classified into two categories: non-adaptive and adaptive algorithms. Non-adaptive algorithms apply the same interpolation technique for all pixels. The nearest neighborhood, bilinear, bicubic, and smooth hue interpolations are typical examples in this category. For example, the bilinear interpolation can be written as a linear filtering

$$\mathbf{Z}_{\text{DM}}^c = \mathbf{H}_{\text{DM}}^c \circledast \mathbf{Z}^c, \quad (2.2)$$

where $\circledast$ denotes the two-dimensional convolution, $\mathbf{Z}_{\text{DM}}^c$ stands for the demosaiced image of the color channel c , and $\mathbf{H}_{\text{DM}}^c$ is the linear filter for the color channel c

$$\mathbf{H}_{\text{DM}}^G = \frac{1}{4} \begin{pmatrix} 0 & 1 & 0 \\ 1 & 4 & 1 \\ 0 & 1 & 0 \end{pmatrix}, \quad \mathbf{H}_{\text{DM}}^R = \mathbf{H}_{\text{DM}}^B = \frac{1}{4} \begin{pmatrix} 1 & 2 & 1 \\ 2 & 4 & 2 \\ 1 & 2 & 1 \end{pmatrix}. \quad (2.3)$$

Although non-adaptive algorithms can provide satisfactory results in smooth regions of an image, they usually fail in textured regions and edges. Therefore, adaptive algorithms, which are more computationally intensive, employ edge information or inter-channel correlation to find an appropriate set of coefficients to minimize the overall interpolation error. Because the CFA interpolation commonly estimates a missing pixel value using its neighbors, it thus creates a correlation between adjacent pixels. This spatial correlation may be amplified during subsequent processing stages.

Furthermore, to improve the visual quality, the RAW image needs to go through another processing step, e.g. *white balancing* [11]. In fact, an object may appear different in color when it is illuminated under different light sources. This is due to the color temperature difference of the light sources, which causes the shift of the

reflection spectrum of the object from the true color. In other words, when a white object is illuminated under a light source with low color temperature, the reflection become reddish. On the other hand, a light source with high color temperature can cause the white object to become bluish. The human visual system can recognize the white color of the white object under different light sources. This phenomenon is called *color constancy*. However, the digital camera does not have such luxury of millions of year of evolution as human visual system. Therefore, the white balance adjustment is implemented in the digital camera to compensate this illumination imbalance so that a captured white object is rendered white in the image. Basically, white balance adjustment is performed by multiplying pixels in each color channel by a different gain factor. For instance, one classical white balancing algorithm is the Gray World assuming that the average value of three color channels will average to a common gray value

$$\bar{z}_{\text{DM}}^{\text{R}} = \bar{z}_{\text{DM}}^{\text{G}} = \bar{z}_{\text{DM}}^{\text{B}}, \quad (2.4)$$

where $\bar{z}_{\text{DM}}^c$ denotes the average intensity of the demosaiced image $\mathbf{Z}_{\text{DM}}^c$

$$\bar{z}_{\text{DM}}^c = \frac{1}{N_{\text{r}} \cdot N_{\text{c}}} \sum_{m=1}^{N_{\text{r}}} \sum_{n=1}^{N_{\text{c}}} z_{\text{DM}}^c(m, n). \quad (2.5)$$

In this algorithm, the green channel is fixed because human eye is more sensitive to this channel (i.e. $g_{\text{WB}}^{\text{G}} = 1$). The gain factor for other color channels is given by

$$g_{\text{WB}}^{\text{R}} = \frac{\bar{z}_{\text{DM}}^{\text{G}}}{\bar{z}_{\text{DM}}^{\text{R}}}, \quad \text{and} \quad g_{\text{WB}}^{\text{B}} = \frac{\bar{z}_{\text{DM}}^{\text{G}}}{\bar{z}_{\text{DM}}^{\text{B}}}, \quad (2.6)$$

where g_{WB}^c denotes the gain factor of the color channel c for white balance adjustment. Therefore, the white-balanced image $\mathbf{Z}_{\text{WB}}^c$ is simply given by

$$\mathbf{Z}_{\text{WB}}^c = g_{\text{WB}}^c \cdot \mathbf{Z}_{\text{DM}}^c. \quad (2.7)$$

Other white-balancing algorithms may be also designed using different gain factors. Actually, the white balance adjustment is a difficult task due to estimation or selection of appropriate gain factors to correct for illumination imbalance. In this task, the prior knowledge of light sources is critical so that the camera knows to select appropriate gain factors. Therefore, some typical light sources such as daylight, incandescent or fluorescent are stored in the camera. The white balance can be done automatically in the camera. Some expensive cameras employ preprogrammed or manual white balance for adapting to illumination conditions correctly.

Generally, the pixel intensity is still linear with respect to scene intensity before gamma correction [11, 12]. However, most displays have non-linear characteristics. The transfer function of these devices can be fairly approximated by a simple power function that relates the luminance L to voltage V

$$L = V^\gamma. \quad (2.8)$$

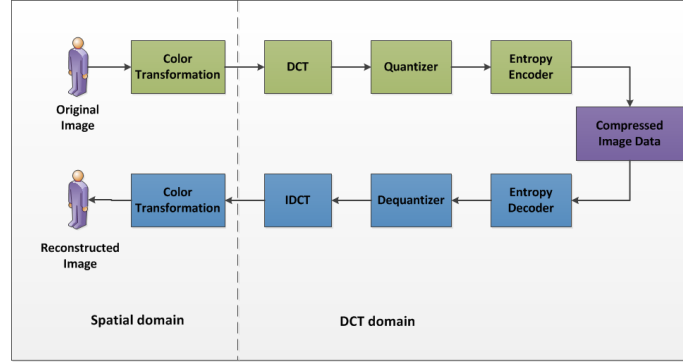


Figure 2.3: JPEG compression chain.

Typically, $\gamma = 2.2$. To compensate this effect and render the luminance into a perceptually uniform domain, the gamma correction is done in the image processing pipeline. Gamma correction is roughly the inverse of Equation (2.8), applying to each input pixel value

$$z_{\text{GM}}^c(m, n) = (z_{\text{WB}}^c(m, n))^{\frac{1}{\gamma}}. \quad (2.9)$$

After going through all post-acquisition processes, a full-color high-quality image, referred to as TIFF image, is rendered. For the sake of simplicity, let $\tilde{\mathbf{Z}}_{\text{TIFF}}$ denote the final full-color TIFF image.

2.2.3 Image Compression

The TIFF format does not make ease of storage or transmission. Therefore, most digital cameras commonly apply lossy compression algorithms to reduce the image data size. Lossy compression algorithms allow to discard information that is not visually significant. Therefore, lossy compression algorithms are irreversible when the image reconstructed from the compressed image data is not as identical as the original image. Moreover, the use of a lossy compression algorithm is a balancing act between storage size and image quality. An image which is compressed with a high compression factor requires little storage space, but it will probably be reconstructed with a poor quality.

Although many lossy compression algorithms have been proposed, most manufacturers predominately utilize JPEG compression. The JPEG compression scheme consists of three fundamental settings: color space, subsampling technique, and quantization table. Even though JPEG was already proposed by the standard Independent JPEG Group [15], manufacturers typically design their own compression scheme for optimal trade-off of image quality versus file size. Fundamental steps of the typical JPEG compression chain are shown in Figure 2.3.

The JPEG compression scheme works in the different color space, typically YCbCr color space, rather than the RGB color space. The transformation to the YCbCr color space is to reduce correlations among red, green and blue components. It allows for more efficient compression. The channel Y represents the luminance of

a pixel, and the channels Cb and Cr represent the chrominance. Each channel Y, Cb and Cr is processed separately. In addition, the channels Cb and Cr are commonly subsampled by a factor of 2 horizontally and vertically. The transformation from the RGB color space to the YCbCr color space is linear

$$\begin{pmatrix} Y \\ Cb \\ Cr \end{pmatrix} = \begin{pmatrix} 0.299 & 0.587 & 0.114 \\ -0.169 & -0.331 & 0.5 \\ 0.5 & -0.419 & 0.081 \end{pmatrix} \begin{pmatrix} R \\ G \\ B \end{pmatrix} + \begin{pmatrix} 0 \\ 128 \\ 128 \end{pmatrix}. \quad (2.10)$$

To avoid introducing too many symbols, let $\tilde{\mathbf{Z}}_{\text{TIF}}F$ denote also the image obtained after this transformation.

The JPEG compression algorithm consists of two key steps: Discrete Cosine Transform (DCT), and quantization. It works separately on each 8×8 block of a color component. The DCT operation converts pixel values from the spatial domain into transform coefficients

$$I(u, v) = \frac{1}{4} T_u T_v \sum_{m=0}^7 \sum_{n=0}^7 \tilde{z}_{\text{TIF}}F(m, n) \times \cos\left(\frac{(2m+1)u\pi}{16}\right) \cos\left(\frac{(2n+1)v\pi}{16}\right), \quad (2.11)$$

where $\tilde{z}_{\text{TIF}}F(m, n)$ is a pixel in a 8×8 block, $0 \leq m, n \leq 7$, $I(u, v)$ denotes the two-dimensional DCT coefficient, $0 \leq u, v \leq 7$, and T_u is the normalized weight

$$T_u = \begin{cases} \frac{1}{\sqrt{2}} & \text{for } u = 0 \\ 1 & \text{for } u > 0 \end{cases}. \quad (2.12)$$

The index of color channel Y, Cb, and Cr is omitted for simplicity as each color channel is processed separately. The coefficient at location $(0, 0)$, called the Direct Current (DC) coefficient, represents the mean value of pixels in the 8×8 block. The remaining 63 coefficients are called the Alternating Current (AC) coefficients. The DCT is known as sub-optimal transform with two important properties: energy compaction and decorrelation. In a natural image, the majority of the energy tends to be more located in low frequencies (i.e. the upper left corner of the 8×8 grid) while high frequencies contains information that is not visually significant.

Then, the DCT coefficients go through the quantization process. The quantization is carried out by simply dividing each coefficient by the corresponding quantization step, and then rounding to the nearest integer

$$D(u, v) = \text{round}\left(\frac{I(u, v)}{\Delta(u, v)}\right), \quad (2.13)$$

where $D(u, v)$ is the quantized DCT coefficient and $\Delta(u, v)$ is the corresponding element in the 8×8 quantization table Δ . The quantization table is designed differently for each color channel. The quantization is irreversible, which results in an impossibility of recovering the original image exactly. The final processing

step is entropy coding that is a form of lossless process. It arranges quantized DCT coefficients into the zig-zag sequence and then employs the Run-Length Encoding (RLE) algorithm and Huffman coding. This step is perfectly reversible.

The JPEG decompression works in the reverse order: entropy decoding, dequantization, and Inverse DCT (IDCT). When the image is decompressed, the entropy is decoded, we obtain the two-dimensional quantized DCT coefficients. The dequantization is performed by multiplying the quantized DCT coefficient $D(u, v)$ by the corresponding quantization step $\Delta(u, v)$

$$\underline{I}(u, v) = \Delta(u, v) \cdot D(u, v), \quad (2.14)$$

where $\underline{I}(u, v)$ stands for the dequantized DCT coefficient. The IDCT operation is applied to dequantized DCT coefficients to return to the spatial domain

$$\begin{aligned} \tilde{z}_{\text{IDCT}}(m, n) &= \sum_{u_h=0}^7 \sum_{u_v=0}^7 \frac{1}{4} T_u T_v \underline{I}(u, v) \\ &\times \cos\left(\frac{(2m+1)u\pi}{16}\right) \cos\left(\frac{(2n+1)v\pi}{16}\right). \end{aligned} \quad (2.15)$$

After upsampling color components and transforming into the RGB color space, the values are rounded to the nearest integers and truncated to a finite dynamic range (typically $[0, 255]$)

$$\tilde{z}_{\text{JPEG}}(m, n) = \text{trunc}\left(\text{round}\left[\tilde{z}_{\text{IDCT}}(m, n)\right]\right), \quad (2.16)$$

where $\tilde{z}_{\text{JPEG}}(m, n)$ is the final decompressed JPEG pixel. In general, the JPEG pixel $\tilde{z}_{\text{JPEG}}(m, n)$ differs from the original TIFF pixel $\tilde{z}_{\text{TIFF}}(m, n)$ due to the quantization, rounding and truncation (R/T) errors in the process. Note that in this image processing pipeline, R/T errors are only take into account one time for the sake of simplification.

The way that JPEG compression works separately on each 8×8 block generates discontinuities across the boundaries of the blocks, which are also known as block artifacts [16]. The blocking artifacts are more severe when the quantization steps are coarser. Moreover, because of the quantization in the DCT domain, the DCT coefficients obtained by applying the DCT operation on the decompressed JPEG image will cluster around integer multiples of $\Delta(u, v)$, even though those DCT coefficients are perturbed by R/T errors. These two artifacts provide a rich source of information for forensic analysis of digital images.

2.3 Passive Image Origin Identification

Basically, when an image is captured by a camera, it is stored with the metadata headers in a memory storage device. The metadata, e.g. Exchangeable Image File (EXIF) and JPEG headers, contain all recording and compression history. Therefore, a simplest way to determine the image's source is to read out directly from the

metadata. However, such metadata headers are not always available in practice if the image is resaved in a different format or recompressed. Another problem is that the metadata headers are not reliable as they can be easily removed or modified using low-cost editing tools. Therefore, the metadata should not be considered in forensic analysis.

The common philosophy of passive approach for image origin identification is to rely on inherent intrinsic fingerprints that the digital camera leaves in a given image. The fingerprint can discriminate different camera brands, camera models, and even camera units. Any method proposed for image origin identification must respond to following questions:

1. Which fingerprints are utilized for origin identification?
2. How to extract these fingerprints accurately from a given image?
3. Under which frameworks is the method designed to exploit the discriminability of fingerprints extracted from images captured by different sources¹ and to calculate the similarity of fingerprints extracted from images captured by the same source?

Every stage from real-world scene acquisition to image storage can provide intrinsic fingerprints for forensic analysis (see Figure 2.1). Although the image processing pipeline is common for most cameras, each processing step is performed according to manufacturers' own design. Thus the information left by each processing step is useful to trace down to the device source. A fingerprint must satisfy three following important requirements:

- Generality: the fingerprint should be present in every image.
- Invariant: the fingerprint does not vary for different image contents.
- Robustness: the fingerprint survives non-linear operations such as lossy compression or gamma correction.

The second question involves a challenge for any forensic method since the fingerprint extraction may be severely contaminated by non-linear operations (e.g. gamma correction and lossy compression).

Generally, the image origin identification problem can be formulated into two frameworks: *supervised classification* [17–19] and *hypothesis testing* [20]. Compared with hypothesis testing framework, supervised classification framework is utilized by most of existing methods in the literature. The construction of a classifier typically consists of two stages: training stage and testing stage. It is assumed that the entire image space $\mathcal{Z}^N$ that includes all images from all the sources in the real world can be

¹The term *source* means an individual camera instance, a camera model, or a camera brand. Other sources such as cell-phone cameras, scanners, computer graphic are not addressed in this thesis.

divided into disjoint subsets in which images with same characteristics from the same source are grouped together. Let $\{\mathcal{S}_1, \mathcal{S}_2, \dots, \mathcal{S}_{N_s}\}$ be N_s different sources that are required to be classified. Typically, each source $\mathcal{S}_n$, $1 \leq n \leq N_s$, is a subset of $\mathcal{Z}^N$. In the training stage, suppose that N_{im} images are collected to be representative for each source. Each image in the source $\mathcal{S}_n$ is denoted by $\mathbf{Z}_{n,i}$. Then a feature vector is extracted from each image. Formally, a feature vector is a mapping $f: \mathcal{Z}^N \rightarrow \mathcal{F}$ where each image $\mathbf{Z}$ is mapped to a N_f -dimensional vector $\mathbf{v} = f(\mathbf{Z})$. Here, $\mathcal{F}$ is called feature space and N_f is the number of selected features, which is also the dimension of the feature space $\mathcal{F}$. The number of features N_f is very small compared with the number of pixels N . Working in a low-dimensional feature space $\mathcal{F}$ that represent the input images is much simpler than working on high-dimensional noisy image space $\mathcal{Z}^N$. The choice of an appropriate feature vector is primordial in supervised classification framework since the accuracy of a classifier highly depends on it. Thus we obtain a set of feature vectors $\{\mathbf{v}_{n,i}\}$ that is representative for each source. In this training stage feature refinement can be also performed such as dimensionality reduction or feature selection to avoid overtraining and redundant features. The knowledge learnt from the set of refined feature vectors helps build a classifier using supervised machine learning algorithms. A classifier typically is a learned function that can map an input feature vector to a corresponding source. Therefore, in the testing stage, the same steps such as feature selection and feature refinement are performed on the testing images. The output of the trained classifier is the result predicted for input testing images. Among many existing powerful machine learning algorithms, Support Vector Machines (SVM) [18, 19] seem to be the most popular choice in passive forensics. SVM are established in a solid mathematical foundation, namely statistical learning theory [18]. Moreover, their implementation is available for download and is easy to use [21].

Supervised classification framework involves two main drawbacks. To achieve high accuracy, supervised classification framework requires an expensive training stage that involves many images with different characteristics (e.g. image content or camera settings) from various sources for representing a real-world situation, which might be unrealistic in practice. Another drawback of this framework is that the trained classifier can not establish the statistical performances analytically since it does not rely on knowledge of *a priori* statistical distribution of images. In the operational context, such as for law enforcement and intelligent agencies, the design of an efficient method might not be sufficient. Forensic analysts also require that the probability of false alarm should be guaranteed and below a prescribed rate. The analytic establishment of statistical performances still remains an open problem in machine learning framework [22].

On the other hand, the problem of image origin identification problem lends itself to a binary hypothesis testing formulation.

Definition 2.1. (*Origin identification problem*). Given an arbitrary image $\mathbf{Z}$ under investigation, to identify the source of the image $\mathbf{Z}$, forensic analysts decide between

two following hypotheses

$$\begin{cases} \mathcal{H}_0 : \mathbf{Z} \text{ is acquired by the source of interest } \mathcal{S}_0 \\ \mathcal{H}_1 : \mathbf{Z} \text{ is acquired by a certain source } \mathcal{S}_1 \text{ that differs from the source } \mathcal{S}_0. \end{cases} \quad (2.17)$$

Suppose that the source $\mathcal{S}_0$ is available, so forensic analysts can have access to its characteristics, or its fingerprints. Therefore, they can make a decision by checking whether the image in question $\mathbf{Z}$ contains the fingerprints of the source. Relying on *a priori* statistical distribution of the image $\mathbf{Z}$ under each source, forensic analysts can establish a test statistic that can give a decision rule according to some criteria of optimality.

Statistical hypothesis testing theory has been considerably studied and applied in many fields. Several statistical tests as well as criteria of optimality have been proposed. While supervised learning framework only requires to find an appropriate set of forensic features, the most challenging part in hypothesis testing framework is to establish a statistical distribution to accurately characterize a high-dimensional real image. In doing so, hypothesis testing framework allows us to establish analytically the performance of the detector and warrant a prescribed false alarm probability, which are two crucial criteria in the operational context that supervised classification framework can not enable. However, hypothesis testing framework is of limited exploitation in forensic analysis. For the sake of clarity, hypothesis testing theory will be more detailed in Chapter A.

There are many passive forensic methods proposed in the literature for image origin identification. In this thesis, we limit the scope of our review to methods for identification of the source of a digital camera (e.g. camera brand, camera model, or individual camera instance). The methods to identify other imaging mechanisms such as cell-phone cameras, scanners, and computer graphics will not be addressed. It is important to distinguish the problem of camera instance identification and the problem of camera model/brand identification. More specifically, fingerprints used for camera instance identification should capture individuality, especially cameras coming from the same brand and model. For camera model/brand identification, it is necessary to exploit fingerprints that are shared between cameras of the same model/brand but discriminative for different camera models/brands.

Existing methods in the literature can be broadly divided into two categories. Methods in the first category exploit differences in image processing techniques and component technologies among camera models and manufacturers such as lens aberration [23], CFA patterns and interpolation [24–26] and JPEG compression [27, 28]. The main challenge in this category is that the image processing techniques remain identical or similar, and the components produced by a few manufacturers are shared among camera models. Methods in the second category aim to identify unique characteristics or fingerprints of the acquisition device such as PRNU [29–36]. The ability to reliably extract this fingerprint from an image is the main challenge in the second category since different image contents and non-linear operations may

severely affect this extraction. Below we will present the methods according to the position of exploited fingerprints in the image acquisition pipeline of a digital camera.

2.3.1 Lens Aberration

Digital cameras use lenses to capture incident light. Due to the imperfection of the design and manufacturing process, lenses cause undesired effects in output images such as spherical aberration, chromatic aberration, or radial distortion. Spherical aberration occurs when all incident light rays end up focusing at different points after passing through a spherical surface, especially light rays passing through the periphery of the spherical lens. Chromatic aberration is a failure of lens to converge different wavelengths at the same position on the image sensor. Radial distortion causes straight lines rendered as curved lines on the image sensor and it occurs when the transverse magnification (ratio of the image distance to the object distance) is not a constant but a function of the off-axis image distance. Among these effects, radial distortion may be the most severe part that lens produces in output images. Different manufacturers design different lens systems to compensate the effect of radial distortion. Moreover, focal length also affects the degree of radial distortion. As a result, each camera brand or model may leave an unique degree of radial distortion on the output images. Therefore, radial distortion of lens is exploited in [23] to identify the source of the image.

The authors in [23] take the center of an image as the center of distortion and model the undistorted radius r_u as a non-linear function of distorted radius r_d

$$r_u = r_d + k_1 r_d^3 + k_2 r_d^5. \quad (2.18)$$

Distortion parameters (k_1, k_2) are estimated using the straight line method [37, 38]. Then the distortion parameters (k_1, k_2) are exploited as forensic features to train a SVM classifier. Although experiments provided a promising result, experiments were only conducted on three different camera brands. Experiments on large database including different devices per camera model and different camera models are more desirable. However, this lens aberration-based classifier would fail for a camera with possibly interchangeable lenses, e.g. Digital Single Lens Reflex (DSLR) camera.

2.3.2 Sensor Imperfections

As discussed in Section 2.2, imperfections during the manufacturing process and inhomogeneity of silicon wafers leads to slight variations in the response of each pixel to incident light. These slight variations are referred as to PRNU, which is unique for each sensor pattern. Thus PRNU can be exploited to trace down to individual camera instance. The FPN was also used in [39] for camera instance identification. However, the FPN can be easily compensated, thus it is not a robust fingerprint and no longer used in later works.

Generally, PRNU is modeled as a multiplicative noise-like signal [30, 32]

$$\mathbf{Z} = \boldsymbol{\mu} + \boldsymbol{\mu}\mathbf{K} + \boldsymbol{\Xi}, \quad (2.19)$$

where $\mathbf{Z}$ is the output noisy image, $\boldsymbol{\mu}$ is the ideal image in the absence of noise, $\mathbf{K}$ represents the PRNU, and $\boldsymbol{\Xi}$ accounts for the combination of other noise sources. All operations in (2.19) are pixel-wise.

Like supervised classification, the PRNU-based method also consists two stages. The training stage involves collecting N_{im} images that are acquired by the camera of interest $\mathcal{S}_0$ and extracting the PRNU $\mathbf{K}_0$ characterizing this camera. This is accomplished by applying a denoising filter on each image to suppress the image content, then performing Maximum Likelihood (ML) estimation of $\mathbf{K}_0$

$$\mathbf{K}_0 = \frac{\sum_{i=1}^{N_{\text{im}}} \mathbf{Z}_i^{\text{res}} \mathbf{Z}_i}{\sum_{i=1}^{N_{\text{im}}} (\mathbf{Z}_i)^2}, \quad (2.20)$$

where $\mathbf{Z}_i^{\text{res}} = \mathbf{Z}_i - \mathcal{D}(\mathbf{Z}_i)$ is the noise residual corresponding to the image $\mathbf{Z}_i$, $1 \leq i \leq N_{\text{im}}$, and $\mathcal{D}$ stands for the denoising filter. Note that when the PRNU is extracted from JPEG images, it may contain artifacts introduced by CFA interpolation and JPEG compression. These artifacts are not unique to each camera instance and shared among different camera units of the same model. To render the PRNU unique to the camera and improve the accuracy of the method, a preprocessing step is performed to suppress these artifacts by subtracting the averages from each row and column, and applying the Wiener filter in the Fourier domain [30]. In the testing stage, given an image under investigation $\mathbf{Z}$, the problem of camera source identification (2.17) is rewritten as follows

$$\begin{cases} \mathcal{H}_0 : & \mathbf{Z}^{\text{res}} = \boldsymbol{\mu}\mathbf{K}_0 + \tilde{\boldsymbol{\Xi}} \\ \mathcal{H}_1 : & \mathbf{Z}^{\text{res}} = \boldsymbol{\mu}\mathbf{K}_1 + \tilde{\boldsymbol{\Xi}}, \end{cases} \quad (2.21)$$

where the noise term $\tilde{\boldsymbol{\Xi}}$ includes the noise $\boldsymbol{\Xi}$ and additional terms introduced by the denoising filter. This formulation must be understood as follows: hypothesis $\mathcal{H}_0$ means that the noise residual $\mathbf{Z}^{\text{res}}$ contains the PRNU $\mathbf{K}_0$ characterizing the camera of interest $\mathcal{S}_0$ while hypothesis $\mathcal{H}_1$ means the opposite. It should be noted that the PRNU detection problem in [30, 32] is formulated in the reverse direction. The sub-optimal detector for the problem (2.21) is the normalized cross correlation between the PRNU term $\boldsymbol{\mu}\mathbf{K}$ and the noise residual $\mathbf{Z}^{\text{res}}$ [30]. In fact, the normalized cross correlation is derived from the Generalized Likelihood Ratio Test (GLRT) by modeling the noise term $\tilde{\boldsymbol{\Xi}}$ as white noise with known variance [40]. A more stable statistic derived in [32] is the Peak to Correlation Energy (PCE) as it is independent of the image size and has other advantages such as its response to the presence of weak periodic signals. Theoretically, the decision threshold for the problem (2.21) is given by $\tau = (\Phi^{-1}(1 - \alpha_0))^2$ where α_0 is the prescribed false alarm probability, $\Phi(\cdot)$ and $\Phi^{-1}(\cdot)$ denotes respectively the cumulative distribution function (cdf) of the standard Gaussian random variable and its inverse. If the

PCE is smaller than a threshold τ , the image $\mathbf{Z}$ is claimed taken by the camera in question. The detection performance can be improved by selecting an appropriate denoising filter [33], attenuating scene details in the test image [34,35], or recognizing the PRNU term with respect to each sub-sample of the CFA pattern [36].

Beside individual camera instance identification, the fingerprint PRNU can be also used for camera model identification [31]. This is based on an assumption that the fingerprint obtained from TIFF or JPEG images contains traces of post-acquisition processes (e.g. CFA interpolation) that carry information about the camera model. In this case, the above preprocessing step that removes the linear pattern from the PRNU will not be performed. The features extracted from the PRNU term including statistical moments, cross correlation, block covariance, and linear pattern, are used to train a SVM classifier.

2.3.3 CFA Pattern and Interpolation

Based on the assumption that different CFA patterns and CFA interpolation algorithms are employed by different manufacturers, even in different camera models, thus they can be used to discriminate camera brands and camera models. Typically, both CFA pattern and interpolation coefficients are unknown in advance. They must be estimated together from a single image. An algorithm has been developed in [24] to jointly estimate CFA pattern and interpolation coefficients, which has shown the robustness to JPEG compression with low quality factors. Firstly, a search space including 36 possible CFA patterns is established based on the observation that most cameras use a RGB type of CFA with a fixed periodicity of 2×2 . Since a camera may employ different interpolation algorithms for different types of regions, it is desirable to classify the given image into three types of regions based on gradient information in a local neighborhood of a pixel: region contains parts of the image with a significant horizontal gradient, region contains parts of the image with a significant vertical gradient, and region includes the remaining smooth parts of the image.

For every CFA pattern $\mathbf{P}_{\text{CFA}}$ in the search space, the interpolation coefficients are computed separately in each region by fitting linear models. Using the final output image $\mathbf{Z}$ and the assumed CFA pattern $\mathbf{P}_{\text{CFA}}$, we can identify the set of pixels that acquired directly from the image sensor and those obtained by interpolation. The interpolated pixels are assumed to be a weighted average of the pixels acquired directly. The interpolation coefficients are then obtained by solving these equations. To overcome the difficulty of noisy pixel values and interference of non-linear post-acquisition processes, singular value decomposition is employed to estimate the interpolation coefficients. These coefficients are then use to re-estimate the output image $\hat{\mathbf{Z}}$, and find the interpolation error $\hat{\mathbf{Z}} - \mathbf{Z}$. The CFA pattern that gives the lowest interpolation error and its corresponding coefficients are chosen as final results [24].

As soon as the interpolation coefficients are estimated from the given image, they are used as forensic features to train a SVM classifier for classification of cam-

era brands and models [24]. The detection performance can be further enhanced by taking into account intra-channel and inter-channel correlations and more sophisticated interpolation algorithms in the estimation methodology [26]. Other features can be used together with interpolation coefficients such as the peak location and magnitudes of the frequency spectrum of the probability map [41].

2.3.4 Image Compression

Image compression is the final step in the image processing pipeline. As discussed in Section 2.2, manufacturers have their own compression scheme for optimal trade-off of image quality versus file size. Different component technologies (e.g. lenses, sensors), different in-camera processing operations (e.g. CFA interpolation, white-balancing), together with different quantization matrices will jointly result in statistical difference of quantized DCT coefficient. Capturing this statistical difference and extracting useful features from it may enable to discriminate different camera brands or camera models.

To this end, instead of extracting statistical features directly from quantized DCT coefficients, features are extracted from the difference JPEG 2-D array [28]. The JPEG 2-D array consists of the magnitudes (i.e. absolute values) of quantized DCT coefficients. Three reasons behind taking absolute values are the followings:

1. The magnitudes of DCT coefficients decrease along the zig-zag order.
2. Taking absolute values can reduce the dynamic range of the resulting array.
3. The signs of DCT coefficients mainly carry information of the outlines and edges of the original spatial-domain image, which does not involve information about camera models. Thus by taking absolute values, all the information regarding camera models remains.

Then to reduce the influence of image content and enhance statistical difference introduced in image processing pipeline, the difference JPEG 2-D array, which is defined by taking the difference between an element and one of its neighbors in the JPEG 2-D array, is introduced. The difference can be calculated along four directions: horizontal, vertical, main diagonal, and minor diagonal. To model the statistical difference of quantized DCT coefficients and take into account the correlation between coefficients, the Markovian transition probability matrix is exploited. Each difference JPEG 2-D array from a direction generates its own transition probability matrix. Each probability value in the transition matrix is given by

$$\mathbb{P}\left[X(u_h + 1, u_v) = k \mid X(u_h, u_v) = l\right] = \frac{\sum_{u_h=1}^{N_h} \sum_{u_v=1}^{N_v} \mathbf{1}_{X(u_h, u_v)=l, X(u_h+1, u_v)=k}}{\sum_{u_h=1}^{N_h} \sum_{u_v=1}^{N_v} \mathbf{1}_{X(u_h, u_v)=l}}, \quad (2.22)$$

where $X(u_h, u_v)$ denotes an element in the difference JPEG 2-D array and $\mathbf{1}_E$ is an indicator function

$$\mathbf{1}_E = \begin{cases} 1 & \text{if } E \text{ is true} \\ 0 & \text{otherwise.} \end{cases} \quad (2.23)$$

These steps are performed for the Y and Cb components of the compressed JPEG image. Totally, we can collect 324 transition probabilities for Y component and 162 transition probabilities for Cb component. The transition probabilities are then used as forensic features for SVM classification. Experiments are then conducted on a large database including 40000 images of 8 different camera models, providing a good classification performance [28]. In this method it is more desirable to perform feature refinement to reduce the number of features and the complexity of the algorithm.

2.4 Passive Image Forgery Detection

Image forgery detection is another fundamental task of forensic analysts, which aims to detect any act of manipulation on image content. The main assumption is that even though a forger with skills and powerful tools does not leave any perceptible trace of manipulation, the manipulation creates itself inconsistencies in image content. Depending on which type of inconsistencies is investigated and how passive forensic methods operate, they can broadly be divided into five categories. A single method can hardly detect all types of forgery, so forensic analysts should use these methods together to reliably detect a wide variety of tampering.

1. *Universal Classifiers*: Any act of manipulation may lead to statistical changes in the underlying image. Instead of capturing these changes directly in a high-dimensional and non-stationary image, which is extremely difficult, one approach is to detect changes in a set of features that represent an image. Based on these features, supervised classification is employed to provide universal classifiers to discriminate between unaltered images and manipulated images. Some typical forensic features are higher-order wavelet statistics [42], image quality and binary similarity measures [43, 44]. These universal classifiers are not only able to detect some basic manipulations such as resizing, splicing, contrast enhancement, but also reveal the existence of hidden messages [45].
2. *Camera Fingerprints-Based*: A typical scenario of forgery is to cut a portion of an image and paste it into a different image, then create the so-called forged image. The forged region may not be taken by the same camera as remaining regions of the image, which results in inconsistencies in camera fingerprints between those regions. Therefore, if these inconsistencies exist in an image, we could assume that the image is not authentic. For authentication, existing methods have exploited many camera fingerprints such as chromatic aberration [46], PRNU [30, 32], CFA interpolation and correlation [25, 47–49], gamma correction [50, 51], Camera Response Function (CRF) [52–54].
3. *Compression and Coding Fingerprints-Based*: Nowadays most commercial cameras export images in JPEG format for ease of storage and transmission. As discussed in Section 2.2, JPEG compression introduces two important artifacts: clustering of DCT coefficients around integer multiples of the quan-

tization step, and blocking artifacts. Checking inconsistencies in these two artifacts can trace the processing history of an image and determine its origin and authenticity. A possible scenario is that while the original image is saved in JPEG format, a forger could save it in a lossless format after manipulation. Existence of these artifacts in an image in a lossless format can show that it has been previously compressed [55–57]. Another scenario is that the forger could save the manipulated image in JPEG format, which means that the image has undergone JPEG compression twice. Detection of double JPEG compression can be performed by checking periodic patterns (e.g. double peaks and missing centroids) in the histogram of DCT coefficients due to different quantization steps [51, 58, 59], which are not present in singly compressed images, or using the distribution of the first digit of DCT coefficients [60, 61]. The detection of double JPEG compression is of greater interest since it can reveal splicing or cut-and-paste forgeries due to the fact that the forged region and remaining regions of the image may not have the same processing history. Inconsistencies can be identified either in DCT domain [62–65] or in spatial domain via blocking artifacts [66, 67]. Furthermore, the detection of double JPEG compression can be applied for detecting hidden messages [58, 59].

4. *Manipulation-Specific Fingerprints-Based*: Each manipulation may leave specific fingerprints itself within an image, which can be used as evidence of tampering. For example, resampling causes specific periodic correlations between neighboring pixels. These correlations can be estimated based on the Expectation Maximization (EM) algorithm [68], and then used to detect the resampling [68, 69]. Furthermore, resampling can be also detected by identifying periodicities in the average of an image's second derivative along its row and columns [70], or periodicities in the variance of an image's derivative [71]. Contrast enhancement creates impulsive peaks and gaps in the histogram of the image's pixel value. These fingerprints can be detected by measuring the amount of high frequency energy introduced into the Fourier transform of an image's pixel value histogram [72]. Median filtering introduces streaking into the signals [73]. Streaks correspond to a sequence of adjacent signal observations all taking the same value. Therefore, median filtering can be detected by analyzing statistical properties of the first difference of an image's pixel values [74–77]. Splicing disrupts higher-order Fourier statistics, which leaves traces to detect splicing [78].
5. *Physical Inconsistencies-Based*: Methods in this category do not make use of any form of fingerprints but exploit properties of lighting environment for forgery detection. The main assumption is that all the objects within an image are typically illuminated under the same light sources, so the same properties of lighting environments. Therefore, difference in lighting across an image can be used as evidence of tampering, e.g. splicing. To this end, it is necessary to estimate the direction of the light source illuminating an object. This can be accomplished by considering two-dimensional [79] or three-dimensional [80]

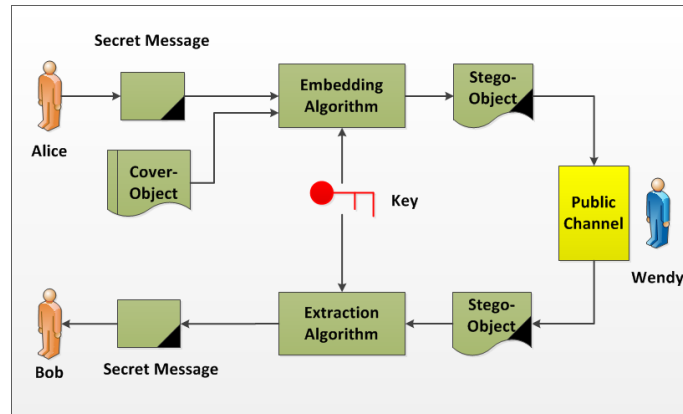


Figure 2.4: Typical steganographic system.

surface normals, and illumination under a single light source [79] or even under multiple light sources [81]. The lighting environment coefficients of all objects in an image are then used for checking inconsistencies.

2.5 Steganography and Steganalysis in Digital Images

Steganography is the art and science of hiding communication. The concept of steganography is used for invisible communication between only two parties, the sender and the receiver, such that the message exchanged between them can not be detected by an adversary. This communication can be illustrated by prisoners' problem [82]. Two prisoners, Alice and Bob, want to develop an escape plan but all communications between them are unfortunately monitored by a warden named Wendy. The escape plan must be kept secret and exchanged without raising Wendy's suspicion. It means that the communication does not only involve the confidentiality of the escape plan but also its undetectability. For this purpose, a practical way is to hide the the escape plan, or the secret message in a certain ordinary object and send it to the intended receiver. By terminology, the original object that is used for message hiding is called *cover-object* and the object that contains the hidden message is called *stego-object*. The hiding technique does not destroy the object content perceptibly to not raise Wendy's suspicion, nor modify the message content so the receiver could totally understand the message.

The advances in information technologies make digital media (e.g. audio, image, or video) ubiquitous. This ubiquity facilitates the choice of a harmless object in which the sender can hide a secret message, so sending such media is inconspicuous. Furthermore, the size of digital media is typically large compared to the size of secret message. Thus the secret message can be easily hidden in digital media without visually destroying digital content. Most of researches focus on digital images, which are also the type of media addressed in this thesis.

A typical steganographic system is shown in Figure 2.4. It consists of two stages:

embedding stage and extraction stage. When Alice wants to send a secret message $\mathbf{M}$, she hides it into a cover-image $\mathbf{C}$ using a key and an embedding algorithm. The secret message $\mathbf{M}$ is a binary sequence of L bits, $\mathbf{M} = (m_1, m_2, \dots, m_L)^T$ with $m_i \in \{0, 1\}$, $1 \leq i \leq L$. The resulting stego-image $\mathbf{S}$ is then transmitted to Bob via an insecure channel. Bob can retrieve the message $\mathbf{M}$ since he knows the embedding algorithm used by Alice and has access to the key used in embedding process. Bob does not absolutely require the original cover-image $\mathbf{C}$ for message extraction. From the Kerckhoffs' principle [83], it is assumed that in digital steganography, steganographic algorithms are public so that all parties including the warden Wendy have access to them. The security of the steganographic system relies solely on the key. The key could be secret key exchanged between Alice and Bob through a secure channel, or public key.

In general, steganographic systems can be evaluated by three basic criteria: capacity, security, and robustness. Capacity is defined as the maximum length of a secret message. The capacity depends on the embedding algorithm and properties of cover-images. The security of a steganographic system is evaluated by the undetectability rather than the difficulty of reading the message content in case of cryptographic system. However, we can see that steganographic systems also exploit the idea of exchange of keys (secret and public) from cryptographic system to reinforce the security. Robustness means the difficulty of removing a hidden message from a stego-image, so the secret message survives some accidental channel distortions or systematic interference of the warden that aims to prevent the use of steganography. It can be noted that longer messages will lead to more changes in the cover image, thus less security. In brief, these three criteria are mutually dependent and are balanced when designing a steganographic system.

The purpose of steganography is to secretly communicate through a public channel. However, this concept has been misused by anti-social elements, criminals, or terrorists. It could lead to important consequences to homeland security or national defence when, for example, two terrorists exchange a terrorist plan. Therefore, it is urgent for law enforcement and intelligence agencies to build up a methodology in order to detect the mere existence of a secret message and break the security of steganographic systems. Embedding a secret message into a cover-image is also an act of manipulating image content, so steganalysis is one of important tasks of forensic analysts, or steganalysts in this case. Unlike in cryptanalysis, the steganalyst Wendy does not require to retrieve the actual message content. As soon as she have detected its existence in an image, she can cut off the communication channel by putting two prisoners in separate cells. This is the failure of steganography. Besides, the task of steganalysis must be accomplished blindly without knowledge of original cover image.

Generally, the steganalyst Wendy can play either *active* or *passive* role. While the active steganalyst is allowed to modify exchanged objects through the public channel in order to prevent the use of steganography, the passive steganalyst is not. The only goal of passive steganalyst is to detect the presence of a hidden message in a given image, which is also the typical scenario on which most of researches mainly

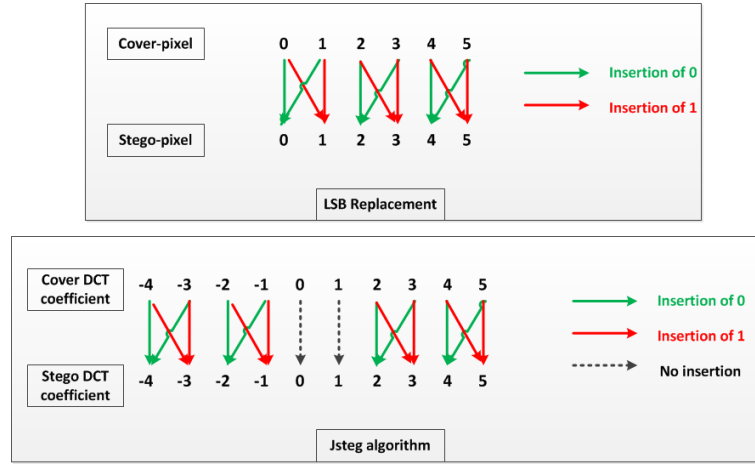


Figure 2.5: Operations of LSB replacement (top) and Jsteg (bottom).

focus. It can be noted that the steganalysis is like the coin-tossing game since the decision of steganalysts is made by telling that the given image is either a cover-image or a stego-image. Hence in any case, steganalysts can get a correct detection probability of 50%. However, steganalysts should establish the problem of hidden message detection in a more formal manner and design a powerful steganalysis tool with higher correct detection probability, rather than a random guess. Apart from detecting the presence of a hidden message, it may be desirable for steganalysts to estimate the message length or brute-force the secret key and retrieve the message content. The estimation of the message length is called *quantitative steganalysis*. Brute-forcing the secret key and extraction of the message content are referred to as *forensic steganalysis*.

As stated above, designing a steganographic system is a trade-off between three basic criteria. Thus many steganographic algorithms have been proposed for different purposes such as mimic natural processing [84–86], preserve a model of cover-images [87, 88], or minimize the distortion function [89, 90]. Among available algorithms, Least Significant Bit (LSB) replacement might be the oldest embedding technique in digital steganography. This algorithm is simple and easy to implement, thus it is available in numerous low-cost steganographic softwares on the Internet despite its relative insecurity. In addition, LSB replacement inspires a majority of other steganographic algorithms (e.g. LSB matching [91], Jsteg [92]). Jsteg algorithm is simply the implementation of LSB replacement in the DCT domain. Therefore, understanding LSB replacement paradigm is a good starting point before addressing more complex embedding paradigms. In this thesis, we only review LSB replacement and Jsteg algorithm, and their powerful steganalysis detectors proposed in the literature. The readers can be referred to [93–95] for other steganographic and steganalysis methods.

2.5.1 LSB Replacement Paradigm and Jsteg Algorithm

Considering the cover-image $\mathbf{C}$ as a column vector, the LSB replacement technique involves choosing a subset of L cover-pixels $\{c_1, c_2, \dots, c_L\}$, and replacing the LSB of each cover pixel by a message bit. The LSB of a cover-pixel c_i is defined as follows

$$\text{LSB}(c_i) = c_i - 2 \left\lfloor \frac{c_i}{2} \right\rfloor, \quad (2.24)$$

where $\lfloor \cdot \rfloor$ is the floor function. The LSB of the cover-pixel c_i takes values in $\{0, 1\}$. Therefore, by embedding a message bit m_i into the cover-pixel c_i , the stego-pixel s_i is given by

$$s_i = 2 \left\lfloor \frac{c_i}{2} \right\rfloor + m_i. \quad (2.25)$$

We see that when $\text{LSB}(c_i) = m_i$, the pixel value does not change after embedding, $s_i = c_i$. By contrast, when $\text{LSB}(c_i) \neq m_i$, the stego-pixel s_i can be defined as a function of the cover-pixel c_i in the following manner

$$s_i = c_i + 1 - 2 \cdot \text{LSB}(c_i) = c_i + (-1)^{c_i} \triangleq \bar{c}_i, \quad (2.26)$$

where $\bar{c}_i$ is the pixel with flipped LSB. In other words, even values are never decremented whereas odd values are never incremented. The absolute difference between a cover-pixel c_i and a stego-pixel s_i is smaller than 1, $|c_i - s_i| \leq 1$, thus the artifact caused by the insertion of secret message $\mathbf{M}$ could be imperceptible under human vision. The operation of LSB replacement technique is illustrated in Figure 2.5. One problem that remains to be solved is the choice of the subset of cover-pixels or the sequence of pixel indices used in embedding process. To increase the complexity of the algorithm, the sender could create a pseudorandom path generated from the secret key shared between the sender and the receiver so that the secret message bits are spread randomly over the cover-image. Therefore, the distance between two embedded bits is also determined pseudorandomly, which would not raise the suspicion of the warden. We can see that the number of message bits that can be embedded does not exceed beyond the number of pixels of the image $\mathbf{Z}$: $L \leq N$, which leads us to define an embedding rate R

$$R = \frac{L}{N}. \quad (2.27)$$

This embedding rate R is a measure of the capacity of the steganographic system based on LSB replacement technique.

Jsteg algorithm is a variant of LSB replacement technique in spatial domain. Jsteg algorithm embeds the secret message into the DCT domain by replacing LSBs of quantized DCT coefficients by message bits. The difference from the LSB replacement technique in spatial domain is that Jsteg algorithm does not embed message bits in the coefficients that are equal to 0 and 1 since artifacts caused by such embedding can be perceptibly and easily detected. The DC coefficient is not used as well for the same reason. The AC coefficients that differ from 0 et 1 are usable

coefficients. Consequently, the embedding rate R in Jsteg algorithm is defined as the ratio of the length L and the number of usable coefficients in the cover-image $\mathbf{C}$

$$R = \frac{L}{\sum_{k=2}^{64} n_k}. \quad (2.28)$$

where n_k is the number of usable coefficients at the frequency k , $2 \leq k \leq 64$.

2.5.2 Steganalysis of LSB Replacement in Spatial Domain

Like the origin identification problem (2.17), the steganalysis problem can be also formulated as a binary hypothesis testing.

Definition 2.2. (*Steganalysis problem*). Given a suspect image $\mathbf{Z}$, to verify whether the image $\mathbf{Z}$ contains a secret message or not, the steganalyst decides between two following hypotheses

$$\begin{cases} \mathcal{H}_0 : & \mathbf{Z} = \mathbf{C}, \text{no hidden message.} \\ \mathcal{H}_1 : & \mathbf{Z} = \mathbf{S}, \text{with hidden message.} \end{cases} \quad (2.29)$$

To solve the steganalysis problem (2.29), several methods have been proposed in the literature. Even though the secret message is imperceptible to human eye, the act of embedding a secret message modifies the cover content and leaves itself artifacts that can be detected. Steganalysis methods of LSB replacement can be roughly divided into four categories: structural detectors, Weighted Stego-image (WS) detectors, statistical detectors, and universal classifiers. Typically, structural detectors and WS detectors are quantitative detectors that provide an estimation of secret message length while statistical detectors and universal classifiers attempt to separate stego-images from cover-images based on changes in statistical properties of cover-images due to message embedding. Below we briefly discuss each category of detectors.

2.5.2.1 Structural Detectors

Structural detectors exploit all combinatorial measures of the artificial dependence between sample differences and the parity structure of the LSB replacement in order to estimate the secret message length. Some representatives in this category are the Regular-Singular (RS) analysis [96], the Sample Pair Analysis (SPA) [97–99], and the Triple/Quadruple analysis [100, 101]. The common framework is to model effects of LSB replacement as a function of embedding rate R , invert these effects to approximate cover-image properties from the stego-image, and find the best candidate $\hat{R}$ to match cover assumptions.

Both RS and SPA methods rely on evaluating groups of spatially adjacent pixels. The observations made in RS analysis were formally justified in SPA. For pedagogical reasons, we discuss the SPA method. For the representation of the SPA method, we

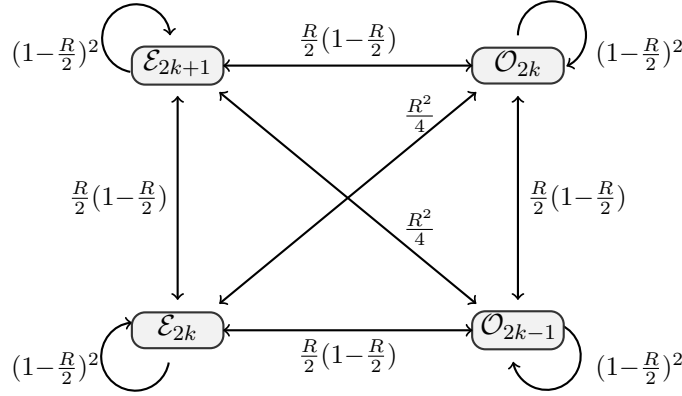


Figure 2.6: Diagram of transition probabilities between trace subsets under LSB replacement.

use the extensible alternative notations in [95, 101]. Given an image $\mathbf{Z}$, we define a trace set $\mathcal{C}_k$ that collect all pairs of adjacent pixels (z_{2i}, z_{2i+1}) as follows

$$\mathcal{C}_k = \left\{ (2i, 2i+1) \in \mathcal{I}^2 \mid \left\lfloor \frac{z_{2i}}{2} \right\rfloor = \left\lfloor \frac{z_{2i+1}}{2} \right\rfloor + k \right\}, \quad (2.30)$$

where $\mathcal{I}$ is the set of pixel indices. Each trace set $\mathcal{C}_k$ is then partitioned into four trace subsets, $\mathcal{C}_k = \mathcal{E}_{2k} \cup \mathcal{E}_{2k+1} \cup \mathcal{O}_{2k} \cup \mathcal{O}_{2k-1}$, where $\mathcal{E}_k$ and $\mathcal{O}_k$ are defined by

$$\begin{cases} \mathcal{E}_k &= \left\{ (2i, 2i+1) \in \mathcal{I}^2 \mid z_{2i} = z_{2i+1} + k, z_{2i+1} \text{ is even} \right\} \\ \mathcal{O}_k &= \left\{ (2i, 2i+1) \in \mathcal{I}^2 \mid z_{2i} = z_{2i+1} + k, z_{2i+1} \text{ is odd} \right\}. \end{cases} \quad (2.31)$$

We can observe that the LSB replacement technique never changes the trace set $\mathcal{C}_k$ of a sample pair but can move sample pairs between trace subsets. Therefore, we establish transition probabilities as functions of the embedding rate R , which is shown in Figure 2.6. Thus we can derive the relation between trace subsets of a stego-image and those of a cover-image

$$\begin{pmatrix} |\mathcal{E}_{2k+1}^s| \\ |\mathcal{E}_{2k}^s| \\ |\mathcal{O}_{2k}^s| \\ |\mathcal{O}_{2k-1}^s| \end{pmatrix} = \begin{pmatrix} (1 - \frac{R}{2})^2 & \frac{R}{2}(1 - \frac{R}{2}) & \frac{R}{2}(1 - \frac{R}{2}) & \frac{R^2}{4} \\ \frac{R}{2}(1 - \frac{R}{2}) & (1 - \frac{R}{2})^2 & \frac{R^2}{4} & \frac{R}{2}(1 - \frac{R}{2}) \\ \frac{R}{2}(1 - \frac{R}{2}) & \frac{R^2}{4} & (1 - \frac{R}{2})^2 & \frac{R}{2}(1 - \frac{R}{2}) \\ \frac{R^2}{4} & \frac{R}{2}(1 - \frac{R}{2}) & \frac{R}{2}(1 - \frac{R}{2}) & (1 - \frac{R}{2})^2 \end{pmatrix} \begin{pmatrix} |\mathcal{E}_{2k+1}^c| \\ |\mathcal{E}_{2k}^c| \\ |\mathcal{O}_{2k}^c| \\ |\mathcal{O}_{2k-1}^c| \end{pmatrix}, \quad (2.32)$$

where $\mathcal{E}_k^c$ and $\mathcal{O}_k^c$ are trace subsets of the cover-image, and $\mathcal{E}_k^s$ and $\mathcal{O}_k^s$ are trace subsets of the stego-image. Here $|\mathcal{S}|$ denotes the cardinality of the set $\mathcal{S}$. After inverting the transition matrix and assuming that $|\mathcal{E}_{2k+1}^c| = |\mathcal{O}_{2k+1}^c|$, we obtain a quadratic equation

$$\begin{aligned} 0 &= R^2 \left(|\mathcal{C}_k| - |\mathcal{C}_{k+1}| \right) + 4 \left(|\mathcal{E}_{2k+1}^s| - |\mathcal{O}_{2k+1}^s| \right) \\ &\quad + 2R \left(|\mathcal{E}_{2k+2}^s| + |\mathcal{O}_{2k+2}^s| - 2|\mathcal{E}_{2k+1}^s| + 2|\mathcal{O}_{2k+1}^s| - |\mathcal{E}_{2k}^s| - |\mathcal{O}_{2k}^s| \right). \end{aligned} \quad (2.33)$$

The solution of Equation (2.33) is an estimator of the embedding rate R . The SPA method was further improved by combining with Least Squares (LS) method [98] and Maximum Likelihood [99], or generalizing from analysis of pairs to analysis of k -tuples [100, 101].

2.5.2.2 WS Detectors

WS detectors were originally proposed by J. Fridrich in [102] and then improved in [103, 104]. The key idea of WS is that the embedding rate can be estimated via the weight that minimizes the distance between the weighted stego image and the cover image [95]. The weighted stego-image with scalar parameter λ of the image $\mathbf{Z}$ is defined by

$$\forall i \in \mathcal{I}, \quad z_i^{(\lambda)} = (1 - \lambda)z_i + \lambda \bar{z}_i, \quad \text{with} \quad \bar{z}_i = z_i + (-1)^{z_i}. \quad (2.34)$$

The estimator $\hat{R}$ can be provided by minimizing the Euclidian distance between the weighed stego-image and the cover image

$$\hat{R} = 2 \arg \min_{\lambda} \sum_{i=1}^N w_i (z_i^{(\lambda)} - c_i)^2, \quad (2.35)$$

where the normalized weight vector $\mathbf{w}$ with $\sum_{i=1}^N w_i = 1$ is taken into account in the minimization problem (2.35) to reflect the heterogeneity in a natural image. By solving the root of the first derivative in (2.35), a simplified estimator is given as

$$\hat{R} = 2 \sum_{i=1}^N w_i (z_i - \bar{z}_i)(z_i - c_i). \quad (2.36)$$

Since the cover-pixels c_i are unknown in advance, a local estimator for each pixel from its spatial neighborhood can be employed, or more generally, a linear filter $\mathcal{D}$, to provide an estimate of cover-image: $\hat{\mathbf{C}} = \mathcal{D}(\mathbf{Z})$. The estimator $\hat{R}$ in (2.36) follows immediately. From above observations, the choices of an appropriate linear filter $\mathcal{D}$ and weight vector $\mathbf{w}$ are crucial for improvement of WS detectors' performance [95, 103, 104].

Both structural detectors and WS detectors are established into quantitative steganalysis framework, which means that instead of indicating a suspect image $\mathbf{Z}$ is either a cover-image or stego-image, the output of those detectors is a real-value estimate of the secret message length. In other words, even no secret message is embedded in the image, i.e. $R = 0$, we could still obtain a negative or positive value. Nevertheless, quantitative detectors offer an additional advantage over statistical detectors, namely that the detection performance can be measured by evaluating the deviation of the estimator $\hat{R}$ from the true embedding rate R . Some criteria can be used as measures of performance such as

- Mean Absolute Error (MAE):

$$\frac{1}{N_{\text{im}}} \sum_{n=1}^{N_{\text{im}}} |\hat{R}_n - R|, \quad (2.37)$$

where N_{im} is number of images.

- Median Absolute Error (mAE):

$$\text{median}_n |\hat{R}_n - R|. \quad (2.38)$$

2.5.2.3 Statistical Detectors

In contrast to structural detectors and WS detectors, statistical detectors rely on changes in statistical properties due to message embedding to detect the presence of the secret message. The output of statistical detectors is a binary decision. Some representatives are χ^2 detector [105] and Bayesian approach-based detector [106]. Another interesting approach is the one proposed in [107] that is based on the statistical hypothesis testing theory. To this end, two preliminary assumptions are given in the following proposition:

Proposition 2.1. *In the LSB replacement embedding technique, we assume that*

1. *The secret message bits are uniformly distributed over the cover-image, namely that the probability of embedding a message bit into every cover-pixel is identical. Moreover, message bits and cover pixels are statistically uncorrelated [107].*
2. *Secret message bits are independent and identically distributed (i.i.d), and each message bit m_i is drawn from the Binomial distribution $\mathcal{B}(1, \frac{1}{2})$*

$$\mathbb{P}[m_i = 0] = \mathbb{P}[m_i = 1] = \frac{1}{2}, \quad (2.39)$$

where $\mathbb{P}[E]$ denotes the probability that an event E occurs.

Therefore, from the mechanism of LSB replacement, we can see that the probability that the pixel does not change after embedding is $1 - \frac{R}{2}$ while the probability that its LSB is flipped is $\frac{R}{2}$

$$\mathbb{P}[s_i = c_i] = 1 - \frac{R}{2} \quad \text{and} \quad \mathbb{P}[s_i = \bar{c}_i] = \frac{R}{2}. \quad (2.40)$$

Let P_0 be the probability distribution of cover-images. Due to message embedding at rate R whose properties are given in Proposition 2.1, the cover image moves from the probability distribution P_0 to a different probability distribution, denoted P_R . Thus the steganalysis problem (2.29) can be rewritten as follows

$$\begin{cases} \mathcal{H}_0 : \mathbf{Z} \sim P_0 \\ \mathcal{H}_1 : \mathbf{Z} \sim P_R. \end{cases} \quad (2.41)$$

Based on the assumption that all pixels are independent and identically distributed, the authors in [107] have developed two schemes depending on the knowledge of the probability distribution P_0 . When the probability distribution P_0 is not known,

the authors study the asymptotically optimal detector (as the number of pixels $N \rightarrow \infty$) according to Hoeffding's test [108]. When the probability distribution P_0 is known in advance, an optimal detector is given in the sense of Neyman-Pearson [20, Theorem 3.2.1]. Although the statistical detector proposed in [107] is interesting from theoretical point of view, its performance in practice is quite moderate due to the fact that the cover model used in [107] is not sufficiently accurate to describe a natural image. The assumption of independence between pixels does not hold since the image structure and the non-stationarity of noises during image acquisition process are not taken into account.

Some later works [109–114] rely on a simplistic local polynomial model in which pixel's expectations are different in order to design a statistical detector, providing high detection performance compared with structural and WS ones. Far from assuming that all the pixels are i.i.d as in [107], those works propose to model each cover-pixel by the Gaussian distribution, $c_i \sim \mathcal{N}(\mu_i, \sigma_i^2)$, in order to design the Likelihood Ratio Test (LRT) in which the Likelihood Ratio (LR) Λ can be given by

$$\Lambda(\mathbf{Z}) \propto \sum_i \frac{1}{\sigma_i^2} (z_i - \bar{z}_i)(z_i - \mu_i). \quad (2.42)$$

The LRT is the most powerful test in the sense of Neyman-Pearson approach [20, Theorem 3.2.1] that can meet simultaneously two criteria of optimality: warranting a prescribed false alarm probability and maximizing the correct detection probability. Moreover, the specificity in this approach is to show that the WS detector [102–104] is indeed a variant of the LRT, which justifies the good detection performance of such *ad hoc* detector. Besides, hypothesis testing theory has been also extended to other complex embedding algorithm, e.g. LSB matching [115, 116].

2.5.2.4 Universal Classifiers

Three previous families of detectors are targeted to a specific steganographic algorithm, namely LSB replacement. In other words, these three families work on an assumption that steganalysts know in advance the embedding algorithm used by the steganographer. Such scenario may not be realistic in the practical context. Universal classifiers are employed by steganalysts to work in a blind manner in order to discriminate stego-images and cover-images. Even though universal classifiers have lower performance than specific embedding-targeted detectors, they are still important because of their flexibility and ability to be adjusted to completely unknown steganographic methods.

Typically, universal classifiers can be divided into two types: supervised and unsupervised. Supervised classification [45, 76, 117–120] has been already discussed in Section 2.3. While supervised classification requires to know in advance the label of each image (i.e. cover-image or stego-image) and then build a classifier based on labeled images, unsupervised classification works in a scenario of unlabeled images and classifies them automatically without user interference. The accuracy of supervised classifiers is limited if the training data is not perfectly representative of

cover source, which may result in mismatch problem [121]. Unsupervised classifiers try to overcome this problem of model mismatch by postponing building a cover model until the classification stage. However, to the best of our knowledge, there has not been yet a reliable method dealing with this scenario in steganalysis.

In universal steganalysis, the design of features is of crucial importance. Features used for classification should be sensitive to changes caused by embedding, yet insensitive to variations between covers including also some non-steganographic processing techniques. In general, the choice of suitable features and machine learning tools remains open problems [121].

2.5.3 Steganalysis of Jsteg Algorithm

Like steganalysis of LSB replacement in spatial domain, existing methods for steganalysis of Jsteg algorithm can be also divided into four categories. Structural detectors detect the presence of secret message by employing the symmetry of the histogram of DCT coefficients in natural images, which is disturbed by the operation of Jsteg embedding. Some representative structural detectors are Zhang and Ping (ZP) detector [122], DCT coefficient-based detector [123], and category attack [124, 125]. Furthermore, the power of structural detectors can be combined with theoretically well-founded ML principle [99] or the concept of Zero Message Hypothesis (ZMH) [96]. These two approaches have been formally analyzed in [126]. Similar to structural detectors for steganalysis of LSB replacement technique, the ZHM framework starts by choosing a feature vector $\mathbf{x}$ of the cover-image (e.g. trace subsets in case of SPA method), establishes the change in the feature vector $\mathbf{x}$ due to embedding algorithm Emb, then inverts embedding effects to provide a hypothetical feature vector $\hat{\mathbf{x}}$

$$\hat{\mathbf{x}} = \text{Emb}^{-1}(\mathbf{x}^r, r), \quad (2.43)$$

where $\mathbf{x}^r$ is the stego vector and r is the change rate defined as the ratio between the number of modified DCT coefficients and the maximum number of usable coefficients, thus $r = \frac{R}{2}$. Using cover assumptions and zero message properties (e.g. natural symmetry of the histogram of DCT coefficients), an appropriate penalty function $\text{znh}(\mathbf{x}) \geq 0$ is defined so that it returns zero on cover features and non-zero otherwise. Therefore, the change rate estimator $\hat{r}$ is defined as the solution of a minimization problem

$$\hat{r} = \arg \min_{r \geq 0} \text{znh}(\hat{\mathbf{x}}) = \arg \min_{r \geq 0} \text{znh}(\text{Emb}^{-1}(\mathbf{x}^r, r)). \quad (2.44)$$

The minimization in (2.44) can be performed either analytically or numerically by implementing a one-dimensional gradient-descent search over r . The main interest in [126] is that all features proposed in [104, 122, 124, 125] have been revisited within ZHM framework. The detector proposed in [123] has been also improved in [126] within ML framework using a more accurate model of DCT coefficients, namely Generalized Cauchy distribution. It can be noted that although ZMH is only a

heuristic framework and less statistically rigorous than ML framework, it has some important advantages in terms of low computational complexity and flexibility.

Although Jsteg algorithm replaces LSBs by secret message bits in DCT domain, the mathematical foundation of WS detectors can be also applied for steganalysis of Jsteg [127, 128]. Given a vector of AC coefficients $\mathbf{D} = \{D_1, D_2, \dots, D_N\}$, the WS-like detector is given by

$$\hat{R} \propto \sum_i w_i (D_i - \bar{D}_i) D_i. \quad (2.45)$$

The difference between the WS detector in (2.36) and the one in (2.45) is that the local predictor for cover AC coefficients is omitted since the expected value of AC coefficients is zero in natural images. The weigh w_i for each coefficient D_i is estimated by taking the coefficients at the same location as D_i but in four adjacent blocks. More details were provided in [128].

The hypothesis testing theory was also applied to the steganalysis of Jsteg algorithm. By relying the Laplacian model of DCT coefficients, a statistical test was designed in [129]. However, a considerable loss of power was revealed due to the fact that the Laplacian model is not accurate enough to characterize DCT coefficients.

2.6 Conclusion

This chapter discusses the emerging field of digital image forensics consisting of two main problems: image origin identification and image forgery detection. To address these problems, active forensic approach has been proposed by generating extrinsically fingerprints and adding them into the digital image in the image formation process, thus creates a trustworthy digital camera. However, active approach is of limited application due to many strict constraints in its protocols. Therefore, passive forensic approach has been considerably evolved to help solve these problems in their entirety. This approach relies on intrinsic traces left by the digital cameras in the image processing pipeline and by the manipulations themselves to gather forensic evidence of image origin or forgery. Some intrinsic fingerprints for identification of image source such as lens aberration, PRNU, CFA pattern and interpolation, and JPEG compression are reviewed. The task of steganalysis that aims to detect the mere presence of a secret message in a digital image is also discussed in this chapter.

The state of the art has shown that most of existing methods have been designed within classification framework. Hypothesis testing framework is of limited exploitation although this framework offers many advantages, namely that the statistical performance of detectors can be analytically established and a prescribed false alarm probability can be guaranteed. Besides, existing methods are designed using simplistic image models, which results in overall poor detection performance. This thesis focuses on applying the hypothesis testing theory in digital image forensics based on an accurate image model, which is established by modeling the main steps in the image processing pipeline. These aspects will be discussed in the rest of the thesis.

Overview on Statistical Modeling of Natural Images

Contents

3.1 Introduction	41
3.2 Spatial-Domain Image Model	41
3.2.1 Poisson-Gaussian and Heteroscedastic Noise Model	42
3.2.2 Non-Linear Signal-Dependent Noise Model	44
3.3 DCT Coefficient Model	45
3.3.1 First-Order Statistics of DCT Coefficients	45
3.3.2 Higher-Order Statistics of DCT Coefficients	46
3.4 Conclusion	46

3.1 Introduction

The application of hypothesis testing theory in digital image forensics requires an accurate statistical image model to achieve high detection performance. For instance, the PRNU-based image origin identification [30, 32] takes into account various noise sources during image acquisition inside a digital camera, which provides an image model allowing to accurately extract the fingerprint for source identification. An inaccurate image model will result in a poor detection performance, e.g. in case of statistical detectors [107, 129]. Therefore in this chapter, the state of the art on statistical modeling of natural images is reviewed. The statistical image modeling can be performed either in spatial domain or DCT domain.

The chapter is organized as follows. Section 3.2 analyzes noise statistics in spatial domain and presents some dominant image models widely used in image processing. Section 3.3 discusses empirical statistical models of DCT coefficients. Finally, Section 3.4 concludes the chapter.

3.2 Spatial-Domain Image Model

In this section, we adopt the representation of an arbitrary image $\mathbf{Z}$ as a column vector of length $N = N_r \times N_c$. The representation as a two-dimensional matrix is of

no interest in the study of statistical noise properties. The index of color channel is omitted for simplicity. Due to the stochastic nature of noise, a pixel is regarded as a random variable. Generally, the random variable z_i , $i \in \mathcal{I}$, can be decomposed as

$$z_i = \mu_{z_i} + \eta_{z_i}, \quad (3.1)$$

where $\mathcal{I} = \{1, \dots, N\}$ denotes the set of pixel indices, μ_{z_i} denotes the expectation of the pixel z_i in the absence of noise, and η_{z_i} accounts for all noise sources that interfere with the original signal. By convention, μ_X and σ_X^2 denote respectively the expectation and variance of a random variable X . Here, the expectation μ_{z_i} is considered deterministic and will not be modeled. However, the expectations differ from each other due to heterogeneity in a natural image. From (3.1), it is easily seen that the variance of noise η_{z_i} is equal to the variance of pixel z_i , i.e. $\sigma_{z_i}^2 = \sigma_{\eta_{z_i}}^2$.

Some models have been proposed in the literature for the noise η_{z_i} in an uncompressed image. They can be classified into two groups: signal-independent and signal-dependent noise models. While signal-independent noise models assume the stationarity of noise in the whole image, regardless original pixel intensity, signal-dependent noise models take into account the proportional dependence of noise variance on the original pixel intensity. A typical example for the group of signal-independent noise is the Additive White Gaussian Noise (AWGN). Besides, signal-dependent noise includes Poisson noise or film-grain noise [130], Poisson-Gaussian noise [131, 132], heteroscedastic noise model [133, 134], and non-linear noise model [135]. Although the AWGN model is widely adopted in image processing because of its simplicity, it ignores the contribution of Poisson noise to the image acquisition chain, which is the case of an image acquired by a digital camera. Noise sources in a natural image are inherently signal-dependent. Therefore, a signal-dependent noise model is more expected to be employed in further applications. Since our works mainly focus on signal-dependent noise, only the group of signal-dependent noise models are discussed in this section.

3.2.1 Poisson-Gaussian and Heteroscedastic Noise Model

The study of noise statistics requires to take into account the impact of Poisson noise related to the stochastic nature of photon-counting process and dark current [131–134, 136]. Let ξ_i denote the number of collected electrons with respect to the pixel z_i . The number of collected electrons ξ_i follows the Poisson distribution with mean λ_i and variance λ_i

$$\xi_i \sim \mathcal{P}(\lambda_i). \quad (3.2)$$

This Poisson noise results in the dependence of noise variance on original pixel intensity. The number of collected electrons is further degraded by the AWGN read-out noise η_r with variance ω^2 . Therefore, the RAW image pixel recorded by the image sensor can be defined as [136]

$$z_i = a \cdot (\xi_i + \eta_r), \quad (3.3)$$

where a is the analog gain controlled by the ISO sensitivity. This leads to the statistical distribution of the RAW pixel z_i

$$z_i \sim a \cdot [\mathcal{P}(\lambda_i) + \mathcal{N}(0, \omega^2)]. \quad (3.4)$$

This model is referred to as Poisson-Gaussian noise model [131, 132]. One interesting property of this model is the linear relation of pixel's expectation and variance. Taking mathematical expectation and variance from (3.4), we obtain

$$\mu_{z_i} = \mathbb{E}[z_i] = a \cdot \lambda_i, \quad (3.5)$$

$$\sigma_{z_i}^2 = \mathbb{V}\text{ar}[z_i] = a^2 \cdot (\lambda_i + \omega^2), \quad (3.6)$$

where $\mathbb{E}[X]$ and $\mathbb{V}\text{ar}[X]$ denote respectively the mathematical expectation and variance with respect to a random variable X . Consequently, the heteroscedastic relation is derived as

$$\sigma_{z_i}^2 = a \cdot \mu_{z_i} + b, \quad (3.7)$$

where $b = a^2\omega^2$.

In some image sensors, the collected electrons ξ_i may be added by a base pedestal parameter p_0 to constitute an offset-from-zero of the output pixel [133]

$$z_i \sim a \cdot [p_0 + \mathcal{P}(\lambda_i - p_0) + \mathcal{N}(0, \omega^2)]. \quad (3.8)$$

Hence, the parameter b is now given by $b = a^2\omega^2 - a^2p_0$. Therefore, the parameter b can be negative when $p_0 > \omega^2$.

To facilitate the application of this signal-dependent noise model, some works [133, 134] have attempted to approximate the Poisson distribution by the Gaussian distribution in virtue of a large number of collected electrons

$$\mathcal{P}(\lambda_i) \approx \mathcal{N}(\lambda_i, \lambda_i). \quad (3.9)$$

In fact, for $\lambda_i \geq 50$, the Gaussian approximation is already very accurate [133] while full-well capacity is largely above 100000 electrons. Finally, the statistical distribution of the RAW pixel z_i can be approximated as

$$z_i \sim \mathcal{N}(\mu_{z_i}, a \cdot \mu_{z_i} + b). \quad (3.10)$$

This model is referred to as heteroscedastic noise model in [134]. The term "heteroscedasticity" means that each pixel exposes a different variability with the other.

Both Poisson-Gaussian and heteroscedastic noise models are more accurate to characterize a RAW image than the conventional AWGN, but they do not take into account yet non-linear post-acquisition operations. Therefore, they are not appropriate for modeling a TIFF or JPEG image. Besides, it should be noted that the Poisson-Gaussian and heteroscedastic noise model assume that the effect of PRNU is negligible, namely that all the pixels respond to the incident light uniformly. The very small variation in pixel's response does not strongly affect its statistical distribution [136].

3.2.2 Non-Linear Signal-Dependent Noise Model

To establish a statistical model of a natural image in TIFF or JPEG format, it is necessary to take into account effects of post-acquisition operations in the image processing pipeline. However, as discussed in Section 2.2, the whole image processing pipeline is not as simple. Some processing steps that are implemented in a digital camera are difficult to model parametrically. One approach is to consider the digital camera as a *black box* in which we attempt to establish a relation between input irradiance and output intensity. This relation is called Camera Response Function (CRF), which is described by a sophisticated non-linear function $f_{\text{CRF}}(\cdot)$ [137]. Gamma correction might be the simplest model for the CRF with only one parameter. Other parametric models have been proposed for CRF such as polynomial model [137] or generalized gamma curve model [138].

Therefore, the pixel z_i can be formally defined as [135, 139]

$$z_i = f_{\text{CRF}}(E_i + \eta_{E_i}), \quad (3.11)$$

where E_i denotes the image irradiance and η_{E_i} accounts for all signal-independent and signal-dependent noise sources. We can note that although some methodologies have been proposed for estimation of CRF [50, 137, 140], it is also difficult to study noise statistics with those sophisticated models.

To facilitate the study of noise statistics, the authors in [135] exploit the first order of Taylor's series expansion

$$z_i = f_{\text{CRF}}(E_i + \eta_{E_i}) \approx f_{\text{CRF}}(E_i) + f'_{\text{CRF}}(E_i)\eta_{E_i}, \quad (3.12)$$

where f'_{CRF} denotes the first derivative of the CRF f_{CRF} . Therefore, a relation between noises before and after transformation by the CRF is obtained

$$\eta_{z_i} = f'_{\text{CRF}}(E_i)\eta_{E_i}. \quad (3.13)$$

It can be noted that even when noise before transformation is independent of the signal, the non-linear transformation f_{CRF} generates a dependence between pixel's expectation and variance.

Based on experimental observations, the authors in [135] obtain a non-linear parametric model

$$z_i = \mu_{z_i} + \mu_{z_i}^{\tilde{\gamma}} \cdot \eta_u, \quad (3.14)$$

where η_u is zero-mean stationary Gaussian noise, $\eta_u \sim \mathcal{N}(0, \sigma_{\eta_u}^2)$, and $\tilde{\gamma}$ is an exponential parameter to account for the non-linearity of the camera response. Here, taking variance on the both sides of (3.14), we obtain

$$\sigma_{z_i}^2 = \mu_{z_i}^{2\tilde{\gamma}} \cdot \text{Var}[\eta_u] = \mu_{z_i}^{2\tilde{\gamma}} \cdot \sigma_{\eta_u}^2. \quad (3.15)$$

In this model, the pixel z_i still follows the Gaussian distribution and the noise variance $\sigma_{\eta_{z_i}}^2$ is non-linearly dependent on the original pixel intensity μ_{z_i}

$$z_i \sim \mathcal{N}(\mu_{z_i}, \mu_{z_i}^{2\tilde{\gamma}} \cdot \sigma_{\eta_u}^2). \quad (3.16)$$

This model allows to represent several kinds of noise such as film-grain, Poisson noise by changing the parameters $\tilde{\gamma}$ and $\sigma_{\eta_u}^2$ (e.g $\tilde{\gamma} = 0.5$).

3.3 DCT Coefficient Model

3.3.1 First-Order Statistics of DCT Coefficients

Apart from modeling an image in the spatial domain, many researches attempt to model it in the DCT domain since the DCT is a fundamental operation in JPEG compression. The model of DCT coefficients has been considerably studied in the literature. However, a majority of DCT coefficient models has just been proposed without giving any mathematical foundation and analysis. Many researches focus on comparing the empirical data with a variety of popular statistical models by conducting the goodness-of-fit (GOF) test, e.g. the Kolmogorov-Smirnov (KS) or χ^2 test. Firstly, the Gaussian model for the DCT coefficients was conjectured in [141]. The Laplacian model was verified in [142] by performing the KS test. This Laplacian model remains a dominant choice in image processing because of its simplicity and relative accuracy. Other possible models such as Gaussian mixture [143] and Cauchy [144] were also proposed. In order to model the DCT coefficients more accurately, the previous models were extended to the generalized versions including the Generalized Gaussian (GG) [145] and the Generalized Gamma (GG) [146] models. It has been recently reported in [146] that the GG model outperforms the Laplacian and GG model. Far from giving a mathematical foundation of DCT coefficient model, these empirical models were only verified using GOF test on a few standard images. Thus, they can not guarantee the accuracy of the chosen model to a wide range of images, which leads to a lack of robustness.

The first mathematical analysis for DCT coefficients is given in [147]. It relies on a doubly stochastic model combining DCT coefficient statistics in a block whose variance is constant with the variability of block variance in a natural image. However, this analysis is incomplete due to the lack of mathematical justification for the block variance model. Nevertheless, it has shown an interest for further improvements. Therefore, here we provide a discussion about this mathematical foundation.

Let I denote AC coefficient and σ_{blk}^2 denote block variance. The DC coefficient is not addressed in this work [147]. The index of frequency is omitted for the sake of clarity. Using the conditional probability, the doubly stochastic model is given by

$$f_I(x) = \int_0^\infty f_{I|\sigma_{\text{blk}}^2}(x|t) f_{\sigma_{\text{blk}}^2}(t) dt \quad x \in \mathbb{R}, \quad (3.17)$$

where $f_X(x)$ denotes the probability density function (pdf) of a random variable X . This doubly stochastic model can be considered as infinite mixture of Gaussian distributions [148, 149]. From the establishment of DCT coefficients in (2.11), it is noted that each DCT coefficient is a weighted sum of random variables. If the block variance σ_{blk}^2 is constant, the AC coefficient I can be approximated as a zero-mean Gaussian random variable based on the Central Limit Theorem (CLT)

$$f_{I|\sigma_{\text{blk}}^2}(x|t) = \frac{1}{\sqrt{2\pi t}} \exp\left(-\frac{x^2}{2t}\right). \quad (3.18)$$

Even though the pixels are spatially correlated in a 8×8 block due to demosaicing algorithms implemented in a digital camera, the CLT can still be used for Gaussian approximation of a sum of correlated random variables [150]. It remains to find the pdf of σ_{blk}^2 to derive the final pdf of the AC coefficient I . To this end, it was reported in [147] that from experimental observations, the block variance σ_{blk}^2 can be modeled by exponential or half-Gaussian distribution. These two distributions can lead to the Laplacian distribution for the DCT coefficient I [147]. However, as stated above, due to the fact that the pdf of block variance σ_{blk}^2 is not mathematically justified, this mathematical framework is incomplete.

3.3.2 Higher-Order Statistics of DCT Coefficients

The above discussion only considers the first-order statistics (i.e. histogram) of DCT coefficients. The DCT coefficients at the same frequency are collected and treated separately. An implicit assumption adopted in this procedure is that the DCT coefficients at the same frequency are i.i.d realizations of a random variable. However, this is not always true in a natural image because DCT coefficients exhibit dependencies (or correlation) between them. There are two fundamental kinds of correlation between DCT coefficients [151], which have been successfully exploited in some applications [126, 151, 152]

1. *intra-block correlation*: A well-known feature of DCT coefficients in a natural image is that the magnitudes of AC coefficients decrease as the frequency increases along the zig-zag order. This correlation reflects the dependence between DCT coefficients within a same 8×8 block. Typically, this correlation is weak since coefficients at different frequencies correspond to different basis functions.
2. *inter-block correlation*: Although the DCT base can provide a good decorrelation, resulting coefficients are still correlated slightly with their neighbors at the same frequency. We refer this kind of correlation as inter-block correlation.

In general, the correlation between DCT coefficients could be captured by adjacency matrix [126].

3.4 Conclusion

This chapter reviews some statistical image models in spatial domain and DCT domain. In spatial domain, two groups of signal-independent and signal-dependent noise models are discussed. From above statistical analysis, we can draw an important insight: noise in natural images is inherently signal-dependent. In DCT domain, some empirical models of DCT coefficient are presented. However, most of DCT coefficient models are given without mathematical justification. It is still necessary to establish an accurate image model that can be exploited in further applications.

Part II

Statistical Modeling and Estimation for Natural Images from RAW Format to JPEG Format

Statistical Image Modeling and Estimation of Model Parameters

Contents

4.1	Introduction	50
4.2	Statistical Modeling of RAW Images	50
4.2.1	Heteroscedastic Noise Model	50
4.2.2	Estimation of Parameters (a, b) in the Heteroscedastic Noise Model	52
4.2.2.1	WLS Estimation	53
4.2.2.2	Statistical Properties of WLS Estimates	54
4.3	Statistical Modeling of TIFF Images	57
4.3.1	Generalized Noise Model	57
4.3.2	Estimation of Parameters $(\tilde{a}, \tilde{b})$ in the Generalized Noise Model	59
4.3.2.1	Edge Detection and Image Segmentation	59
4.3.2.2	Maximum Likelihood Estimation	60
4.3.3	Application to Image Denoising	61
4.3.4	Numerical Experiments	62
4.4	Statistical Modeling in DCT Domain	65
4.4.1	Statistical Model of Quantized DCT Coefficients	65
4.4.1.1	Statistical Model of Block Variance and Unquantized DCT Coefficients	65
4.4.1.2	Impact of Quantization	68
4.4.2	Estimation of Parameters (α, β) from Unquantized DCT Coefficients	69
4.4.3	Estimation of Parameters (α, β) from Quantized DCT Coefficients	70
4.4.4	Numerical Experiments	71
4.5	Conclusion	73

4.1 Introduction

Chapter 3 has presented an overview on statistical modeling of natural images in spatial domain and DCT domain. Most of existing models in the literature were provided empirically. The goal of this chapter is to establish a mathematical framework of studying statistical properties of natural images along image processing pipeline of a digital camera. The study is performed in spatial domain and DCT domain. In the spatial domain, the heteroscedastic noise model is firstly recalled, and a method for estimating the parameters of the heteroscedastic noise model following the Weighted Least Square (WLS) approach is proposed in Section 4.2. The analytic establishment of WLS estimates allows us to study their statistical properties, which is of importance for designing statistical tests. The WLS estimation of parameters (a, b) has been presented in [134].

Next, Section 4.3 presents the study of noise statistics in a TIFF image by starting from the heteroscedastic noise model and take into account the effect of gamma correction, resulting in the generalized signal-dependent noise model. It is shown that the generalized noise model is also relevant to characterize JPEG images with moderate-to-high quality factors ($Q \geq 70$). This section also proposes a method that can estimate the parameters of the generalized noise model accurately from a single image. Numerical results on a large image database show the relevance of the proposed method. The generalized noise model could be useful in many applications. A direct application for image denoising is proposed in this section. The foundation of generalized noise model and estimation of model parameters have been presented in [153].

Section 4.4 describes the mathematical framework of modeling the statistical distribution of DCT coefficients. To simplify the study, the approach is based on the main assumption that the pixels are identically distributed (not necessarily independent) within a 8×8 block. Consequently, the statistical distribution of block variance can be approximated, thus the model of unquantized DCT coefficients is provided. Moreover, it is proposed to take into account the quantization operation to provide a final model of quantized DCT coefficients. The parameters of DCT coefficient model can be estimated following the ML approach. Numerical results show that the proposed model outperforms other existing models including Laplacian, GG, and GF model. Section 4.5 concludes the chapter. The foundation of DCT coefficient model has been presented in [154].

4.2 Statistical Modeling of RAW Images

4.2.1 Heteroscedastic Noise Model

The RAW image acquisition has been discussed in Section 2.2.1. Let $\mathbf{Z} = (z_i)_{i \in \mathcal{I}}$ denote a RAW image acquired by the image sensor. Typically, the model of RAW pixel consists of a Poissonian part that addresses the photon shot noise and dark current and a Gaussian part for the remaining stationary disturbances, e.g. read-

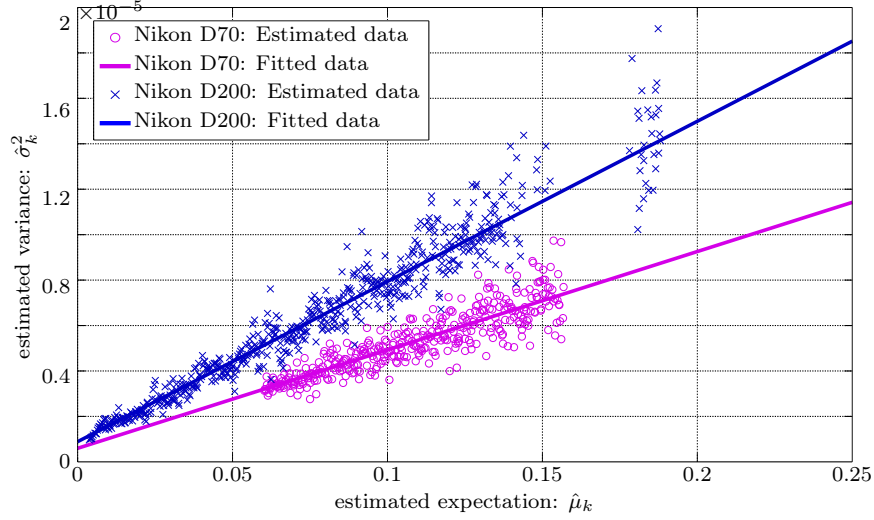


Figure 4.1: Scatter-plot of pixels' expectation and variance from a natural RAW image with ISO 200 captured by *Nikon D70* and *Nikon D200* cameras. The image is segmented into homogeneous segments. In each segment, the expectation and variance are calculated and the parameters (a, b) are estimated as proposed in Section 4.2.2. The dash line is drawn using the estimated parameters (a, b) . Only the red channel is used in this experiment.

out noise. For the sake of simplification, the Gaussian approximation of the Poisson distribution can be exploited because of a large number of collected electrons, which leads to the heteroscedastic noise model [133, 134]

$$z_i \sim \mathcal{N}(\mu_i, a\mu_i + b), \quad (4.1)$$

where μ_i denotes the expectation of the pixel z_i . The heteroscedastic noise model, which gives the noise variance as a linear function of pixel's expectation, characterizes a RAW image more accurately than the conventional AWGN model. The heteroscedastic noise model (4.1) is illustrated in Figure 4.1. It is assumed that the noise corrupting each RAW pixel is statistically independent of those of neighbor pixels [133, 136]. In this section it is assumed that the phenomenon of clipping is absent from a natural RAW image for the sake of simplification, i.e. the probability that one observation z_i exceeds over the boundary 0 or $B = 2^\nu - 1$ is negligible. More details about the phenomenon of clipping are given in [133, 155] and in Chapter 8.

In practice, the PRNU weakly affects the parameter a in the heteroscedastic noise model (4.1). Nevertheless, in the problem of camera model identification, the PRNU is assumed to be negligible, i.e. the parameter a remains constant for every pixel.

4.2.2 Estimation of Parameters (a, b) in the Heteroscedastic Noise Model

Estimation of noise model parameters can be performed from a single image or multiple images. From a practical point of view, we mainly focus on noise model parameter estimation from a single image. Several methods have been proposed in the literature for estimation of signal-dependent noise model parameters, see [133–135, 139, 156]. They rely on similar basic steps but differ in details. The common methodology starts from obtaining local estimates of noise variance and image content, then performing the curve fitting to the scatter-plot based on the prior knowledge of noise model. The existing methods involve two main difficulties: influence of image content and spatial correlation of noise in a natural image. In fact, homogeneous regions where local expectations and variances are estimated are obtained by performing edge detection and image segmentation. However, the accuracy of those local estimates may be contaminated due to the presence of outliers (textures, details and edges) in the homogeneous regions. Moreover, because of the spatial correlation between pixels, the local estimates of noise variance can be overestimated. Overall, the two difficulties may result in inaccurate estimation of noise parameters.

For the design of subsequent tests, the parameters (a, b) should be estimated following the ML approach and statistical properties of ML estimates should be analytically established. One interesting method is proposed in [133] for ML estimation of parameters (a, b) . However, that method can not provide an analytic expression of ML estimates due to the difficulty of resolving the complicated system of partial derivatives. Therefore, ML estimates are only numerically solved by using the Nelder-Mead optimization method [157]. Although ML estimates given by that method are relatively accurate, they involve three main drawbacks. First, the convergence of the maximization process and the sensitivity of the solution to initial conditions have not been analyzed yet. Second, the Bayesian approach used in [133] with a fixed uniform distribution might be doubtful in practice. Finally, it seems impossible to establish statistical properties of the estimates.

This section proposes a method for estimation of parameters (a, b) from a single image. The proposed method relies on the same technique of image segmentation used in [133] in order to obtain local estimates in homogeneous regions. Subsequently, the proposed method is based on the WLS approach to take into account heteroscedasticity and statistical properties of local estimates. One important advantage is that WLS estimates can be analytically provided, which allows us to study statistical properties of WLS estimates. Moreover, the WLS estimates are asymptotically equivalent to the ML estimates in large samples when the weights are consistently estimated, as explained in [158, 159].

4.2.2.1 WLS Estimation

The RAW image $\mathbf{Z}$ is first transformed into the wavelet domain and then segmented into K non-overlapping homogeneous segments, denoted S_k , of size n_k , $k \in \{1, \dots, K\}$. The readers are referred to [133] for more details of segmentation technique. In each segment S_k , pixels are assumed to be i.i.d, thus they have the same expectation and variance. Let $\mathbf{z}_k^{\text{wapp}} = (z_{k,i}^{\text{wapp}})_{i \in \{1, \dots, n_k\}}$ and $\mathbf{z}_k^{\text{wdet}} = (z_{k,i}^{\text{wdet}})_{i \in \{1, \dots, n_k\}}$ be respectively the vector of wavelet approximation coefficients and wavelet detail coefficients representing the segment S_k . Because the transformation is linear, the coefficients $z_{k,i}^{\text{wapp}}$ and $z_{k,i}^{\text{wdet}}$ also follow the Gaussian distribution

$$z_{k,i}^{\text{wapp}} \sim \mathcal{N}(\mu_k, \|\varphi\|_2^2 \sigma_k^2), \quad (4.2)$$

$$z_{k,i}^{\text{wdet}} \sim \mathcal{N}(0, \sigma_k^2), \quad (4.3)$$

where μ_k denotes expectation of all pixels in the segment S_k , $\sigma_k^2 = a\mu_k + b$, and φ is the 2-D normalized wavelet scaling function. Hence, the ML estimates of local expectation μ_k and local variance σ_k^2 are given by

$$\hat{\mu}_k = \frac{1}{n_k} \sum_{i=1}^{n_k} z_{k,i}^{\text{wapp}}, \quad (4.4)$$

$$\hat{\sigma}_k^2 = \frac{1}{n_k - 1} \sum_{i=1}^{n_k} (z_{k,i}^{\text{wdet}} - \bar{z}_k^{\text{wdet}})^2, \quad \text{with} \quad \bar{z}_k^{\text{wdet}} = \frac{1}{n_k} \sum_{i=1}^{n_k} z_{k,i}^{\text{wdet}}. \quad (4.5)$$

The estimate $\hat{\mu}_k$ is unbiased and follows the Gaussian distribution

$$\hat{\mu}_k \sim \mathcal{N}\left(\mu_k, \frac{\|\varphi\|_2^2}{n_k} \sigma_k^2\right), \quad (4.6)$$

while the estimate $\hat{\sigma}_k^2$ follows a scaled chi-square distribution with $n_k - 1$ degrees of freedom. This distribution can also be accurately approximated as the Gaussian distribution for large n_k [160]:

$$\hat{\sigma}_k^2 \sim \mathcal{N}\left(\sigma_k^2, \frac{2}{n_k - 1} \sigma_k^4\right), \quad (4.7)$$

Figure 4.1 illustrates a scatter-plot of all the pairs $\{(\hat{\mu}_k, \hat{\sigma}_k^2)\}$ extracted from real natural RAW images of *Nikon D70* and *Nikon D200* cameras.

The parameters (a, b) are estimated by considering all the pairs $\{(\hat{\mu}_k, \hat{\sigma}_k^2)\}_{k=1}^K$ where the local variance $\hat{\sigma}_k^2$ is treated as a heteroscedastic model of the local expectation $\hat{\mu}_k$. This model is formulated as follows

$$\hat{\sigma}_k^2 = a\hat{\mu}_k + b + s_k \epsilon_k, \quad (4.8)$$

where ϵ_k are independent and identically distributed as standard Gaussian variable and s_k is a function of local mean μ_k . A direct calculation from (4.8) shows that

$$\begin{aligned} s_k^2 &= \text{Var}[\hat{\sigma}_k^2] - \text{Var}[a\hat{\mu}_k + b] \\ &= \frac{2}{n_k - 1} \sigma_k^4 - a^2 \frac{\|\varphi\|_2^2}{n_k} \sigma_k^2 = \frac{2}{n_k - 1} (a\mu_k + b)^2 - a^2 \frac{\|\varphi\|_2^2}{n_k} (a\mu_k + b). \end{aligned} \quad (4.9)$$

The heteroscedasticity in the model (4.8) is governed by different residuals s_k . The modeling in (4.8) also takes into account stochastic errors associated with estimation of local estimates $\hat{\mu}_k$ and $\hat{\sigma}_k^2$. It can be noted that the residuals s_k can not be defined since the parameters μ_k and (a, b) are unknown in practice. In order to apply WLS approach, a popular strategy is to first perform the Ordinary Least Squares (OLS) approach to obtain initial estimates of the parameters (a, b) allowing to provide the estimates $\hat{s}_k^2$, then utilize the weights $\hat{w}_k = \frac{1}{\hat{s}_k^2}$ for WLS estimation. The OLS estimates $(\hat{a}_L, \hat{b}_L)$ are given by [161]:

$$\begin{pmatrix} \hat{a}_L \\ \hat{b}_L \end{pmatrix} = (\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T \mathbf{v}, \quad (4.10)$$

where

$$\mathbf{H} = \begin{pmatrix} \hat{\mu}_1 & 1 \\ \vdots & \vdots \\ \hat{\mu}_K & 1 \end{pmatrix}, \quad \text{and} \quad \mathbf{v} = \begin{pmatrix} \hat{\sigma}_1^2 \\ \vdots \\ \hat{\sigma}_K^2 \end{pmatrix},$$

where $\mathbf{H}^T$ and $\mathbf{H}^{-1}$ are the transpose and inverse of the matrix $\mathbf{H}$, respectively. Therefore, the consistent estimates $\hat{s}_k^2$ can be directly computed as

$$\hat{s}_k^2 = \frac{2}{n_k - 1} (\hat{a}_L \hat{\mu}_k + \hat{b}_L)^2 - \hat{a}_L^2 \frac{\|\varphi\|_2^2}{n_k} (\hat{a}_L \hat{\mu}_k + \hat{b}_L). \quad (4.11)$$

By using the estimated weights $\hat{w}_k = \frac{1}{\hat{s}_k^2}$, the WLS estimates $(\hat{a}, \hat{b})$ are given by

$$\begin{pmatrix} \hat{a} \\ \hat{b} \end{pmatrix} = (\mathbf{H}^T \mathbf{W} \mathbf{H})^{-1} \mathbf{H}^T \mathbf{W} \mathbf{v}, \quad (4.12)$$

where $\mathbf{W} = \text{diag}(\hat{w}_1, \dots, \hat{w}_K)$ is the diagonal weight matrix.

4.2.2.2 Statistical Properties of WLS Estimates

As discussed above, the analytic establishment of WLS estimates $(\hat{a}, \hat{b})$ allows us to study their statistical properties. According to [158, 159], the WLS estimates $(\hat{a}, \hat{b})$ are consistent and follow the bivariate Gaussian distribution

$$\begin{pmatrix} \hat{a} \\ \hat{b} \end{pmatrix} \sim \mathcal{N} \left(\begin{pmatrix} a \\ b \end{pmatrix}, \begin{pmatrix} \sigma_a^2 & \sigma_{ab} \\ \sigma_{ab} & \sigma_b^2 \end{pmatrix} \right), \quad (4.13)$$

where σ_a^2 , σ_b^2 , σ_{ab} denote respectively the variance of $\hat{a}$, the variance of $\hat{b}$ and the covariance between $\hat{a}$ and $\hat{b}$. The covariance matrix also needs to be defined.

From (4.12), the WLS estimates $(\hat{a}, \hat{b})$ can be alternatively written as

$$\hat{a} = \frac{\sum_{k=1}^K \hat{w}_k (\hat{\mu}_k - \hat{\bar{\mu}}) \hat{\sigma}_k^2}{\sum_{k=1}^K \hat{w}_k (\hat{\mu}_k - \hat{\bar{\mu}})^2} \quad \text{and} \quad \hat{b} = \hat{\bar{\sigma}}^2 - \hat{a} \hat{\bar{\mu}}, \quad (4.14)$$

where $\widehat{\mu}$ is the weighted mean and $\widehat{\sigma}^2$ is the weighted variance

$$\widehat{\mu} = \frac{\sum_{k=1}^K \hat{w}_k \hat{\mu}_k}{\sum_{k=1}^K \hat{w}_k} \quad \text{and} \quad \widehat{\sigma}^2 = \frac{\sum_{k=1}^K \hat{w}_k \hat{\sigma}_k^2}{\sum_{k=1}^K \hat{w}_k}. \quad (4.15)$$

For brevity, let denote

$$U_1 = \sum_{k=1}^K \hat{w}_k (\hat{\mu}_k - \widehat{\mu}) \hat{\sigma}_k^2 \quad \text{and} \quad U_2 = \sum_{k=1}^K \hat{w}_k (\hat{\mu}_k - \widehat{\mu})^2, \quad (4.16)$$

such that $\hat{a} = U_1/U_2$. It is proposed to study statistical properties of U_1 and U_2 , then employ the Delta method in Lemma A.3 to calculate the covariance matrix.

It can be noted that $\hat{\mu}_k$ and $\hat{\sigma}_k^2$ are mutually independent due to the orthogonality of the wavelets [133]. Moreover, the pair $(\hat{\mu}_i, \hat{\sigma}_i^2)$ and $(\hat{\mu}_j, \hat{\sigma}_j^2)$ are also mutually independent because the corresponding segment S_i and S_j are non-overlapping. These observations can facilitate the study of covariance matrix.

Because $(\hat{a}_L, \hat{b}_L)$ are the consistent estimates of (a, b) , i.e. they asymptotically converge in probability to their true value: $\hat{a}_L \xrightarrow{P} a$ and $\hat{b}_L \xrightarrow{P} b$, from the Continuous Mapping Theorem [20, theorem 11.2.13], we can derive $\hat{w}_k \xrightarrow{P} w_k$. It follows from the Slutsky's Theorem [20, theorem 11.2.11] that

$$\widehat{\mu} \xrightarrow{d} \mathcal{N} \left(\bar{\mu}, \frac{\sum_{k=1}^K w_k^2 \frac{\|\varphi\|_2^2}{n_k} \sigma_k^2}{\left(\sum_{k=1}^K w_k \right)^2} \right), \quad (4.17)$$

$$\widehat{\sigma}^2 \xrightarrow{d} \mathcal{N} \left(\bar{\sigma}^2, \frac{\sum_{k=1}^K w_k^2 \frac{2}{n_k-1} \sigma_k^4}{\left(\sum_{k=1}^K w_k \right)^2} \right), \quad (4.18)$$

where

$$\bar{\mu} = \frac{\sum_{k=1}^K w_k \mu_k}{\sum_{k=1}^K w_k}, \quad (4.19)$$

$$\bar{\sigma}^2 = \frac{\sum_{k=1}^K w_k \sigma_k^2}{\sum_{k=1}^K w_k} = a\bar{\mu} + b. \quad (4.20)$$

Based on the linearity property of the Gaussian distribution, from (4.6) and (4.17), it is easily shown that

$$\hat{\mu}_k - \widehat{\mu} \sim \mathcal{N} \left(\mu_k - \bar{\mu}, \frac{\|\varphi\|_2^2}{n_k} \sigma_k^2 + \text{Var}[\widehat{\mu}] \right), \quad (4.21)$$

where $\text{Var}[\widehat{\mu}]$ is given in (4.17). Combining (4.7) and (4.21), a direct calculation yields to

$$\mathbb{E}[(\hat{\mu}_k - \widehat{\mu}) \hat{\sigma}_k^2] = (\mu_k - \bar{\mu}) \sigma_k^2 \quad (4.22)$$

$$\text{Var}[(\hat{\mu}_k - \widehat{\mu}) \hat{\sigma}_k^2] = \sigma_k^4 \left[\frac{2}{n_k-1} (\mu_k - \bar{\mu})^2 + \frac{\|\varphi\|_2^2}{n_k} \sigma_k^2 \text{Var}[\widehat{\mu}] \right] + o(n_k^{-2}), \quad (4.23)$$

where the notation $x = o(y)$, with $y > 0$, means that $\frac{x}{y}$ tends to 0 as y tends to 0. Hence, the mathematical expectation and variance of U_1 can be given by

$$\mathbb{E}[U_1] = \sum_{k=1}^K w_k (\mu_k - \bar{\mu}) \sigma_k^2, \quad (4.24)$$

$$\mathbb{V}\text{ar}[U_1] = \sum_{k=1}^K w_k^2 \sigma_k^4 \left[\frac{2}{n_k - 1} (\mu_k - \bar{\mu})^2 + \frac{\|\varphi\|_2^2}{n_k} \sigma_k^2 + \mathbb{V}\text{ar}[\hat{\mu}] \right]. \quad (4.25)$$

From (4.21), based on the definitions of the mathematical expectation and the variance, one obtains

$$\mathbb{E}[(\hat{\mu}_k - \hat{\mu})^2] = (\mu_k - \bar{\mu})^2 + \frac{\|\varphi\|_2^2}{n_k} \sigma_k^2 + \mathbb{V}\text{ar}[\hat{\mu}], \quad (4.26)$$

$$\mathbb{V}\text{ar}[(\hat{\mu}_k - \hat{\mu})^2] = 4(\mu_k - \bar{\mu})^2 \left(\frac{\|\varphi\|_2^2}{n_k} \sigma_k^2 + \mathbb{V}\text{ar}[\hat{\mu}] \right) + o(n_k^{-2}). \quad (4.27)$$

Hence, the mathematical expectation and variance of U_2 are given by

$$\mathbb{E}[U_2] = \sum_{k=1}^K w_k \left[(\mu_k - \bar{\mu})^2 + \frac{\|\varphi\|_2^2}{n_k} \sigma_k^2 + \mathbb{V}\text{ar}[\hat{\mu}] \right], \quad (4.28)$$

$$\mathbb{V}\text{ar}[U_2] = \sum_{k=1}^K 4w_k^2 (\mu_k - \bar{\mu})^2 \left(\frac{\|\varphi\|_2^2}{n_k} \sigma_k^2 + \mathbb{V}\text{ar}[\hat{\mu}] \right). \quad (4.29)$$

On the other hand, from the definition of the covariance, a direct calculation yields to

$$\begin{aligned} \mathbb{C}\text{ov}[(\hat{\mu}_k - \hat{\mu}) \hat{\sigma}_k^2, (\hat{\mu}_k - \hat{\mu})^2] &= \mathbb{E}[(\hat{\mu}_k - \hat{\mu})^3 \hat{\sigma}_k^2] - \mathbb{E}[(\hat{\mu}_k - \hat{\mu}) \hat{\sigma}_k^2] \mathbb{E}[(\hat{\mu}_k - \hat{\mu})^2] \\ &= \sigma_k^2 \mathbb{E}[(\hat{\mu}_k - \hat{\mu})^3] - \sigma_k^2 \mathbb{E}[(\hat{\mu}_k - \hat{\mu})] \mathbb{E}[(\hat{\mu}_k - \hat{\mu})^2] \\ &= 2\sigma_k^2 (\mu_k - \bar{\mu}) \left(\frac{\|\varphi\|_2^2}{n_k} \sigma_k^2 + \mathbb{V}\text{ar}[\hat{\mu}] \right). \end{aligned} \quad (4.30)$$

Hence, the covariance between U_1 and U_2 can be given by

$$\begin{aligned} \mathbb{C}\text{ov}[U_1, U_2] &= \sum_{k=1}^K w_k^2 \mathbb{C}\text{ov}[(\hat{\mu}_k - \hat{\mu}) \hat{\sigma}_k^2, (\hat{\mu}_k - \hat{\mu})^2] \\ &= \sum_{k=1}^K 2w_k^2 \sigma_k^2 (\mu_k - \bar{\mu}) \left(\frac{\|\varphi\|_2^2}{n_k} \sigma_k^2 + \mathbb{V}\text{ar}[\hat{\mu}] \right). \end{aligned} \quad (4.31)$$

Therefore, from the Delta method in Lemma A.3 the variance σ_a^2 can be approximated as

$$\begin{aligned} \sigma_a^2 &= \mathbb{V}\text{ar} \left[\frac{U_1}{U_2} \right] \\ &= \frac{\mathbb{V}\text{ar}[U_1]}{\mathbb{E}^2[U_2]} - \frac{2\mathbb{E}[U_1]}{\mathbb{E}^3[U_2]} \mathbb{C}\text{ov}[U_1, U_2] + \frac{\mathbb{E}^2[U_1]}{\mathbb{E}^4[U_2]} \mathbb{V}\text{ar}[U_2] + \left(\left(\sum_{k=1}^K n_k^2 \right)^{-2} \right). \end{aligned} \quad (4.32)$$

Additionally, the variance σ_b^2 and the covariance σ_{ab} are given by

$$\sigma_b^2 = \mathbb{V}\text{ar}[\hat{\sigma}^2 - \hat{a}\hat{\mu}] = \mathbb{V}\text{ar}[\hat{\sigma}^2] + a^2\mathbb{V}\text{ar}[\hat{\mu}] + \bar{\mu}^2\sigma_a^2 + \sigma_a^2\mathbb{V}\text{ar}[\hat{\mu}] \quad (4.33)$$

$$\sigma_{ab} = \mathbb{C}\text{ov}[\hat{a}, \hat{\sigma}^2 - \hat{a}\hat{\mu}] = -\bar{\mu}\sigma_a^2. \quad (4.34)$$

4.3 Statistical Modeling of TIFF Images

4.3.1 Generalized Noise Model

To establish a model of TIFF images, it is proposed to start from the heteroscedastic noise model (4.1) and take into account the effects of post-acquisition operations in the image processing pipeline. For the sake of simplicity, the operations of demosaicing and white-balancing are assumed to be linear (see more details in Section 2.2.2), i.e. a white-balanced pixel, say $\dot{z}_i$, can be written as: $\dot{z}_i = \varpi z_i$ where ϖ is a factor representing the effects of demosaicing and white-balancing. Therefore, from (4.1), it is easily shown that the white-balanced pixel $\dot{z}_i$ also follows the Gaussian distribution

$$\dot{z}_i \sim \mathcal{N}(\dot{\mu}_i, \tilde{a}\dot{\mu}_i + \tilde{b}), \quad \text{with} \quad \tilde{a} = \varpi a \quad \text{and} \quad \tilde{b} = \varpi^2 b, \quad (4.35)$$

where $\dot{\mu}_i$ denotes the expectation of the white-balanced pixel $\dot{z}_i$. The relation between the expectation and variance of the white-balanced pixel $\dot{z}_i$ remains linear. The gamma correction is defined by the element-wise power-law expression $\ddot{z}_i = \dot{z}_i^{\frac{1}{\gamma}}$. By applying the change of variables theorem, the pdf of the gamma-corrected pixel $\ddot{z}_i$ is given by

$$f_{\ddot{z}_i}(t) = \frac{\gamma t^{\gamma-1}}{\sqrt{2\pi(\tilde{a}\dot{\mu}_i + \tilde{b})}} \exp\left(-\frac{(t^\gamma - \dot{\mu}_i)^2}{2(\tilde{a}\dot{\mu}_i + \tilde{b})}\right). \quad (4.36)$$

It is desirable to establish the relation between expectation and variance of the gamma-corrected pixel $\ddot{z}_i$. Equivalently, Equation (4.35) can be rewritten as

$$\dot{z}_i = \dot{\mu}_i + \eta_{\dot{z}_i}, \quad \text{with} \quad \eta_{\dot{z}_i} \sim \mathcal{N}(0, \tilde{a}\dot{\mu}_i + \tilde{b}), \quad (4.37)$$

where $\eta_{\dot{z}_i}$ accounts for zero-mean signal-dependent Gaussian noise after white-balancing. Therefore, the gamma-corrected $\ddot{z}_i$ is given by

$$\ddot{z}_i = \dot{z}_i^{\frac{1}{\gamma}} = (\dot{\mu}_i + \eta_{\dot{z}_i})^{\frac{1}{\gamma}} = \dot{\mu}_i^{\frac{1}{\gamma}} \left(1 + \frac{\eta_{\dot{z}_i}}{\dot{\mu}_i}\right)^{\frac{1}{\gamma}}. \quad (4.38)$$

By using the Taylor's series expansion of $(1+x)^{\frac{1}{\gamma}}$ at $x=0$, it follows that

$$\ddot{z}_i = \dot{\mu}_i^{\frac{1}{\gamma}} + \frac{1}{\gamma} \dot{\mu}_i^{\frac{1}{\gamma}-1} \eta_{\dot{z}_i} + o\left(\frac{\eta_{\dot{z}_i}}{\dot{\mu}_i}\right) \approx \ddot{\mu}_i + \frac{1}{\gamma} \ddot{\mu}_i^{1-\gamma} \eta_{\dot{z}_i}, \quad (4.39)$$

where $\ddot{\mu}_i = \dot{\mu}_i^{\frac{1}{\gamma}}$ is the expectation of the gamma-corrected pixel $\ddot{z}_i$. Taking mathematical expectation and variance on the both sides of the equation (4.39), the

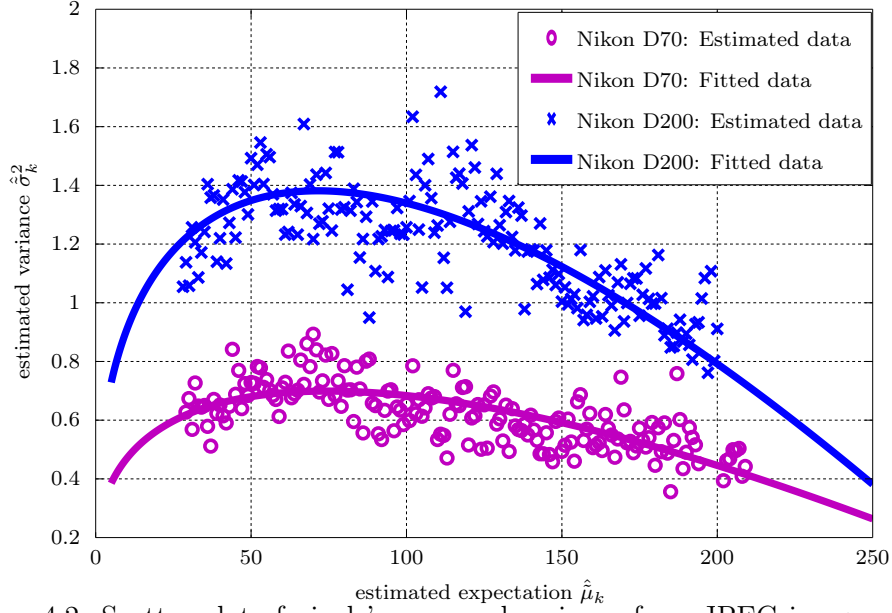


Figure 4.2: Scatter-plot of pixels' mean and variance from JPEG images with ISO 200 issued from *Nikon D70* and *Nikon D200* cameras. The red channel is used in this experiment. The image is segmented into homogeneous segments to estimate local means and variances. The generalized noise model is used to fit to the data.

relation between gamma-corrected pixel's expectation and variance is given as

$$\tilde{\sigma}_i^2 = \frac{1}{\gamma^2} \ddot{\mu}_i^{2-2\gamma} \text{Var}[\eta_{\tilde{z}_i}^2] = \frac{1}{\gamma^2} \ddot{\mu}_i^{2-2\gamma} (\tilde{a} \ddot{\mu}_i^\gamma + \tilde{b}). \quad (4.40)$$

If the camera exports the output image in TIFF format, the gamma-corrected image undergoes the quantization $\mathcal{Q}_\kappa$ with step κ in the image acquisition pipeline. It should be noted that if the camera exports output image in JPEG format, the quantization operation is assumed to be performed after operations in the DCT domain. Under mild assumptions [162], the quantization noise can be modeled as an additive noise that is uniformly distributed on the interval $[-\frac{\kappa}{2}, \frac{\kappa}{2}]$ and uncorrelated with the input signal. Therefore, the output TIFF pixel $\tilde{z}_i$ can be defined as

$$\tilde{z}_i = \ddot{z}_i + \eta_{\mathcal{Q}_\kappa} \quad (4.41)$$

where $\eta_{\mathcal{Q}_\kappa}$ denotes the quantization noise. The variance of the uniformly-distributed quantization noise is $\frac{\kappa^2}{12}$. Taking into account this variance, the generalized noise model in a TIFF image is derived as

$$\tilde{\sigma}_i^2 = \text{Var}[\tilde{z}_i] + \text{Var}[\eta_{\mathcal{Q}_\kappa}] = \frac{1}{\gamma^2} \tilde{\mu}_i^{2-2\gamma} (\tilde{a} \tilde{\mu}_i^\gamma + \tilde{b}) + \frac{\kappa^2}{12}, \quad (4.42)$$

where $\tilde{\mu}_i = \ddot{\mu}_i$ since the quantization noise $\eta_{\mathcal{Q}_\kappa}$ is zero-mean. For the sake of simplification, it is assumed that the quantization step is unitary, i.e. $\kappa = 1$. Now let denote the generalized noise model (4.42) by the function f_{gen} : $\tilde{\sigma}_i^2 = f_{\text{gen}}(\tilde{\mu}_i; \tilde{a}, \tilde{b}, \gamma)$.

Since this generalized noise model also accounts for heteroscedasticity of noise, it is more appropriate to characterize a natural image in TIFF format than existing non-linear models used in [135, 156]. The impact of JPEG compression is discussed in Section 4.3.4. It is shown that the generalized noise model (4.42) is also relevant for JPEG images with moderate-to-high quality factors. The generalized noise model (4.42) for natural JPEG images is illustrated in Figure 4.2.

4.3.2 Estimation of Parameters $(\tilde{a}, \tilde{b})$ in the Generalized Noise Model

The generalized noise model (4.42) is non-linear, which causes a difficulty of estimating noise model parameters. When the gamma factor γ is known in advance, an obvious approach is to invert the gamma correction for obtaining again the heteroscedastic relation (4.35), and then perform the WLS estimation as proposed in 4.2.2. Unfortunately, this approach leads to many problems in practice [163]. Firstly, the value of γ can not be known. One method for estimation of gamma factor γ blindly without calibration information or knowledge of imaging device was proposed in [50]. However the stability of this method on a large real image database is still questioned. Secondly, even when the value of γ is exactly known, the effect of the quantization Q_κ makes the inversion of the gamma correction ill-conditioned. Finally, this non-linear inversion would introduce artefacts into the signal, which prevents from obtaining a subsequent good estimation of parameters. Therefore, the goal of this section is to develop a method that estimate the model parameters by working directly on the non-linear generalized noise model (4.42).

The proposed method consists of three fundamental steps: edge detection, image segmentation, and ML estimation of parameters. Edge detection and image segmentation aim at detecting homogeneous blocks and partition the image into non-overlapping homogeneous segments in which the pixels are assumed to be i.i.d. Thus local expectations and local variances in each segment can be calculated, allowing to estimate noise model parameters simultaneously.

4.3.2.1 Edge Detection and Image Segmentation

Let $\tilde{\mathbf{Z}}$ be a two-dimensional matrix representing a natural image. Firstly, an estimation of image structure is performed using a denoising filter $\mathcal{D}$: $\tilde{\mathbf{Z}}^{\text{app}} = \mathcal{D}(\tilde{\mathbf{Z}})$ where $\tilde{\mathbf{Z}}^{\text{app}}$ is an approximate image structure. The residual image $\tilde{\mathbf{Z}}^{\text{res}}$, which is the difference between the noisy image $\tilde{\mathbf{Z}}$ and the denoised image $\tilde{\mathbf{Z}}^{\text{app}}$, is further used to estimate local noise variances. Since it is desirable that the proposed method can be further applied on JPEG images, and JPEG compression works separately on each 8×8 block, it is proposed to decompose the image $\tilde{\mathbf{Z}}$ (accordingly $\tilde{\mathbf{Z}}^{\text{app}}$ and $\tilde{\mathbf{Z}}^{\text{res}}$) into 64 vectors of pixels $\tilde{\mathbf{z}}_l = (\tilde{z}_{l,1}, \dots, \tilde{z}_{l,N_{\text{blk}}})$, where $l \in \{1, \dots, 64\}$ denotes the location index in the 8×8 grid and N_{blk} is the number of blocks. Therefore, the vector $\tilde{\mathbf{z}}_l$ contains all the pixels at the same location of the 8×8 grid and the pixels $(\tilde{z}_{1,u}, \dots, \tilde{z}_{64,u})$ are in the same block u .

For edge detection, instead of identifying pixels at which a local discontinuity exists, the approach aims to identify if a 8×8 block is homogeneous or contains an edge or discontinuity. To this end, the standard deviation of each block is calculated and compared with a threshold ϑ . The median of absolute deviations (MAD), which is considered as a robust estimator of standard deviation [164], is employed for the calculation of standard deviation

$$\text{MAD}(x_1, \dots, x_n) = \text{median}_i(|x_i - \text{median}_j(x_j)|). \quad (4.43)$$

Therefore, the standard deviation of block u is calculated in the DCT domain as follows

$$\varsigma_u = 1.4826 \cdot \text{MAD}(\text{DCT}(z_{1,u}^{\text{app}}, \dots, z_{64,u}^{\text{app}})), \quad (4.44)$$

where 1.4826 is the scaling factor to make the estimator unbiased [133]. Here, the denoised image $\tilde{\mathbf{Z}}^{\text{app}}$ is employed instead of the noisy image $\tilde{\mathbf{Z}}$ because the noise may contaminate the calculation of standard deviation. Moreover, only 63 AC coefficients are used in (4.44). The DC coefficient is excluded. The block u is selected if the standard deviation ς_u is smaller than the threshold ϑ . Hence the set of homogeneous blocks is defined by

$$S_{\text{homo}} = \{1 \leq u \leq N_{\text{blk}} : \varsigma_u \leq \vartheta\}. \quad (4.45)$$

After detecting homogeneous blocks, it is proposed to use only a sub-image $\tilde{\mathbf{z}}_l$ for partitioning into K non-overlapping segments. Each segment $\tilde{S}_k$, $k \in \{1, \dots, K\}$ is defined by

$$\tilde{S}_k = \left\{ \tilde{z}_{l,u} : \tilde{z}_{l,u}^{\text{app}} \in \left[t_k - \frac{\kappa_k}{2}, t_k + \frac{\kappa_k}{2} \right], u \in S_{\text{homo}} \right\}. \quad (4.46)$$

In other words, the dynamic range of the image is uniformly divided into K intervals of length Δ_k . The number of segments K is set to the number of quantization levels, e.g. $K = 2^\nu$ and $\kappa_k = 1$. For the sake of clarity, the pixel in each segment $\tilde{S}_k$ is now denoted as $\tilde{z}_{k,i}$, $i \in \{1, \dots, \tilde{n}_k\}$ where $\tilde{n}_k$ is the number of pixels in segment $\tilde{S}_k$. Analogously, $\tilde{z}_{k,i}^{\text{app}}$ and $\tilde{z}_{k,i}^{\text{res}}$ denote respectively its denoised value and residual value.

4.3.2.2 Maximum Likelihood Estimation

Consequently, the estimates of local expectation and local variance in each segment $\tilde{S}_k$ are given by

$$\hat{\mu}_k = \frac{1}{\tilde{n}_k} \sum_{i=1}^{\tilde{n}_k} \tilde{z}_{k,i}^{\text{app}} \quad (4.47)$$

$$\hat{\sigma}_k^2 = \frac{1}{\tilde{n}_k - 1} \sum_{i=1}^{\tilde{n}_k} (\tilde{z}_{k,i}^{\text{res}} - \bar{z}_k^{\text{res}})^2 \text{ with } \bar{z}_k^{\text{res}} = \frac{1}{\tilde{n}_k} \sum_{i=1}^{\tilde{n}_k} \tilde{z}_{k,i}^{\text{res}}. \quad (4.48)$$

Because the local estimate $\hat{\mu}_k$ is calculated as the average of all denoised value in each segment, it is assumed that its variance is negligible when the number of pixels

is large, i.e. the local mean $\hat{\mu}_k$ is very close to the true value $\tilde{\mu}_k$: $\hat{\mu}_k \cong \tilde{\mu}_k$ where $\tilde{\mu}_k$ is the expectation of all pixels in the segment $\tilde{S}_k$. Meanwhile, the variance of $\hat{\sigma}_k^2$ is more crucial and needs to be treated carefully. In virtue of Lindeberg CLT, for a very large number of pixels $\tilde{n}_k$, the local variance $\hat{\sigma}_k^2$ follows the Gaussian distribution

$$\hat{\sigma}_k^2 \sim \mathcal{N}\left(\tilde{\sigma}_k^2, \frac{2}{\tilde{n}_k} \tilde{\sigma}_k^4\right), \quad (4.49)$$

where $\tilde{\sigma}_k^2 = f_{\text{gen}}(\tilde{\mu}_k; \tilde{a}, \tilde{b}, \gamma)$ is the true variance with respect to the expectation $\tilde{\mu}_k$. The scatter-plot of couples $\{\hat{\mu}_k, \hat{\sigma}_k^2\}_{k=1}^K$ is illustrated in Figure 4.2.

The ML approach is used to fit the global parametric model $\tilde{\sigma}_k^2 = f_{\text{gen}}(\tilde{\mu}_k; \tilde{a}, \tilde{b}, \gamma)$ to the scatter-plot of couples $\{\hat{\mu}_k, \hat{\sigma}_k^2\}_{k=1}^K$. The log-likelihood function of K segments is given by

$$\mathcal{L}(\tilde{a}, \tilde{b}, \gamma) = -\frac{1}{2} \sum_{k=1}^K \left[\log \left(2\pi \frac{2}{\tilde{n}_k} f_{\text{gen}}^2(\hat{\mu}_k; \tilde{a}, \tilde{b}, \gamma) \right) + \frac{\hat{\sigma}_k^2 - f_{\text{gen}}(\hat{\mu}_k; \tilde{a}, \tilde{b}, \gamma)}{\frac{2}{\tilde{n}_k} f_{\text{gen}}^2(\hat{\mu}_k; \tilde{a}, \tilde{b}, \gamma)} \right]. \quad (4.50)$$

Here, because the true value $\tilde{\mu}_k$ is unknown in practice, it is proposed to replace $\tilde{\mu}_k$ by $\hat{\mu}_k$ in the log-likelihood function $\mathcal{L}$. The ML estimates of $(\tilde{a}, \tilde{b}, \gamma)$ are obtained by maximizing the log-likelihood function $\mathcal{L}$

$$(\hat{\tilde{a}}, \hat{\tilde{b}}, \hat{\gamma}) = \arg \max_{(\tilde{a}, \tilde{b}, \gamma)} \mathcal{L}(\tilde{a}, \tilde{b}, \gamma). \quad (4.51)$$

Because there is no closed form for ML estimates, the problem (4.51) is proposed to be solved numerically by using the Nelder-Mead method [157].

4.3.3 Application to Image Denoising

The generalized noise model could be useful in many applications. One direct application of the generalized noise model is for image denoising. The generalized noise model (4.42) is proposed to apply with the Local Linear Minimum Mean Square Error (LLMMSE) filter [130] to derive an efficient image denoising method. The LLMMSE filter is based on the non-stationary mean, non-stationary variance image model. From (4.39), the pixel $\tilde{z}$ can be decomposed as

$$\tilde{z} = y + \frac{1}{\gamma} y^{1-\gamma} \eta, \quad (4.52)$$

where y denotes the original pixel and η is the zero-mean signal-dependent noise, $\sigma_\eta^2 = \tilde{a} \mu_y^\gamma + \tilde{b}$. The index of pixel is omitted for the sake of clarity. The original pixel y involves non-stationary mean and non-stationary variance. The non-stationary mean describes the gross structure of an image and the non-stationary variance characterizes edge information of the image [130]. In the decomposition (4.52), the quantization noise is assumed to be negligible. As explained in [130], the LLMMSE filter for any signal-dependent noise model is formulated as

$$\hat{y} = \left(1 - \frac{\sigma_y^2}{\sigma_{\tilde{z}}^2}\right) \mu_{\tilde{z}} + \frac{\sigma_y^2}{\sigma_{\tilde{z}}^2} \tilde{z}, \quad (4.53)$$

where $\mu_{\tilde{z}}$ and $\sigma_{\tilde{z}}^2$ are respectively local mean and local variance of the pixel $\tilde{z}$, μ_y and σ_y^2 are respectively local mean and local variance of the original pixel y . Since the noise η is zero-mean, it follows from (4.52) that $\mu_{\tilde{z}} = \mu_y$. The LLMMSE estimate is the weighted sum of original signal mean μ_y and the noisy observation $\tilde{z}$ where the weight is determined as the ratio of the original signal variance and noise variance. For a low signal-to-noise (SNR) ratio, the LLMMSE filter puts more weight on the prior mean μ_y because the image is too noisy to make an accurate estimate of the original image. Conversely, for high SNR, this filter puts more weight on noisy observations to preserve the edge sharpness. In the LLMMSE filter, a simple technique to obtain local statistics $\mu_{\tilde{z}}$ and $\sigma_{\tilde{z}}^2$ is to calculate over a sliding window of size $(2r + 1) \times (2c + 1)$

$$\mu_{\tilde{z}}(m, n) = \frac{1}{(2r + 1)(2c + 1)} \sum_{i=m-r}^{m+r} \sum_{j=n-c}^{n+c} \tilde{z}(i, j), \quad (4.54)$$

$$\sigma_{\tilde{z}}^2(m, n) = \frac{1}{(2r + 1)(2c + 1)} \sum_{i=m-r}^{m+r} \sum_{j=n-c}^{n+c} (\tilde{z}(i, j) - \mu_{\tilde{z}}(m, n))^2. \quad (4.55)$$

Therefore, it remains to calculate the variance σ_y^2 . From (4.52), the variance $\sigma_{\tilde{z}}^2$ can be given by

$$\sigma_{\tilde{z}}^2 = \sigma_y^2 + \frac{1}{\gamma^2} \mathbb{E}[y^{2-2\gamma}] \sigma_{\eta}^2. \quad (4.56)$$

By using the Taylor series expansion of $y^{2-2\gamma}$ around μ_y , the expression of $\mathbb{E}[y^{2-2\gamma}]$ can be simplified as

$$\mathbb{E}[y^{2-2\gamma}] = \mu_y^{2-2\gamma} + (1 - \gamma)(1 - 2\gamma)\mu_y^{-2\gamma}\sigma_y^2. \quad (4.57)$$

Combining (4.56) and (4.57), the variance σ_y^2 is then derived as

$$\sigma_y^2 = \frac{\sigma_{\tilde{z}}^2 - \frac{1}{\gamma^2} \mu_y^{2-2\gamma} \sigma_{\eta}^2}{1 + \frac{1}{\gamma^2} (1 - \gamma)(1 - 2\gamma) \mu_y^{-2\gamma} \sigma_{\eta}^2}. \quad (4.58)$$

The LLMSSE filter (4.53) follows immediately.

4.3.4 Numerical Experiments

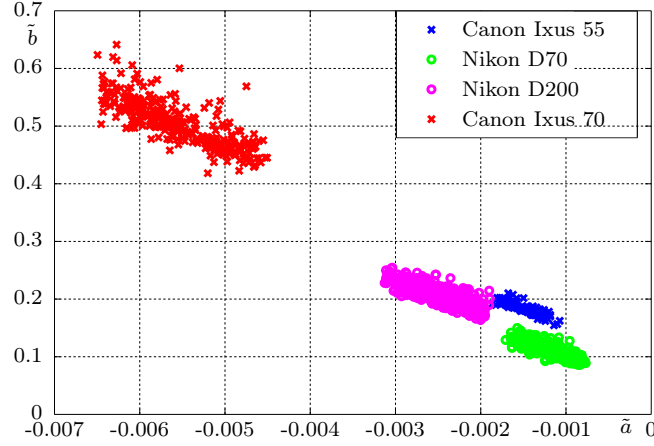
To estimate the image structure, the wavelet-based denoising filter proposed in [29, 165] is employed because of its relative accuracy and computational efficiency. Besides, the proposed method requires an appropriate threshold ϑ such that we have sufficient statistics for estimation process. Therefore, the threshold ϑ is defined as the median of absolute deviations of all residual pixels $\tilde{z}_i^{\text{res}}$

$$\vartheta = 1.4826 \cdot \text{MAD}(\tilde{z}_1^{\text{res}}, \dots, \tilde{z}_N^{\text{res}}). \quad (4.59)$$

This threshold is simple and efficient for rejecting blocks with strong edges. Besides, the sub-image used segmentation and parameter estimation corresponds to the location (4, 4) of the 8×8 grid since compression error is higher for pixels near block boundaries, and especially high at block corners [166].

	TIFF images		JPEG images ($Q = 90$)		JPEG images ($Q = 80$)		JPEG images ($Q = 70$)	
	Avg	Std Dev	Avg	Std Dev	Avg	Std Dev	Avg	Std Dev
$\tilde{a} = -0.0012$	-0.0012	$2.2e^{-4}$	-0.0012	$2.5e^{-4}$	-0.0013	$2.8e^{-4}$	-0.00095	$4.1e^{-4}$
$\tilde{b} = 0.11$	0.114	0.0097	0.117	0.0102	0.126	0.015	0.132	0.025
$\gamma = 0.8$	0.807	0.053	0.811	0.057	0.79	0.058	0.814	0.11
$\tilde{a} = -0.0025$	-0.0024	$3.5e^{-4}$	-0.0024	$4.2e^{-4}$	-0.0026	$5.4e^{-4}$	-0.0023	$5.8e^{-4}$
$\tilde{b} = 0.20$	0.196	0.0084	0.191	0.0098	0.215	0.0117	0.191	0.021
$\gamma = 0.85$	0.845	0.049	0.853	0.058	0.845	0.061	0.842	0.086

Table 4.1: Parameter estimation on synthetic images

Figure 4.3: Estimated parameters $(\tilde{a}, \tilde{b})$ on JPEG images issued from different camera models in Dresden image database.

Reference images from TID2008 database [167] are chosen to evaluate the accuracy of the proposed estimation method. The parameters $(\tilde{a}, \tilde{b}, \gamma)$, which are estimated from natural JPEG images that are acquired by *Nikon D70* and *Nikon D200* cameras (see Figure 4.2), are used to generate synthetic images according to the generalized noise model (4.42). The synthetic images are then compressed with different quality factors $\{70, 80, 90\}$. The Table 4.1 shows the statistics of estimated parameters on non-compressed TIFF images and JPEG images with different quality factors. It is noted that the estimated parameters are close to the ground truth. Moreover, the distortion caused by JPEG compression with moderate-to-high quality factors ($Q \geq 70$) seems to weakly interfere in the estimation process. Since the generalized noise model (4.42) has not been proposed yet in the literature, there is no existing method for noise parameter estimation to compare with the proposed one.

To highlight the relevance of the proposed approach, experiments are then conducted on a large image database. The Dresden database [168] that contains different camera devices, different imaged scenes, different camera settings and different environmental conditions is chosen for this experiment. All images of the database are acquired with the highest available JPEG quality setting and maximum available resolution. Figure 4.3 shows estimated parameters $(\tilde{a}, \tilde{b})$ over 1000 JPEG images of different camera models. As expected, the estimated parameters of the same camera

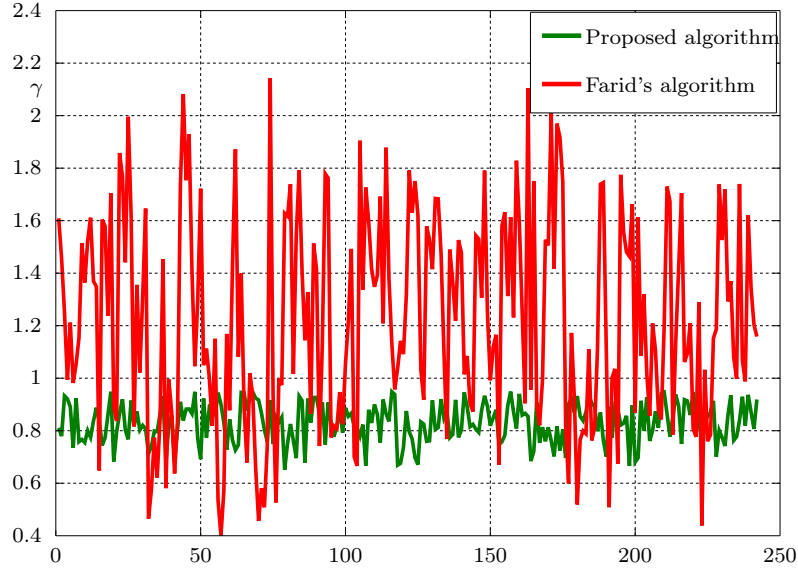


Figure 4.4: Comparison between the proposed method and Farid's for estimation of gamma factor on JPEG images issued from *Nikon D200* camera model.

	Simple Wiener filter	Extended LLMMSE filter
$\tilde{a} = -0.0012$	30.06	50.97
$\tilde{b} = 0.11$		
$\gamma = 0.8$		
$\tilde{a} = -0.0025$	30.05	48.61
$\tilde{b} = 0.20$		
$\gamma = 0.85$		

Table 4.2: PSNR of the extended LLMMSE filter

model are close to each other. Furthermore, estimated gamma values provided by the proposed method are compared with the Farid's [50]. Figure 4.4 shows estimated gamma of the two methods on the JPEG images taken from *Nikon D200* camera model. It can be noted that the variability of gamma estimated by the proposed method is considerably smaller than Farid's [50], thus the value of gamma can be estimated more efficiently.

To evaluate the denoising performance of the extended LLMMSE filter, it is proposed to apply on the synthetic non-compressed images generated from TID2008 database. The parameters $(\tilde{a}, \tilde{b}, \gamma)$ are estimated on each synthetic image and then used in the LLMMSE filter. Its averaged Peak Signal-to-Noise Ratio (PSNR) is shown in Table 4.2. It is noted that the extended LLMMSE filter provides better denoising performance.

4.4 Statistical Modeling in DCT Domain

4.4.1 Statistical Model of Quantized DCT Coefficients

This section presents a complete mathematical framework of studying the statistical distribution of DCT coefficients. The framework is based on the doubly stochastic model as describe in Section 3.3. To facilitate the study, the framework is based on the main assumption that the pixels are identically distributed within a small 8×8 block but not necessarily independent. Firstly, the model of block variance and unquantized DCT coefficients are analytically established based on that assumption. Then, it is proposed to take into account the impact of quantization in the DCT domain and provide a final model of quantized DCT coefficients.

4.4.1.1 Statistical Model of Block Variance and Unquantized DCT Coefficients

The block variance σ_{blk}^2 can be defined as

$$\sigma_{\text{blk}}^2 = \frac{1}{63} \sum_{m=0}^7 \sum_{n=0}^7 (\tilde{z}_{m,n} - \bar{\tilde{z}})^2, \quad (4.60)$$

where $\tilde{z}_{m,n}$ denotes a TIFF pixel in a 8×8 block, $0 \leq m \leq 7$, $0 \leq n \leq 7$, and $\bar{\tilde{z}}$ denotes the averaged pixel

$$\bar{\tilde{z}} = \frac{1}{64} \sum_{m=0}^7 \sum_{n=0}^7 \tilde{z}_{m,n}. \quad (4.61)$$

It can be noted that $\bar{\tilde{z}}$ is itself a random variable. Then we have

$$\tilde{z}_{m,n} - \bar{\tilde{z}} = \frac{1}{64} \sum_{i=0}^7 \sum_{j=0}^7 (\tilde{z}_{m,n} - \tilde{z}_{i,j}). \quad (4.62)$$

By invoking the CLT for correlated random variables [150], the distribution of $\tilde{z}_{m,n} - \bar{\tilde{z}}$ approaches to the Gaussian distribution with zero-mean. The fact of "Gaussianization" is important since it allows us to simplify the study of the sum of 64 random variables, of which pdf (4.36) is complicated. Because the square of a standard Gaussian random variable follows the chi-square distribution of one degree of freedom and a chi-square random variable scaled by a constant follows the Gamma distribution, it follows that

$$\frac{1}{63} (\tilde{z}_{m,n} - \bar{\tilde{z}})^2 \xrightarrow{d} \mathcal{G}\left(\frac{1}{2}, \beta_{m,n}\right), \quad (4.63)$$

where $\mathcal{G}(\cdot)$ denotes the Gamma distribution and $\beta_{m,n}$ is a scale parameter depending on the variance of $\tilde{z}_{m,n} - \bar{\tilde{z}}$.

Consequently, from (4.60) the block variance σ_{blk}^2 can be considered as a sum of correlated Gamma random variables. The exact distribution of a sum of correlated

Gamma variables was analytically established in [169]. However, this exact distribution is too complicated to establish the pdf f_I (3.17) of the DCT coefficient I . One alternative approach is to approximate the distribution of the sum of correlated Gamma variables. According to [169], the Moment-Generating Function (MGF) of σ_{blk}^2 can be expressed as

$$M_{\sigma_{\text{blk}}^2}(t) = \left[\det(\mathbf{Id}_{64} - t \cdot \mathbf{Dg} \cdot \mathbf{Co}) \right]^{-\frac{1}{2}}, \quad (4.64)$$

where $\det(\cdot)$ denotes the determinant operator, $\mathbf{Id}_{64}$ is the 64×64 identity matrix, $\mathbf{Dg}$ is the 64×64 diagonal matrix with the entries $\{\beta_{m,n}\}$, and $\mathbf{Co}$ is the 64×64 covariance matrix defined by

$$\mathbf{Co} = \begin{pmatrix} 1 & \sqrt{\rho_{1,2}} & \cdots & \sqrt{\rho_{1,64}} \\ \sqrt{\rho_{2,1}} & 1 & \cdot & \sqrt{\rho_{2,64}} \\ \cdot & \cdot & \cdot & \cdot \\ \sqrt{\rho_{64,1}} & \cdots & \cdots & 1 \end{pmatrix}. \quad (4.65)$$

Here, $\rho_{i,j}$ is the correlation coefficient between two pixels within a block. Let $\{\lambda_i\}_{i=1}^{64}$ be the eigenvalues of the matrix $\mathbf{Dg} \cdot \mathbf{Co}$, the MGF $M_{\sigma_{\text{blk}}^2}(t)$ can be rewritten as

$$M_{\sigma_{\text{blk}}^2}(t) = \prod_{i=1}^{64} (1 - t\lambda_i)^{-\frac{1}{2}}. \quad (4.66)$$

This MGF has a similar form of the one of the sum of independent Gamma variables $\mathcal{G}(\frac{1}{2}, \lambda_i)$, $i = \{1, \dots, 64\}$.

It is proposed to approximate the distribution of the sum of independent Gamma variables $\mathcal{G}(\frac{1}{2}, \lambda_i)$ by a Gamma distribution $\mathcal{G}(\alpha, \beta)$ using the moment matching method. By matching the first two moments of two distributions, the parameters (α, β) are given by

$$\alpha = \frac{(\sum_{i=1}^{64} \lambda_i)^2}{2 \sum_{i=1}^{64} \lambda_i^2} \quad \text{and} \quad \beta = \frac{\sum_{i=1}^{64} \lambda_i^2}{\sum_{i=1}^{64} \lambda_i}. \quad (4.67)$$

As a result, the block variance σ_{blk}^2 can be approximately modeled by the Gamma distribution $\mathcal{G}(\alpha, \beta)$

$$f_{\sigma_{\text{blk}}^2}(t) = \frac{t^{\alpha-1}}{\beta^\alpha \Gamma(\alpha)} \exp\left(-\frac{t}{\beta}\right), \quad (4.68)$$

where α is a positive shape parameter, β is a positive scale parameter, and $\Gamma(\cdot)$ denotes the gamma function.

Finally, it follows from (3.17), (3.18) and (4.68) that the pdf of the unquantized DCT coefficient I is given by

$$f_I(x) = \frac{1}{\sqrt{2\pi}\beta^\alpha \Gamma(\alpha)} \int_0^\infty \exp\left(-\frac{t}{\beta} - \frac{x^2}{2t}\right) t^{\alpha-\frac{3}{2}} dt. \quad (4.69)$$

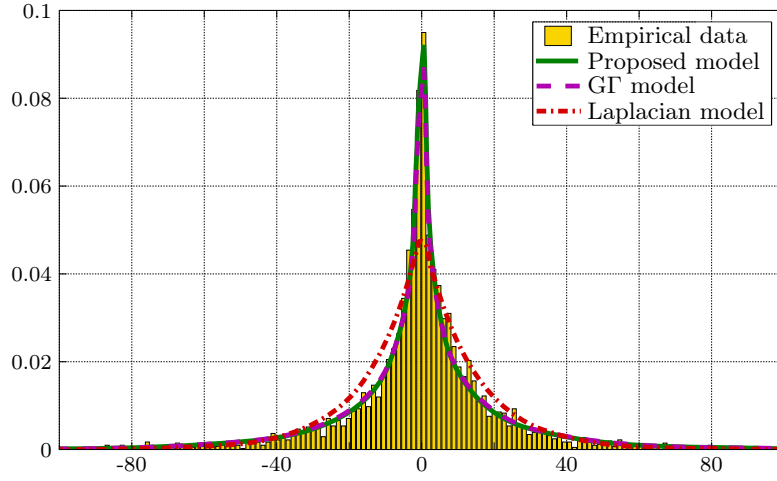


Figure 4.5: Comparison between the Laplacian, GT and proposed model of DCT coefficients.

From [170], the integral representation of the modified Bessel $K_\nu(\cdot)$ yields to

$$f_I(x) = \sqrt{\frac{2}{\pi}} \frac{\left(|x|\sqrt{\frac{\beta}{2}}\right)^{\alpha-\frac{1}{2}}}{\beta^\alpha \Gamma(\alpha)} K_{\alpha-\frac{1}{2}}\left(|x|\sqrt{\frac{2}{\beta}}\right). \quad (4.70)$$

The Figure 4.5 illustrates the empirical distribution of the third DCT coefficient, extracted from an image in the BOSS Base [171], compared with the proposed model of unquantized DCT coefficients, Laplacian model and GT model.

One technique to study the shape of a statistical distribution is to determine the kurtosis coefficient. Since the distribution of I is symmetric, the odd moments vanish. Based on the law of total expectation, the variance of the unquantized DCT coefficient I is given by

$$\text{Var}_I[I] = \mathbb{E}_I[I^2] = \mathbb{E}_{\sigma_{\text{blk}}^2} \left[\mathbb{E}_{I|\sigma_{\text{blk}}^2} [I^2 | \sigma_{\text{blk}}^2] \right] = \mathbb{E}_{\sigma_{\text{blk}}^2} [\sigma_{\text{blk}}^2] = \alpha\beta. \quad (4.71)$$

Then, the kurtosis coefficient of the proposed statistical distribution is defined by

$$\gamma_2 = \frac{\mathbb{E}_I[I^4]}{\text{Var}_I^2[I]} = \frac{\mathbb{E}_{\sigma_{\text{blk}}^2} [3\sigma_{\text{blk}}^4]}{\mathbb{E}_{\sigma_{\text{blk}}^2}^2 [\sigma_{\text{blk}}^2]} = 3 \frac{\alpha\beta^2(\alpha+1)}{\alpha^2\beta^2} = 3 \left(1 + \frac{1}{\alpha}\right). \quad (4.72)$$

It can be noted that the proposed model includes the Laplacian and Gaussian as special cases. In fact, as $\alpha \rightarrow \infty$, then $\gamma_2 \rightarrow 3$. The DCT coefficient I tends to be distributed as Gaussian variable. Similarly, as $\alpha = 1$, $\gamma_2 = 6$, then the Gamma distribution of block variance reduces back to the exponential distribution, the Laplacian model for AC coefficient I can be obtained [147]. Consequently, the proposed model of I outperforms the Laplacian, yet at the expense of more complex expressions and extra computational cost.

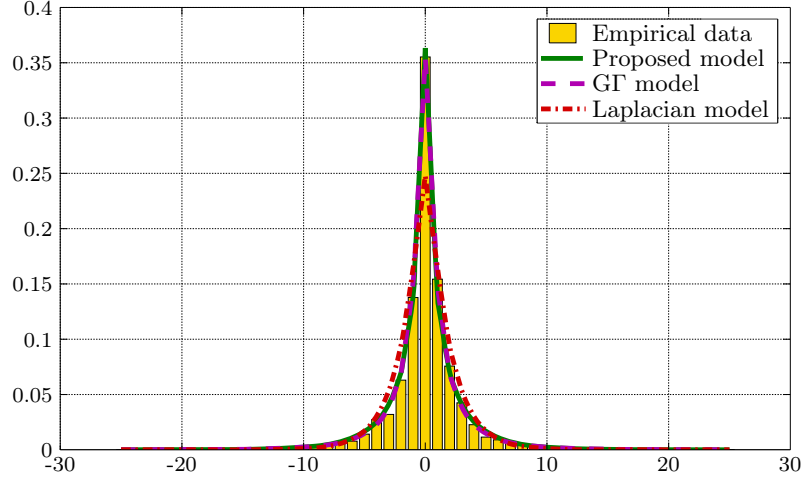


Figure 4.6: Comparison between the quantized Laplacian, quantized GF and proposed model for quantized AC coefficient.

4.4.1.2 Impact of Quantization

Let $p_D(l)$, $l \in \mathbb{Z}$, be the probability mass function (pmf) of the quantized DCT coefficient D with the corresponding quantization step Δ . The uniform quantization operation with step Δ can be written as follows

$$D = l \iff I \in \left[\Delta \left(l - \frac{1}{2} \right), \Delta \left(l + \frac{1}{2} \right) \right]. \quad (4.73)$$

Therefore, the pmf $p_D(l)$ is defined by

$$p_D(l) = \mathbb{P}[D = l] = \int_{\Delta(l-\frac{1}{2})}^{\Delta(l+\frac{1}{2})} f_I(x) dx. \quad (4.74)$$

Because the pmf $p_D(l)$ is symmetric, it is sufficient to consider $l \geq 0$. Let define the function $G(l)$ as

$$G(l) = \int_0^{\Delta(l+\frac{1}{2})} f_I(x) dx \quad \forall l \in \mathbb{Z}^+. \quad (4.75)$$

By changing the variable $x = \Delta(l + \frac{1}{2}) \cdot t$, a direct calculation from (4.70) and (4.75) yields to

$$G(l) = \sqrt{\frac{2}{\pi}} \frac{\left(\sqrt{\frac{\beta}{2}} \right)^{\alpha-\frac{1}{2}} \left(\Delta(l + \frac{1}{2}) \right)^{\alpha+\frac{1}{2}}}{\beta^\alpha \Gamma(\alpha)} \times \int_0^1 t^{\alpha-\frac{1}{2}} K_{\alpha-\frac{1}{2}} \left[t \cdot \Delta \left(l + \frac{1}{2} \right) \sqrt{\frac{2}{\beta}} \right] dt. \quad (4.76)$$

It follows from [170] that:

$$G(l) = \frac{1}{2} g(l) \left[K_{\alpha-\frac{1}{2}}(g(l)) \mathbf{L}_{\alpha-\frac{3}{2}}(g(l)) + K_{\alpha-\frac{3}{2}}(g(l)) \mathbf{L}_{\alpha-\frac{1}{2}}(g(l)) \right], \quad (4.77)$$

where $g(l) = \Delta(l + \frac{1}{2})\sqrt{\frac{2}{\beta}}$ and $\mathbf{L}_\nu(\cdot)$ is the modified Struve function. Finally, the pmf $P_D(l)$ is given by

$$p_D(l) = \begin{cases} G(|l|) - G(|l| - 1) & \forall l \in \mathbb{Z}_* \\ 2G(0) & l = 0. \end{cases} \quad (4.78)$$

The Figure 4.6 illustrates the empirical data compared with the proposed model of quantized AC coefficients, quantized Laplacian and quantized GF model.

4.4.2 Estimation of Parameters (α, β) from Unquantized DCT Coefficients

The above mathematical framework does not explain the difference in scale of the distributions across the DCT coefficients. In fact, because of heterogeneity in a natural image, the energy tends to be more concentrated in the lower frequency than in the higher frequency. Consequently, DCT coefficients at different frequencies do not share the same parameters (α, β) . The estimation of parameters (α, β) should be performed on each frequency separately.

For the sake of clarity, the DCT coefficients are arranged into 64 vectors of coefficients according to the zig-zag order. Let $\mathbf{I}_k = (I_{k,1}, \dots, I_{k,N_{\text{blk}}})$, $k \in \{1, \dots, 64\}$, be the vector of length N_{blk} representing the unquantized DCT coefficient at the frequency k and the coefficients $(I_{1,u}, \dots, I_{64,u})$ are in the same block u . The parameters characterizing the distribution of unquantized DCT coefficient $\mathbf{I}_k$ are now denoted by (α_k, β_k) . Analogously, quantized DCT coefficients at the frequency k are denoted by $\mathbf{D}_k = (D_{k,1}, \dots, D_{k,N_{\text{blk}}})$ and Δ_k denotes the corresponding quantization step.

This section proposes to estimate the parameters (α_k, β_k) following the ML approach from unquantized DCT coefficients $\mathbf{I}_k$. By definition, ML estimates of the parameters (α_k, β_k) are defined as the solution of the maximization problem

$$(\hat{\alpha}_k^{\text{ML}}, \hat{\beta}_k^{\text{ML}}) = \arg \max_{(\alpha_k, \beta_k)} \sum_{u=1}^{N_{\text{blk}}} \log f_{\mathbf{I}_k}(I_{k,u}). \quad (4.79)$$

It can be noted that the likelihood function is differentiable but computing the derivative seems complicated. Since there is no closed form for ML estimates, it is proposed to resolve the maximization problem (4.79) numerically by using the Nelder-Mead method [157]. This method requires a starting point. This is accomplished by estimating the parameters (α_k, β_k) following the Method of Moments (MM). It follows from (4.71) and (4.72) that

$$\beta_k = \frac{\mathbb{E}[I_k^4]}{3\mathbb{E}[I_k^2]} - \mathbb{E}[I_k^2] \quad \text{and} \quad \alpha_k = \frac{\mathbb{E}[I_k^2]}{\beta_k}. \quad (4.80)$$

Therefore, MM estimates of the parameters (α_k, β_k) can be given by

$$\hat{\beta}_k^{\text{MM}} = \frac{\bar{I}_{k,4}}{3\bar{I}_{k,2}} - \bar{I}_{k,2} \quad \text{and} \quad \hat{\alpha}_k^{\text{MM}} = \frac{\bar{I}_{k,2}}{\hat{\beta}_k^{\text{MM}}}, \quad (4.81)$$

where $\bar{I}_{k,2}$ and $\bar{I}_{k,4}$ are the empirical second and fourth moments of $\mathbf{I}_k$

$$\bar{I}_{k,2} = \frac{1}{N_{\text{blk}}} \sum_{u=1}^{N_{\text{blk}}} I_{k,u}^2 \quad \text{and} \quad \bar{I}_{k,4} = \frac{1}{N_{\text{blk}}} \sum_{u=1}^{N_{\text{blk}}} I_{k,u}^4. \quad (4.82)$$

4.4.3 Estimation of Parameters (α, β) from Quantized DCT Coefficients

Due to the quantization in the DCT domain, the DCT coefficients of the original image can not be perfectly recovered. For a practical use, the parameters (α_k, β_k) need to be estimated from the quantized DCT coefficients $\mathbf{D}_k$. It is proposed to rely on the same approach that has been previously proposed for estimation of the parameters (α_k, β_k) from unquantized DCT Coefficients, i.e. ML estimates of the parameters (α_k, β_k) are numerically provided using the Nelder-Mead method and the MM estimates are taken as initial solution.¹ Therefore, it is sufficient to determine the MM estimates of the parameters (α_k, β_k) from quantized DCT coefficients $\mathbf{D}_k$.

According to the theory of quantization [162], the effect of uniform quantization can be modeled by an additive noise that is uniformly distributed and uncorrelated with the input signal. The quantized DCT coefficient D_k can be given by

$$D_k = \frac{I_k}{\Delta_k} + \eta_Q, \quad (4.83)$$

where η_Q denotes the quantization noise, $\eta_Q \sim \mathcal{U}\left[-\frac{1}{2}, \frac{1}{2}\right]$, and $\mathcal{U}$ stands for the uniform distribution. Based on the definitions of the mathematical expectation, the second and fourth moments of D_k are given by

$$\mathbb{E}[D_k^2] = \frac{1}{\Delta_k^2} \mathbb{E}[I_k^2] + \mathbb{E}[\eta_Q^2] = \frac{\alpha_k \beta_k}{\Delta_k^2} + \frac{1}{12} \quad (4.84)$$

$$\mathbb{E}[D_k^4] = \frac{1}{\Delta_k^4} \mathbb{E}[I_k^4] + \frac{6}{\Delta_k^2} \mathbb{E}[I_k^2] \mathbb{E}[\eta_Q^2] + \mathbb{E}[\eta_Q^4] = \frac{3}{\Delta_k^4} \alpha_k \beta_k^2 (\alpha_k + 1) + \frac{1}{2\Delta_k^2} \alpha_k \beta_k + \frac{1}{80}. \quad (4.85)$$

It follows that the parameters (α_k, β_k) can be expressed as

$$\alpha_k = \frac{\left(\mathbb{E}[D_k^2] - \frac{1}{12}\right)^2}{\frac{1}{3}\mathbb{E}[D_k^4] - \mathbb{E}^2[D_k^2] + \frac{1}{360}} \quad \text{and} \quad \beta_k = \frac{\Delta_k^2 \left(\mathbb{E}[D_k^2] - \frac{1}{12}\right)}{\alpha_k}. \quad (4.86)$$

The MM estimates of (α_k, β_k) are then derived as

$$\hat{\alpha}_k^{\text{MM}} = \frac{\left(\bar{D}_{k,2} - \frac{1}{12}\right)^2}{\frac{1}{3}\bar{D}_{k,4} - \bar{D}_{k,2}^2 + \frac{1}{360}} \quad \text{and} \quad \hat{\beta}_k^{\text{MM}} = \frac{\Delta_k^2 \left(\bar{D}_{k,2} - \frac{1}{12}\right)}{\hat{\alpha}_k^{\text{MM}}}, \quad (4.87)$$

¹The algorithm of estimation of parameters (α, β) has been published on <http://remi.cogranne.pagesperso-orange.fr/>.

	Barbara				Boat				Lena			
	Lap	GG	GF	Prpsd Mod	Lap	GG	GF	Prpsd Mod	Lap	GG	GF	Prpsd Mod
V_{01}	32	4	3	2	157	75	61	26	70	34	25	8
V_{10}	213	70	67	39	57	60	52	51	1087	8	6	7
V_{20}	4665	129	129	127	160	226	78	61	5496	365	353	332
V_{11}	177	33	33	12	180	241	225	224	680	277	233	202
V_{02}	83	8	7	5	4097	230	58	28	623	54	47	41
V_{03}	1185	301	295	281	5573	690	683	566	21113	1537	1514	1290
V_{12}	4046	302	296	289	831	351	349	289	4198	175	170	156
V_{21}	644	161	158	161	176	49	8	7	6808	336	336	286
V_{03}	2456	655	635	629	2120	96	94	86	7289	406	241	234

Table 4.3: χ^2 test statistics of Laplacian, GG, GF, and proposed model for the first 9 quantized coefficients of 3 testing standard images.

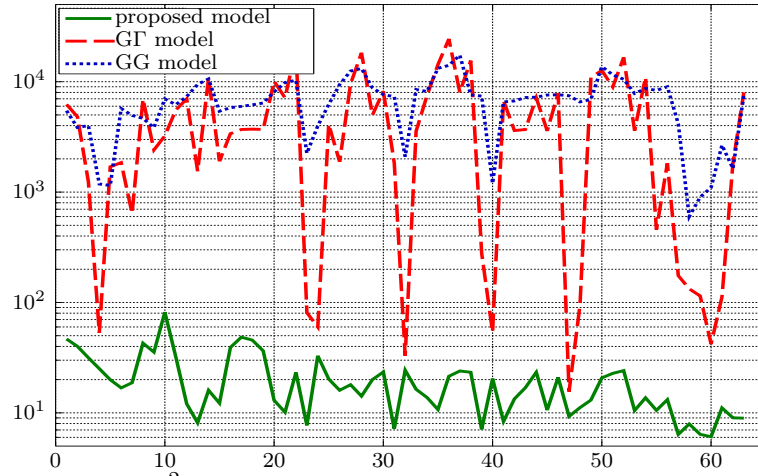


Figure 4.7: Averaged χ^2 test statistics of GG, GF and proposed model for 63 quantized AC coefficients

where $\overline{D}_{k,2}$ and $\overline{D}_{k,4}$ are the empirical second and fourth moments of D_k

$$\overline{D}_{k,2} = \frac{1}{N_{\text{blk}}} \sum_{u=1}^{N_{\text{blk}}} D_{k,u}^2 \quad \overline{D}_{k,4} = \frac{1}{N_{\text{blk}}} \sum_{u=1}^{N_{\text{blk}}} D_{k,u}^4. \quad (4.88)$$

4.4.4 Numerical Experiments

Although many models have been proposed for DCT coefficients, Laplacian and Generalized Gaussian (GG) model remain popular choice in several applications. Besides, it has been recently reported in [146] that the Generalized Gamma (GF) model outperforms the GG model and Laplacian model. To emphasize the relevance of the proposed model, experiments are conducted to compare it with the Laplacian, GG, and GF model. Firstly, the comparison is performed on some standard images: lena, boat, and barbara. These three images are converted to JPEG format using *imagemagick* with quality factor of 70. The χ^2 GOF test is employed for the comparison. The model whose the χ^2 value is smaller is more relevant to characterize the distribution of DCT coefficients. Table 4.3 shows the χ^2 test statistics of all models for AC coefficients in low frequencies. Obviously, the proposed model outperforms quantized Laplacian, quantized GG, and quantized GF models. The

Laplacian model is a special case of the GG, GF and the proposed model, thus it is considerably less accurate.

Furthermore, experiment are conducted on 10000 images in the Dresden image database [168]. Conducting experiments on such a large real database rather than on some standard images allows to evaluate the robustness of the proposed model to a wide variety of images. The averaged χ^2 GOF test statistics of all models for 63 AC coefficients is illustrated in Figure 4.7. These results obviously show the better accuracy of the proposed model compared with GG and GF model.

Remark 4.1. It can be noted that the whole image processing pipeline that goes from the RAW image acquisition to JPEG compression is a complicated process. It is too difficult to take into account all operations performed in the pipeline to establish a model of DCT coefficients that reflects the reality perfectly. To simplify the study, our mathematical framework is mainly based on the assumption that pixels are identically distributed within a 8×8 block, i.e. the structure of block is assumed to be relatively smooth. Although this assumption may not hold for all blocks in practice, the proposed model outperforms other existing models including the state-of-the-art GF model. The accuracy of the proposed model is also highlighted when applying in the steganalysis of Jsteg algorithm (see more results in Chapter 9). The proposed model of DCT coefficient is a trade-off between reality and exploitability.

The above limitation could be solved by taking into account the block structure. In our framework, all the pixels in the block $\tilde{\mathbf{z}}_{\text{blk}} = \{\tilde{z}_1, \dots, \tilde{z}_n\}$ with $n = 64$ are assumed to have the same expectation, say $\mu_{\tilde{z}_i} = t$ for all i , or equivalently

$$\boldsymbol{\mu}_{\text{blk}} = \begin{pmatrix} \mu_{\tilde{z}_1} \\ \vdots \\ \mu_{\tilde{z}_n} \end{pmatrix} = t \cdot \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}. \quad (4.89)$$

The block structure can be taken into account by using the parametric model $\boldsymbol{\mu}_{\text{blk}} = \mathbf{H}_{\text{blk}} \times \mathbf{t}$ where the $n \times q$ full-rank matrix $\mathbf{H}_{\text{blk}}$ ($q < n$) has the form [172]

$$\mathbf{H}_{\text{blk}} = \begin{pmatrix} 1 & \frac{1}{n} & \dots & \left(\frac{1}{n}\right)^{q-1} \\ 1 & \frac{2}{n} & \dots & \left(\frac{2}{n}\right)^{q-1} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & \frac{n-1}{n} & \dots & \left(\frac{n-1}{n}\right)^{q-1} \\ 1 & 1 & 1 & 1 \end{pmatrix}, \quad (4.90)$$

and $\mathbf{t} \in \mathbb{R}^q$ is an unknown parameter vector. The simple Least Squares method can give estimates of $\boldsymbol{\mu}_{\text{blk}}$ from the random vector $\tilde{\mathbf{z}}_{\text{blk}}$

$$\hat{\boldsymbol{\mu}}_{\text{blk}} = \mathbf{H}_{\text{blk}}(\mathbf{H}_{\text{blk}}^T \mathbf{H}_{\text{blk}})^{-1} \mathbf{H}_{\text{blk}}^T \tilde{\mathbf{z}}_{\text{blk}}. \quad (4.91)$$

Another limitation is that our framework only considers the first-order statistics of DCT coefficients and neglects the fact that DCT coefficients still exhibit intra-block correlation and inter-block correlation [151]. These correlations can be captured by

co-occurrence matrix [152] or adjacency matrix [126], which can lead to a generalized model of DCT coefficients. These considerations could be the subject of future researches.

4.5 Conclusion

This chapter establishes a statistical framework of modeling a natural image from RAW format to JPEG format, both in spatial domain and DCT domain. A methodology for estimation of model parameters from a single image is developed. The proposed models are more relevant to characterize a natural image than other existing models in the literature. Numerical results on a large image database show the relevance of the proposed approaches. Since a natural image can be accurately modeled and the model parameters can be efficiently estimated, this chapter builds up a solid foundation for applications in the field of digital image forensics. All these works have been valorized in our publications [134, 153, 154].

Part III

Camera Model Identification in Hypothesis Testing Framework

Camera Model Identification Based on the Heteroscedastic Noise Model

Contents

5.1	Introduction	77
5.2	Camera Fingerprint	78
5.3	Optimal Detector for Camera Model Identification Problem	80
5.3.1	Hypothesis Testing Formulation	80
5.3.2	LRT for Two Simple Hypotheses	80
5.4	GLRT with Unknown Image Parameters	83
5.5	GLRT with Unknown Image and Camera Parameters	86
5.6	Numerical Experiments	89
5.6.1	Detection Performance on Simulated Database	89
5.6.2	Detection Performance on Two <i>Nikon D70</i> and <i>Nikon D200</i> Camera Models	90
5.6.3	Detection Performance on a Large Image Database	91
5.7	Conclusion	94
5.8	Appendix	95
5.8.1	Expectation and Variance of the GLR $\bar{\Lambda}_{\text{het}}(z_{k,i}^{\text{wapp}})$ under Hy- pothesis $\mathcal{H}_j$	95
5.8.2	Expectation and Variance of the GLR $\tilde{\Lambda}_{\text{het}}(z_{k,i}^{\text{wapp}})$ under Hy- pothesis $\mathcal{H}_j$	96
5.8.2.1	Statistical Properties of ζ_k	96
5.8.2.2	Statistical Properties of ϱ_k	97
5.8.2.3	Statistical Properties of $\tilde{\Lambda}_{\text{het}}(z_{k,i}^{\text{wapp}})$	98

5.1 Introduction

The goal of this chapter is to design a statistical test within hypothesis testing framework for camera model identification from RAW images based on the heteroscedastic noise model (4.1). The main contributions are the following:

- The approach is based on the heteroscedastic noise model which describes more accurately natural images in RAW format. The parameters (a, b) are proposed to be exploited as camera fingerprint for camera model identification.
- Stating the camera model identification problem into hypothesis testing framework, this chapter studies an optimal detector given by the LRT in an ideal context where all model parameters are known in advance. This optimal detector serves as an upper-bound of any statistical test for the camera model identification problem based on the heteroscedastic noise model.
- In the practical context, the model parameters are unknown. Two GLRTs are designed to deal with the difficulty of unknown parameters. The statistical performance of these GLRTs is analytically established. Moreover, the proposed GLRTs allow the guaranteeing of a prescribed false-alarm rate and the setting of decision threshold independently of the image content. Numerical experiments also show that the loss of power of GLRTs compared with the LRT is negligible.

In general, the fact of relying on the heteroscedastic noise model is to exploit the difference in noise characteristics from different sources. Previously, noise characteristics have been used only empirically in digital image forensics, e.g. for image forgery detection [173, 174]. This chapter goes further by using a parametric noise model and handling the camera model identification problem within an analytic framework. The work in this chapter has been published in [134].

The chapter is organized as follows. Section 5.2 studies properties of the parameters (a, b) that are further exploited as camera fingerprint for camera model identification. Section 5.3 formally states the camera model identification problem in the framework of hypothesis testing theory and studies an optimal detector assuming that all model parameters are known. Section 5.4 designs a GLRT to address the difficulty of unknown image parameters. Section 5.5 addresses the context where the image parameters and the camera parameters are all unknown. Section 5.6 presents numerical results of the proposed tests on simulated and real natural RAW images. Section 5.7 discusses strengths and limitations of the proposed approach and concludes the chapter.

5.2 Camera Fingerprint

In Section 3.2.1, the heteroscedastic noise model has been analyzed by studying the RAW image acquisition and taking into account various noise sources introduced in this stage. This model can accurately capture statistics of RAW images acquired by a digital camera. Moreover, the parameters (a, b) of the heteroscedastic noise model exhibit some meaningful properties. From the relation (3.7), it can be noted that the parameters (a, b) mainly depend on the ISO sensitivity. If the digital imaging sensor does not add a pedestal parameter p_0 , the parameter a is proportional to ISO sensitivity while the parameter b is proportional to its square. Other camera

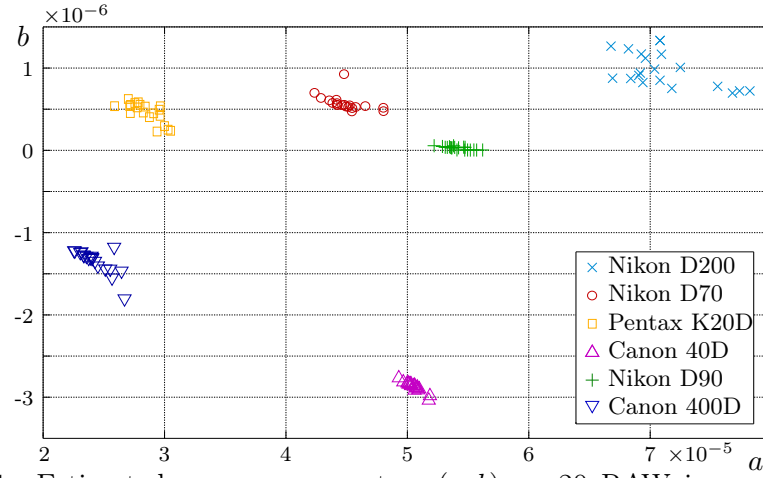


Figure 5.1: Estimated camera parameters (a, b) on 20 RAW images of different camera model with ISO 200 and different camera settings.

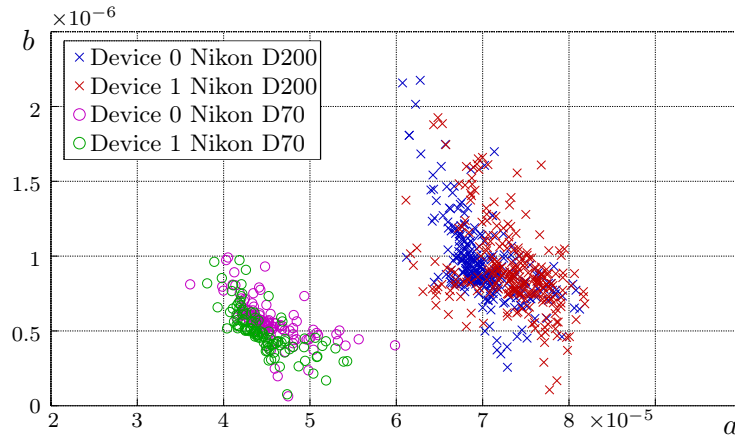


Figure 5.2: Estimated camera parameters (a, b) of different devices per camera model with ISO 200 and different camera settings.

settings such as integration time, shutter speed or focal length can also affect the parameters (a, b) . However, those effects are very small compared with the effect of ISO sensitivity. For a fixed ISO sensitivity, the parameters (a, b) are discriminative for different camera models, see Figure 4.1 and Figure 5.1. Besides, the parameters (a, b) are similar for different devices of the same model because of the same sensor characteristics, see Figure 5.2. Therefore, the parameters (a, b) of the heteroscedastic noise model can be exploited as camera fingerprint for camera model identification.

5.3 Optimal Detector for Camera Model Identification Problem

5.3.1 Hypothesis Testing Formulation

Let analyze two camera models $\mathcal{S}_0$ and $\mathcal{S}_1$. Each camera model $\mathcal{S}_j$, $j \in \{0, 1\}$, is characterized by two parameters (a_j, b_j) for a fixed ISO sensitivity. For obvious reasons, it is assumed that $(a_0, b_0) \neq (a_1, b_1)$. In a binary hypothesis testing framework as described in Definition 2.1, the inspected RAW image $\mathbf{Z} = (z_1, \dots, z_N)^T$ is either acquired by camera model $\mathcal{S}_0$ or camera model $\mathcal{S}_1$. The goal of the test is to decide between two hypotheses defined by

$$\begin{cases} \mathcal{H}_0 &= \left\{ z_i \sim \mathcal{N}(\mu_i, \sigma_{i,0}^2), \forall i \in \{1, \dots, N\} \right\} \\ \mathcal{H}_1 &= \left\{ z_i \sim \mathcal{N}(\mu_i, \sigma_{i,1}^2), \forall i \in \{1, \dots, N\} \right\}, \end{cases} \quad (5.1)$$

where μ_i and $\sigma_{i,j}^2$ respectively denote the expectation and variance of the RAW pixel z_i under hypothesis $\mathcal{H}_j$, $\sigma_{i,j}^2 = a_j \mu_i + b_j$. The parameters μ_i represent the image content in the absence of noise, which do not contain information about the camera model. They can be considered as nuisance parameters in the camera model identification problem (5.1). Moreover, they are unknown in practice. As previously explained in Chapter A, we focus on designing a test that allows to guarantee a prescribed false-alarm probability. Hence, let

$$\mathcal{K}_{\alpha_0} = \left\{ \delta : \sup_{(\boldsymbol{\mu}, a_0, b_0)} \mathbb{P}_{\mathcal{H}_0} [\delta(\mathbf{Z}) = \mathcal{H}_1] \leq \alpha_0 \right\}$$

be the class of tests whose false alarm probability is upper-bounded by the rate α_0 where $\boldsymbol{\mu} = (\mu_1, \dots, \mu_N)$ is the mean vector. Among all the tests in the class $\mathcal{K}_{\alpha_0}$, it is aimed at finding a test δ which maximizes the power function. The problem (5.1) highlights the difficulties of the camera model identification because the parameters (μ_i, a_j, b_j) are unknown in practice. It is assumed that the camera model $\mathcal{S}_0$ is available, thus forensic analysts can have access to its characteristics, or its fingerprints, i.e. the parameters (a_0, b_0) can be known. Meanwhile, the camera parameters (a_1, b_1) could be known or unknown. The main goal of this chapter is to study the LRT and to design the GLRTs to address the difficulty of unknown parameters (μ_i, a_1, b_1) .

5.3.2 LRT for Two Simple Hypotheses

In the ideal context where all model parameters are known, in virtue of the Neyman-Pearson lemma, the MP test δ_{het}^* solving the problem (5.1) is the LRT given by the

following decision rule

$$\delta_{\text{het}}^*(\mathbf{Z}) = \begin{cases} \mathcal{H}_0 & \text{if } \Lambda_{\text{het}}(\mathbf{Z}) = \sum_{i=1}^N \Lambda_{\text{het}}(z_i) < \tau_{\text{het}}^* \\ \mathcal{H}_1 & \text{if } \Lambda_{\text{het}}(\mathbf{Z}) = \sum_{i=1}^N \Lambda_{\text{het}}(z_i) \geq \tau_{\text{het}}^* \end{cases} \quad (5.2)$$

the LR $\Lambda_{\text{het}}(z_i)$ is defined by

$$\begin{aligned} \Lambda_{\text{het}}(z_i) &= \log \frac{\frac{1}{\sqrt{2\pi\sigma_{i,1}^2}} \exp \left[-\frac{(z_i - \mu_i)^2}{2\sigma_{i,1}^2} \right]}{\frac{1}{\sqrt{2\pi\sigma_{i,0}^2}} \exp \left[-\frac{(z_i - \mu_i)^2}{2\sigma_{i,0}^2} \right]} \\ &= \frac{1}{2} \left(\log(\sigma_{i,0}^2) - \log(\sigma_{i,1}^2) \right) + \frac{1}{2} \left(\frac{1}{\sigma_{i,0}^2} - \frac{1}{\sigma_{i,1}^2} \right) (z_i - \mu_i)^2. \end{aligned} \quad (5.3)$$

For brevity, let denote two functions $h_1(x)$ and $h_2(x)$ as

$$h_1(x) = \log(a_0x + b_0) - \log(a_1x + b_1) \quad \text{and} \quad h_2(x) = \frac{1}{a_0x + b_0} - \frac{1}{a_1x + b_1}, x \in \mathbb{R}_+,$$

such that

$$\Lambda_{\text{het}}(z_i) = \frac{1}{2} h_1(\mu_i) + \frac{1}{2} h_2(\mu_i) (z_i - \mu_i)^2. \quad (5.4)$$

As discussed in Section A.3, it is proposed to exploit the Lindeberg CLT to determine the statistical distribution the LR $\Lambda_{\text{het}}(\mathbf{Z})$. To this end, it is necessary to calculate the first two moments of $\Lambda_{\text{het}}(z_i)$.

Proposition 5.1. *Under hypothesis $\mathcal{H}_j$, the first two moments of $\Lambda_{\text{het}}(z_i)$ are given by*

$$\mathbb{E}_{\mathcal{H}_j} [\Lambda_{\text{het}}(z_i)] = \frac{1}{2} h_1(\mu_i) + \frac{1}{2} h_2(\mu_i) \sigma_{i,j}^2 \quad (5.5)$$

$$\mathbb{V}\text{ar}_{\mathcal{H}_j} [\Lambda_{\text{het}}(z_i)] = \frac{1}{2} h_2^2(\mu_i) \sigma_{i,j}^4. \quad (5.6)$$

Proof. It follows from $z_i \sim \mathcal{N}(\mu_i, \sigma_{i,j}^2)$ under each hypothesis $\mathcal{H}_j$, $j \in \{0, 1\}$, that

$$\frac{(z_i - \mu_i)^2}{\sigma_{i,j}^2} \sim \chi_1^2.$$

Therefore, the first two moments of $(z_i - \mu_i)^2$ under $\mathcal{H}_j$ can be given by

$$\begin{aligned} \mathbb{E}_{\mathcal{H}_j} [(z_i - \mu_i)^2] &= \sigma_{i,j}^2 \\ \mathbb{V}\text{ar}_{\mathcal{H}_j} [(z_i - \mu_i)^2] &= 2\sigma_{i,j}^4. \end{aligned}$$

The proof follows immediately. □

In virtue of the Lindeberg CLT, the statistical distribution of the LR $\Lambda_{\text{het}}(\mathbf{Z})$ under hypothesis $\mathcal{H}_j$ is given by

$$\Lambda_{\text{het}}(\mathbf{Z}) \xrightarrow{d} \mathcal{N}\left(m_{\text{het}}^{(j)}, v_{\text{het}}^{(j)}\right) \quad (5.7)$$

where the expectation $m_{\text{het}}^{(j)}$ and variance $v_{\text{het}}^{(j)}$ are given by

$$m_{\text{het}}^{(j)} = \sum_{i=1}^N \left[\frac{1}{2} h_1(\mu_i) + \frac{1}{2} h_2(\mu_i) \sigma_{i,j}^2 \right] \quad (5.8)$$

$$v_{\text{het}}^{(j)} = \sum_{i=1}^N \frac{1}{2} h_2^2(\mu_i) \sigma_{i,j}^4. \quad (5.9)$$

Since a natural RAW image is heterogeneous, it is proposed to normalize the LR $\Lambda_{\text{het}}(\mathbf{Z})$ in order to set the decision threshold independently of the image content. The normalized LR is defined by

$$\Lambda_{\text{het}}^*(\mathbf{Z}) = \frac{\Lambda_{\text{het}}(\mathbf{Z}) - m_{\text{het}}^{(0)}}{\sqrt{v_{\text{het}}^{(0)}}}. \quad (5.10)$$

Consequently, it follows from (5.7) that

$$\begin{cases} \Lambda_{\text{het}}^*(\mathbf{Z}) \xrightarrow{d} \mathcal{N}(0, 1) & \text{under } \mathcal{H}_0, \\ \Lambda_{\text{het}}^*(\mathbf{Z}) \xrightarrow{d} \mathcal{N}\left(\frac{m_{\text{het}}^{(1)} - m_{\text{het}}^{(0)}}{\sqrt{v_{\text{het}}^{(0)}}}, \frac{v_{\text{het}}^{(1)}}{v_{\text{het}}^{(0)}}\right) & \text{under } \mathcal{H}_1. \end{cases} \quad (5.11)$$

The corresponding LRT δ_{het}^* is rewritten as follows

$$\delta_{\text{het}}^*(\mathbf{Z}) = \begin{cases} \mathcal{H}_0 & \text{if } \Lambda_{\text{het}}^*(\mathbf{Z}) < \tau_{\text{het}}^* \\ \mathcal{H}_1 & \text{if } \Lambda_{\text{het}}^*(\mathbf{Z}) \geq \tau_{\text{het}}^* \end{cases} \quad (5.12)$$

where the decision threshold τ_{het}^* is the solution of the equation $\mathbb{P}_{\mathcal{H}_0}[\Lambda_{\text{het}}^*(\mathbf{Z}) \geq \tau_{\text{het}}^*] = \alpha_0$ to ensure the LRT δ_{het}^* to be in the class $\mathcal{K}_{\alpha_0}$. The fact of normalizing the LR $\Lambda_{\text{het}}(\mathbf{Z})$ allows the test δ_{het}^* to be applicable to any natural RAW image since the normalized LR $\Lambda_{\text{het}}^*(\mathbf{Z})$ follows the standard Gaussian distribution under hypothesis $\mathcal{H}_0$. The decision threshold τ_{het}^* and the power function $\beta(\delta_{\text{het}}^*)$ are given in the following theorem.

Theorem 5.1. *In an ideal context where all the model parameters $(\boldsymbol{\mu}, a_j, b_j)$ are known in advance, the decision threshold and the power function of the LRT δ_{het}^* are given by*

$$\tau_{\text{het}}^* = \Phi^{-1}(1 - \alpha_0) \quad (5.13)$$

$$\beta(\delta_{\text{het}}^*) = 1 - \Phi\left(\frac{m_{\text{het}}^{(0)} - m_{\text{het}}^{(1)} + \tau_{\text{het}}^* \sqrt{v_{\text{het}}^{(0)}}}{\sqrt{v_{\text{het}}^{(1)}}}\right). \quad (5.14)$$

Proof. is given as in Example A.1. $\square$

The test power $\beta(\delta_{\text{het}}^*)$ serves as an upper-bound of any statistical test for the camera model identification problem. The test δ_{het}^* allows to warrant a prescribed false alarm rate and maximizes the detection probability. Since its statistical performance is analytically established, it can provide an analytically predictable result for any false alarm probability α_0 . Besides, it can be noted that the detection performance of the test depends on the distance between (a_0, b_0) and (a_1, b_1) . The smaller the distance between two points (a_0, b_0) and (a_1, b_1) is, the more difficult the camera model identification is.

5.4 GLRT with Unknown Image Parameters

In practice, the parameters μ_i and (a_1, b_1) are unknown, which compromises the LRT. A possible approach to deal with unknown parameters consists in eliminating them by using the invariance principle [20]. However, in the heteroscedastic noise model, the image parameter μ_i appears in both expectation and variance of the RAW pixel z_i . The invariance approach may not be applied due to a difficulty of finding a group of transformation under which the problem remains invariant. Alternatively, it is proposed to design a GLRT to deal with the difficulty of unknown parameters by replacing unknown parameters by their ML estimates in the LR $\Lambda_{\text{het}}(z_i)$ (5.3).

The GLRT designed in this section deals with the difficulty of unknown image parameters μ_i when the camera parameters (a_0, b_0) and (a_1, b_1) are still known. This GLRT can be interpreted as a closed hypothesis testing since the decision is made only between two known camera models $\mathcal{S}_0$ and $\mathcal{S}_1$.

Since a natural image is high-dimensional, the number of unknown image parameters μ_i grows with the number of pixels N . Therefore, a preprocessing stage consisting in segmenting the inspected image $\mathbf{Z}$ into K non-overlapping homogeneous segments S_k of size n_k is performed to reduce the number of unknown parameters. It is proposed to employ the segmentation technique proposed in [133] (see also Section 4.2.2). The camera model identification problem (5.1) can be rewritten as

$$\begin{cases} \mathcal{H}_0 &= \left\{ z_{k,i}^{\text{wapp}} \sim \mathcal{N}(\mu_k, \|\varphi\|_2^2 \sigma_{k,0}^2), \forall k \in \{1, \dots, K\}, \forall i \in \{1, \dots, n_k\} \right\} \\ \mathcal{H}_1 &= \left\{ z_{k,i}^{\text{wapp}} \sim \mathcal{N}(\mu_k, \|\varphi\|_2^2 \sigma_{k,1}^2), \forall k \in \{1, \dots, K\}, \forall i \in \{1, \dots, n_k\} \right\}. \end{cases} \quad (5.15)$$

with $\sigma_{k,j}^2 = a_j \mu_k + b_j$. Using the ML estimates $\hat{\mu}_k$ defined in (4.4), the GLRT $\bar{\delta}_{\text{het}}$ solving the problem (5.15) is designed as follows

$$\bar{\delta}_{\text{het}} = \begin{cases} \mathcal{H}_0 & \text{if } \bar{\Lambda}_{\text{het}}(\mathbf{Z}) = \sum_{k=1}^K \sum_{i=1}^{n_k} \bar{\Lambda}_{\text{het}}(z_{k,i}^{\text{wapp}}) < \bar{\tau}_{\text{het}} \\ \mathcal{H}_1 & \text{if } \bar{\Lambda}_{\text{het}}(\mathbf{Z}) = \sum_{k=1}^K \sum_{i=1}^{n_k} \bar{\Lambda}_{\text{het}}(z_{k,i}^{\text{wapp}}) \geq \bar{\tau}_{\text{het}} \end{cases} \quad (5.16)$$

where the GLR $\bar{\Lambda}_{\text{het}}(z_{k,i}^{\text{wapp}})$ is given by

$$\begin{aligned}\bar{\Lambda}_{\text{het}}(z_{k,i}^{\text{wapp}}) &= \frac{1}{2} \left(\log(a_0 \hat{\mu}_k + b_0) - \log(a_1 \hat{\mu}_k + b_1) \right) \\ &\quad + \frac{1}{2} \left(\frac{1}{a_0 \hat{\mu}_k + b_0} - \frac{1}{a_1 \hat{\mu}_k + b_1} \right) \frac{(z_{k,i}^{\text{wapp}} - \hat{\mu}_k)^2}{\|\varphi\|_2^2} \\ &= \frac{1}{2} h_1(\hat{\mu}_k) + \frac{1}{2} h_2(\hat{\mu}_k) \frac{(z_{k,i}^{\text{wapp}} - \hat{\mu}_k)^2}{\|\varphi\|_2^2}.\end{aligned}\quad (5.17)$$

Proposition 5.2. *Under hypothesis $\mathcal{H}_j$, the first two moments of the GLR $\bar{\Lambda}_{\text{het}}(z_{k,i}^{\text{wapp}})$ are given by*

$$\mathbb{E}_{\mathcal{H}_j} \left[\bar{\Lambda}_{\text{het}}(z_{k,i}^{\text{wapp}}) \right] = \frac{1}{2} h_1(\mu_k) + \frac{1}{2} h_2(\mu_k) \sigma_{k,j}^2 \left(1 + \frac{1}{n_k} \right) \quad (5.18)$$

$$\begin{aligned}\text{Var}_{\mathcal{H}_j} \left[\bar{\Lambda}_{\text{het}}(z_{k,i}^{\text{wapp}}) \right] &= \frac{1}{4} (h'_1(\mu_k))^2 \frac{\|\varphi\|_2^2}{n_k} \sigma_{k,j}^2 + \frac{1}{2} (h_2(\mu_k))^2 \sigma_{k,j}^4 \left(1 + \frac{1}{n_k} \right)^2 \\ &\quad + \frac{3}{4} (h'_2(\mu_k))^2 \frac{\|\varphi\|_2^2}{n_k} \sigma_{k,j}^6 \left(1 + \frac{1}{n_k} \right)^2,\end{aligned}\quad (5.19)$$

where h'_1 and h'_2 respectively denote the first derivative of the function h_1 and h_2

$$\begin{aligned}h'_1(x) &= \frac{a_0}{a_0 x + b_0} - \frac{a_1}{a_1 x + b_1} \\ h'_2(x) &= \frac{a_1}{(a_1 x + b_1)^2} - \frac{a_0}{(a_0 x + b_0)^2}.\end{aligned}$$

Proof. of Proposition 5.2 is given in Appendix 5.8.1. $\square$

By invoking the Lindeberg CLT, the statistical distribution of the GLR $\bar{\Lambda}_{\text{het}}(\mathbf{Z})$ under hypothesis $\mathcal{H}_j$ is defined as

$$\bar{\Lambda}_{\text{het}}(\mathbf{Z}) \xrightarrow{d} \mathcal{N} \left(\bar{m}_{\text{het}}^{(j)}, \bar{v}_{\text{het}}^{(j)} \right), \quad (5.20)$$

where the expectation $\bar{m}_{\text{het}}^{(j)}$ and variance $\bar{v}_{\text{het}}^{(j)}$ are given by

$$\bar{m}_{\text{het}}^{(j)} = \sum_{k=1}^K \sum_{i=1}^{n_k} \left[\frac{1}{2} h_1(\mu_k) + \frac{1}{2} h_2(\mu_k) \sigma_{k,j}^2 \left(1 + \frac{1}{n_k} \right) \right] \quad (5.21)$$

$$\begin{aligned}\bar{v}_{\text{het}}^{(j)} &= \sum_{k=1}^K \sum_{i=1}^{n_k} \left[\frac{1}{4} (h'_1(\mu_k))^2 \frac{\|\varphi\|_2^2}{n_k} \sigma_{k,j}^2 + \frac{1}{2} (h_2(\mu_k))^2 \sigma_{k,j}^4 \left(1 + \frac{1}{n_k} \right)^2 \right. \\ &\quad \left. + \frac{3}{4} (h'_2(\mu_k))^2 \frac{\|\varphi\|_2^2}{n_k} \sigma_{k,j}^6 \left(1 + \frac{1}{n_k} \right)^2 \right].\end{aligned}\quad (5.22)$$

Similarly to (5.10), the normalized GLR $\bar{\Lambda}_{\text{het}}^*(\mathbf{Z})$ is defined by

$$\bar{\Lambda}_{\text{het}}^*(\mathbf{Z}) = \frac{\bar{\Lambda}_{\text{het}}(\mathbf{Z}) - \bar{m}_{\text{het}}^{(0)}}{\sqrt{\bar{v}_{\text{het}}^{(0)}}}. \quad (5.23)$$

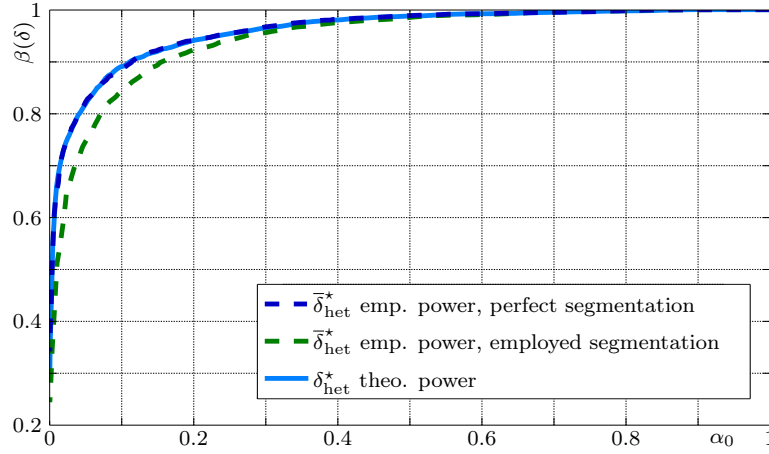


Figure 5.3: The detection performance of the GLRT $\bar{\delta}_{het}^*$ with 50 pixels selected randomly from simulated images.

However, since the image parameters μ_k are unknown in practice, the expectation $\bar{m}_{het}^{(0)}$ and variance $\bar{v}_{het}^{(0)}$ can not be defined. Therefore, it is proposed to replace μ_k by $\hat{\mu}_k$ in (5.21) and (5.22) to obtain the estimates of $\bar{m}_{het}^{(0)}$ and $\bar{v}_{het}^{(0)}$, denoted $\hat{\bar{m}}_{het}^{(0)}$ and $\hat{\bar{v}}_{het}^{(0)}$, respectively. Consequently, the normalized GLR can be given in practice as

$$\bar{\Lambda}_{het}^*(\mathbf{Z}) = \frac{\bar{\Lambda}_{het}(\mathbf{Z}) - \hat{\bar{m}}_{het}^{(0)}}{\sqrt{\hat{\bar{v}}_{het}^{(0)}}}. \quad (5.24)$$

Since the estimates $\hat{\bar{m}}_{het}^{(0)}$ and $\hat{\bar{v}}_{het}^{(0)}$ are consistent, it follows from the Slutsky's theorem [20, theorem 11.2.11] that

$$\begin{cases} \bar{\Lambda}_{het}^*(\mathbf{Z}) \xrightarrow{d} \mathcal{N}(0, 1) & \text{under } \mathcal{H}_0, \\ \bar{\Lambda}_{het}^*(\mathbf{Z}) \xrightarrow{d} \mathcal{N}\left(\frac{\bar{m}_{het}^{(1)} - \bar{m}_{het}^{(0)}}{\sqrt{\bar{v}_{het}^{(0)}}}, \frac{\bar{v}_{het}^{(1)}}{\bar{v}_{het}^{(0)}}\right) & \text{under } \mathcal{H}_1. \end{cases} \quad (5.25)$$

The corresponding GLRT $\bar{\delta}_{het}^*$ is rewritten as follows

$$\bar{\delta}_{het}^*(\mathbf{Z}) = \begin{cases} \mathcal{H}_0 & \text{if } \bar{\Lambda}_{het}^*(\mathbf{Z}) < \bar{\tau}_{het}^* \\ \mathcal{H}_1 & \text{if } \bar{\Lambda}_{het}^*(\mathbf{Z}) \geq \bar{\tau}_{het}^* \end{cases} \quad (5.26)$$

where again to ensure the GLRT $\bar{\delta}_{het}^*$ to be in the class $\mathcal{K}_{\alpha_0}$, $\bar{\tau}_{het}^*$ is the solution of the equation $\mathbb{P}_{\mathcal{H}_0}[\bar{\Lambda}_{het}^*(\mathbf{Z}) \geq \bar{\tau}_{het}^*] = \alpha_0$.

Theorem 5.2. *For testing between two known camera models $\mathcal{S}_0$ and $\mathcal{S}_1$, i.e. the parameters (a_0, b_0) and (a_1, b_1) are known in advance, the decision threshold and the*

power of the GLRT $\bar{\delta}_{\text{het}}^*$ are given by

$$\bar{\tau}_{\text{het}}^* = \Phi^{-1}(1 - \alpha_0) \quad (5.27)$$

$$\beta(\bar{\delta}_{\text{het}}^*) = 1 - \Phi \left(\frac{\bar{m}_{\text{het}}^{(0)} - \bar{m}_{\text{het}}^{(1)} + \bar{\tau}_{\text{het}}^* \sqrt{\bar{v}_{\text{het}}^{(0)}}}{\sqrt{\bar{v}_{\text{het}}^{(1)}}} \right). \quad (5.28)$$

Proof. From (5.25), the proof follows immediately. $\square$

The detection performance of the GLRT $\bar{\delta}_{\text{het}}^*$ for 50 pixels selected randomly from simulated images is illustrated in Figure 5.3. A small loss of power compared with the LRT δ_{het}^* can be noted due to the slight inaccuracy of employed segmentation. More details are given in Section 5.6.

5.5 GLRT with Unknown Image and Camera Parameters

The goal of this section is to design a GLRT to deal with the difficulty of unknown image parameters μ_k and unknown camera parameters (a_1, b_1) . This GLRT can be interpreted as open hypothesis testing when it aims to verify whether the inspected image $\mathbf{Z}$ is acquired by the camera model $\mathcal{S}_0$ or not. In other words, the inspected image $\mathbf{Z}$ is allowed to be taken from an unknown camera model.

Before designing the GLRT, the WLS estimation of camera parameters (a_1, b_1) is performed on the inspected image $\mathbf{Z}$ as proposed in Section 4.2.2. Since WLS estimates $(\hat{a}_1, \hat{b}_1)$ are asymptotically equivalent to ML estimates [158, 159], they are also relevant to be exploited in the design of the GLRT. Using the ML estimates $\hat{\mu}_k$ and WLS estimates $(\hat{a}_1, \hat{b}_1)$, the GLRT $\tilde{\delta}_{\text{het}}$ solving the problem (5.15) is given as follows

$$\tilde{\delta}_{\text{het}} = \begin{cases} \mathcal{H}_0 & \text{if } \tilde{\Lambda}_{\text{het}}(\mathbf{Z}) = \sum_{k=1}^K \sum_{i=1}^{n_k} \tilde{\Lambda}_{\text{het}}(z_{k,i}^{\text{wapp}}) < \tilde{\tau}_{\text{het}} \\ \mathcal{H}_1 & \text{if } \tilde{\Lambda}_{\text{het}}(\mathbf{Z}) = \sum_{k=1}^K \sum_{i=1}^{n_k} \tilde{\Lambda}_{\text{het}}(z_{k,i}^{\text{wapp}}) \geq \tilde{\tau}_{\text{het}} \end{cases} \quad (5.29)$$

where the GLR $\tilde{\Lambda}_{\text{het}}(z_{k,i}^{\text{wapp}})$ is defined as

$$\begin{aligned} \tilde{\Lambda}_{\text{het}}(z_{k,i}^{\text{wapp}}) = & \frac{1}{2} \left(\log(a_0 \hat{\mu}_k + b_0) - \log(\hat{a}_1 \hat{\mu}_k + \hat{b}_1) \right) \\ & + \frac{1}{2} \left(\frac{1}{a_0 \hat{\mu}_k + b_0} - \frac{1}{\hat{a}_1 \hat{\mu}_k + \hat{b}_1} \right) \frac{(z_{k,i}^{\text{wapp}} - \hat{\mu}_k)^2}{\|\varphi\|_2^2}. \end{aligned} \quad (5.30)$$

The WLS estimates $(\hat{a}_1, \hat{b}_1)$ exhibit a certain variability. Let $\sigma_{a_1}^2$, $\sigma_{b_1}^2$, $\sigma_{a_1 b_1}$ denote respectively the variance of $\hat{a}_1$, variance of $\hat{b}_1$ and covariance between $\hat{a}_1$ and $\hat{b}_1$, which have been defined in Section 4.2.2.2. The statistical properties of WLS estimates $(\hat{a}_1, \hat{b}_1)$ need to be taken into account in the establishment of statistical performance of the GLRT $\tilde{\delta}_{\text{het}}$.

Proposition 5.3. *Under hypothesis $\mathcal{H}_j$, the first two moments of the GLR $\tilde{\Lambda}_{\text{het}}(z_{k,i}^{\text{wapp}})$ are given by*

$$\mathbb{E}_{\mathcal{H}_j} [\tilde{\Lambda}_{\text{het}}(z_{k,i}^{\text{wapp}})] = \frac{1}{2}h_1(\mu_k) + \frac{1}{2}h_2(\mu_k)\sigma_{k,j}^2\left(1 + \frac{1}{n_{k,j}}\right) \quad (5.31)$$

$$\begin{aligned} \text{Var}_{\mathcal{H}_j} [\tilde{\Lambda}_{\text{het}}(z_{k,i}^{\text{wapp}})] &= \frac{1}{4} \left[\frac{a_0^2}{\sigma_{k,0}^4} \frac{\|\varphi\|_2^2}{n_k} \sigma_{k,j}^2 + \frac{\text{Var}_{\mathcal{H}_j} [\hat{a}_1 \hat{\mu}_k + \hat{b}_1]}{\sigma_{k,1}^4} \right] \\ &\quad + \frac{1}{2} (h_2(\mu_k))^2 \sigma_{k,j}^4 \left(1 + \frac{1}{n_k}\right)^2 \\ &\quad + \frac{3}{4} \sigma_{k,j}^4 \left(1 + \frac{1}{n_k}\right)^2 \left[\frac{a_0^2}{\sigma_{k,0}^8} \frac{\|\varphi\|_2^2}{n_k} \sigma_{k,j}^2 + \frac{\text{Var}_{\mathcal{H}_j} [\hat{a}_1 \hat{\mu}_k + \hat{b}_1]}{\sigma_{k,1}^8} \right]. \end{aligned} \quad (5.32)$$

with

$$\text{Var}_{\mathcal{H}_j} [\hat{a}_1 \hat{\mu}_k + \hat{b}_1] = a_1^2 \frac{\|\varphi\|_2^2}{n_k} \sigma_{k,j}^2 + \left(\mu_k^2 + \frac{\|\varphi\|_2^2}{n_k} \sigma_{k,j}^2 \right) \sigma_{a_1}^2 + 2\mu_k \sigma_{a_1 b_1} + \sigma_{b_1}^2.$$

Proof. of Proposition 5.3 is given in Appendix 5.8.2. $\square$

In virtue of Lindeberg CLT, the GLR $\tilde{\Lambda}_{\text{het}}(\mathbf{Z})$ follows the Gaussian distribution under hypothesis $\mathcal{H}_j$

$$\tilde{\Lambda}_{\text{het}}(\mathbf{Z}) \xrightarrow{d} \mathcal{N}(\tilde{m}_{\text{het}}^{(j)}, \tilde{v}_{\text{het}}^{(j)}) \quad (5.33)$$

where the expectation $\tilde{m}_{\text{het}}^{(j)}$ and variance $\tilde{v}_{\text{het}}^{(j)}$ are defined by

$$\tilde{m}_{\text{het}}^{(j)} = \sum_{k=1}^K \sum_{i=1}^{n_k} \mathbb{E}_{\mathcal{H}_j} [\tilde{\Lambda}_{\text{het}}(z_{k,i}^{\text{wapp}})] \quad (5.34)$$

$$\tilde{v}_{\text{het}}^{(j)} = \sum_{k=1}^K \sum_{i=1}^{n_k} \text{Var}_{\mathcal{H}_j} [\tilde{\Lambda}_{\text{het}}(z_{k,i}^{\text{wapp}})]. \quad (5.35)$$

with $\mathbb{E}_{\mathcal{H}_j} [\tilde{\Lambda}_{\text{het}}(z_{k,i}^{\text{wapp}})]$ and $\text{Var}_{\mathcal{H}_j} [\tilde{\Lambda}_{\text{het}}(z_{k,i}^{\text{wapp}})]$ given in Proposition 5.3.

Similarly, the normalized GLR $\tilde{\Lambda}_{\text{het}}^*(\mathbf{Z})$ is defined by

$$\tilde{\Lambda}_{\text{het}}^*(\mathbf{Z}) = \frac{\tilde{\Lambda}_{\text{het}}(\mathbf{Z}) - \hat{\tilde{m}}_{\text{het}}^{(0)}}{\sqrt{\hat{\tilde{v}}_{\text{het}}^{(0)}}} \quad (5.36)$$

where $\hat{\tilde{m}}_{\text{het}}^{(0)}$ and $\hat{\tilde{v}}_{\text{het}}^{(0)}$ are the estimates of the expectation $\tilde{m}_{\text{het}}^{(0)}$ and variance $\tilde{v}_{\text{het}}^{(0)}$ by replacing (μ_k, a_1, b_1) by $(\hat{\mu}_k, \hat{a}_1, \hat{b}_1)$ in (5.34) and (5.35). The GLRT $\tilde{\delta}_{\text{het}}^*$ based on the normalized GLR $\tilde{\Lambda}_{\text{het}}^*(\mathbf{Z})$ is given by

$$\tilde{\delta}_{\text{het}}^*(\mathbf{Z}) = \begin{cases} \mathcal{H}_0 & \text{if } \tilde{\Lambda}_{\text{het}}^*(\mathbf{Z}) < \tilde{\tau}_{\text{het}}^* \\ \mathcal{H}_1 & \text{if } \tilde{\Lambda}_{\text{het}}^*(\mathbf{Z}) \geq \tilde{\tau}_{\text{het}}^* \end{cases} \quad (5.37)$$

where the decision threshold $\tilde{\tau}_{\text{het}}^*$ is the solution of the equation $\mathbb{P}_{\mathcal{H}_0} [\tilde{\Lambda}_{\text{het}}^*(\mathbf{Z}) \geq \tilde{\tau}_{\text{het}}^*] = \alpha_0$.

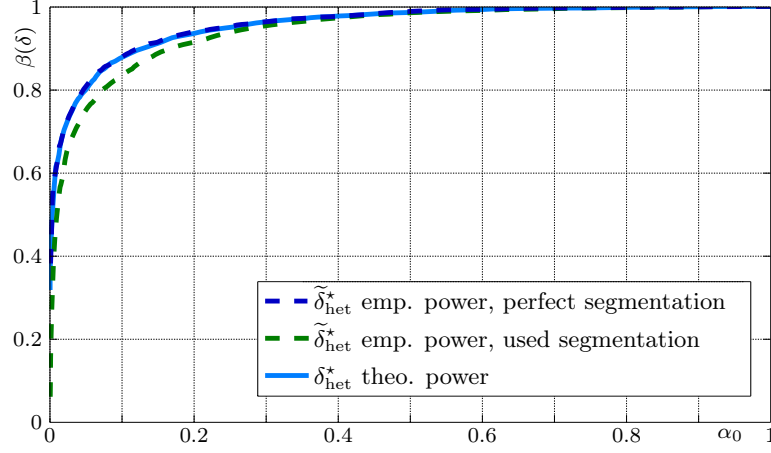


Figure 5.4: The detection performance of the GLRT $\tilde{\delta}_{\text{het}}^*$ with 50 pixels selected randomly on simulated images.

Theorem 5.3. *For testing against a known camera model $\mathcal{S}_0$, i.e. the parameters (a_0, b_0) are known, the decision threshold and the power of the GLRT $\tilde{\delta}_{\text{het}}^*$ are given by*

$$\tilde{\tau}_{\text{het}}^* = \Phi^{-1}(1 - \alpha_0) \quad (5.38)$$

$$\beta(\tilde{\delta}_{\text{het}}^*) = 1 - \Phi \left(\frac{\tilde{m}_{\text{het}}^{(0)} - \tilde{m}_{\text{het}}^{(1)} + \tilde{\tau}_{\text{het}}^* \sqrt{\tilde{v}_{\text{het}}^{(0)}}}{\sqrt{\tilde{v}_{\text{het}}^{(1)}}} \right). \quad (5.39)$$

Proof. From the Slutsky's theorem, the proof follows immediately. $\square$

The philosophy presented in the GLRT $\tilde{\delta}_{\text{het}}^*$ is quite different from the one in the LRT δ_{het}^* or the GLRT $\bar{\delta}_{\text{het}}^*$. While the LRT δ_{het}^* and the GLRT $\bar{\delta}_{\text{het}}^*$ deal with the scenario of two known camera models, the GLRT $\tilde{\delta}_{\text{het}}^*$ needs to deal with the scenario of one camera model in hand and verify whether the given image is captured by that camera model under hypothesis $\mathcal{H}_0$. The camera model under $\mathcal{H}_1$ is unknown, so the its fingerprint are also unknown. Consequently, the GLRT $\bar{\delta}_{\text{het}}^*$ can not be applied in this case. Therefore, the GLRT $\tilde{\delta}_{\text{het}}^*$ is derived by using the parameters (a, b) estimated from the given image. If the given image is captured by the camera model under $\mathcal{H}_0$, the estimates $(\hat{a}, \hat{b})$ should be in the neighborhood of (a_0, b_0) . The test performance depends on the accuracy of parameter estimation.

The Figure 5.4 illustrates the detection performance of the test $\tilde{\delta}_{\text{het}}^*$ on simulated images with 50 pixels. Here gain, the GLRT $\tilde{\delta}_{\text{het}}^*$ shows a small loss of power due to the imperfect employed segmentation.

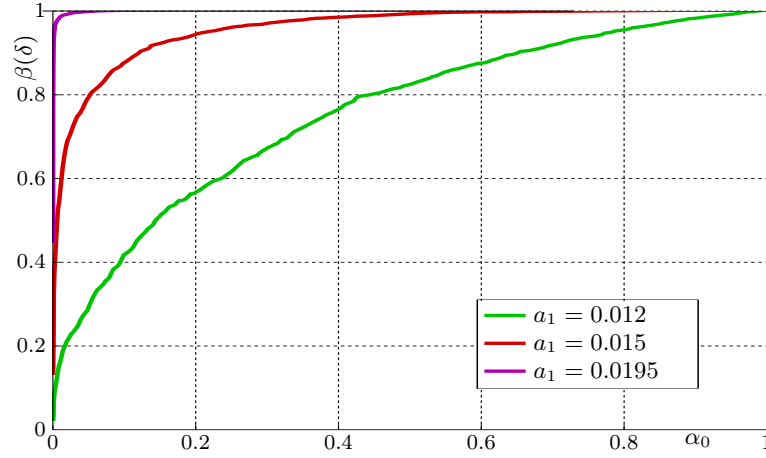


Figure 5.5: The detection performance of the test $\bar{\delta}_{\text{het}}^*$ with 200 pixels selected randomly on simulated images for $a_0 = 0.0115$ and different parameters a_1 .

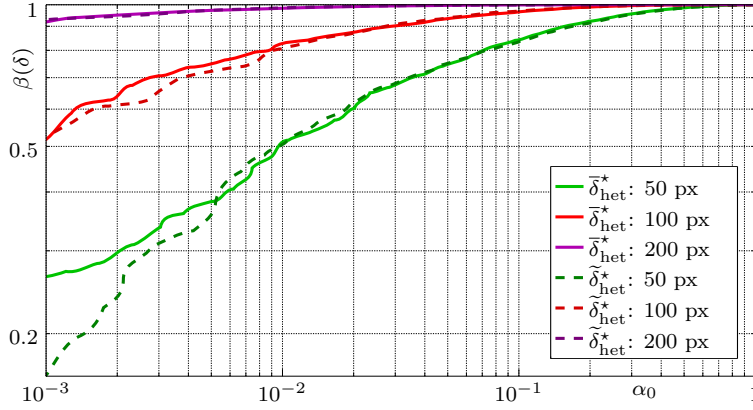


Figure 5.6: The detection performance of the test $\bar{\delta}_{\text{het}}^*$ and $\tilde{\delta}_{\text{het}}^*$ on simulated images for different numbers of pixels.

5.6 Numerical Experiments

5.6.1 Detection Performance on Simulated Database

The detection performance of the proposed GLRTs $\bar{\delta}_{\text{het}}^*$ and $\tilde{\delta}_{\text{het}}^*$ is first theoretically studied on a simulated database. The camera model $\mathcal{S}_0$ and $\mathcal{S}_1$ are respectively characterized by $(a_0, b_0) = (0.0115, 0.0002)$ and $(a_1, b_1) = (0.0195, 0.00025)$. These parameters respectively correspond to *Nikon D70* and *Nikon D200* camera models with ISO 200 estimated from the Dresden image database [168], see Figure 4.1. Moreover, the values correspond to an 8-bit image in the normalized $[0,1]$ scale. The camera parameters are used with 8-bit synthetic image of size 512×512 to generate randomly 5000 images for camera model $\mathcal{S}_0$ and 5000 images for camera model $\mathcal{S}_1$. The number of segments K is set to $K = 2^8$. The Figure 5.3 and Figure 5.4 illustrate respectively the detection performance of the GLRTs $\bar{\delta}_{\text{het}}^*$ and $\tilde{\delta}_{\text{het}}^*$ with 50 pixels selected randomly on the synthetic images. The segmentation method [133] used to

obtain homogeneous segments is certainly not perfect. Therefore, a slight error in the estimation of local expectations in each segment leads to a small loss of optimality. A perfect segmentation may not be available in practice due to the difficulty of controlling noise in natural images and the influence of image content. The perfect segmentation can be performed in this simulation since the original synthetic image is available. The empirical power with perfect segmentation fits accurately to the theoretical power. This also indicates that the variability of estimates are well taken into account.

Additionally, it is desirable to observe the detection performance for different camera parameters. The test $\bar{\delta}_{\text{het}}^*$ is conducted by keeping the parameters (a_0, b_0) and setting a_1 to $\{0.0195, 0.015, 0.012\}$. As expected, the Figure 5.5 shows that when the parameter a_1 tends to a_0 , the power function of the test $\bar{\delta}_{\text{het}}^*$ declines and the ROC curves tends to $\beta(\bar{\delta}_{\text{het}}^*) = \alpha_0$. In other words, the detection performance of the proposed tests depends on the discriminability of camera parameters (a, b) .

The Figure 5.6 illustrates in log-log scale the detection performance of the GLRT $\bar{\delta}_{\text{het}}^*$ and $\tilde{\delta}_{\text{het}}^*$ for $\{50, 100, 200\}$ pixels. A small loss of power is obviously revealed for 50 and 100 pixels between $\bar{\delta}_{\text{het}}^*$ and $\tilde{\delta}_{\text{het}}^*$ due to insufficient statistics. Both tests are identical with 200 pixels. Besides, the proposed GLRTs are perfect with 500 pixels, i.e. no error of detection has been observed from 5000 simulated images of camera model $\mathcal{S}_0$ and 5000 simulated images of camera model $\mathcal{S}_1$. Actually, the fact of selecting a number of pixels (e.g. 50, 100, 200 pixels) for the tests on the simulated data allows a better visibility since the empirical power function of the tests is perfect (e.g. $\beta(\delta) = 1$) with only 500 pixels. The proposed approach only requires a small number of pixels to achieve perfect detection performance while other methods usually exploit all the pixels without this performance. This emphasizes the sharpness of the proposed approach.

5.6.2 Detection Performance on Two *Nikon D70* and *Nikon D200* Camera Models

The experiments on simulated data give us an important insight: the employed segmentation is probably not perfect, which leads to an error in the estimation of pixels' expectation. In practice, for real images, this task is more difficult since the presence of edges or details can cause poor estimates. Those outliers need to be removed because they can dramatically affect the detection performance of the proposed GLRTs. Therefore, it is proposed to remove those outliers by the classical three-sigma rule [175]. Under normality, a pixel is considered as non-outlier if both following conditions are satisfied:

$$\begin{cases} |z_{k,i}^{\text{wapp}} - \hat{\mu}_k| & \leq 3\|\varphi\|_2\sqrt{\hat{a}\hat{\mu}_k + \hat{b}} \\ |z_{k,i}^{\text{wdet}}| & \leq 3\sqrt{\hat{a}\hat{\mu}_k + \hat{b}} \end{cases} \quad (5.40)$$

After outlier removal, all remaining pixels are used for the proposed tests.

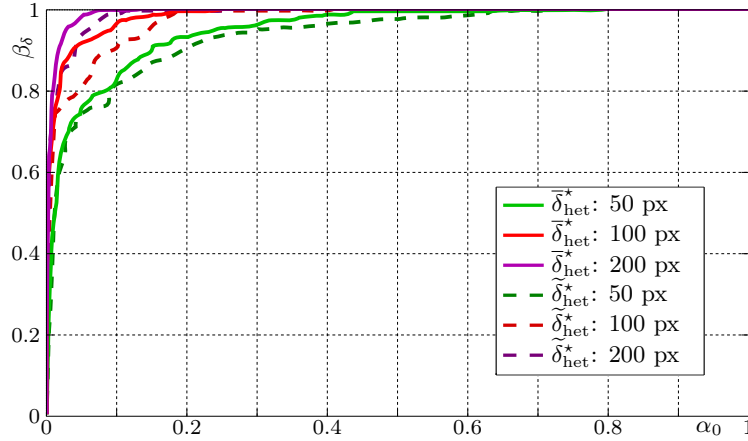


Figure 5.7: The detection performance of the GLRTs $\bar{\delta}_{\text{het}}^*$ and $\tilde{\delta}_{\text{het}}^*$ on the Dresden database for different numbers of pixels.

To highlight the relevance of the proposed tests, two camera models from the same brand *Nikon D70* and *Nikon D200* of the Dresden image database [168] are chosen to conduct experiments since two camera models of the same brand are expected to exhibit similar characteristics. These cameras are set at the same ISO 200. Prior to experiments, each RAW image was converted to an uncompressed format using Dcraw (with parameters -D -T -4 -j -v) and was decomposed into 4 sub-images. Only the red color channel is used in experiments. The *Nikon D70* and *Nikon D200* cameras are respectively set at $\mathcal{H}_0$ and $\mathcal{H}_1$. The camera parameters are estimated on each image following the WLS approach. The reference parameters (a_0, b_0) and (a_1, b_1) are obtained by averaging the previously estimated values over 50 images. The Figure 5.7 shows the detection performance of the GLRTs $\bar{\delta}_{\text{het}}^*$ and $\tilde{\delta}_{\text{het}}^*$ for different numbers of pixels. We can note a similar behavior to the ROC on simulated database. Obviously, there is a small loss of power between the two power functions since the test $\tilde{\delta}_{\text{het}}^*$ takes into account different estimates $(\hat{a}_1, \hat{b}_1)$ influenced by the image content. Nevertheless, two proposed tests are almost perfect for 500 pixels. Besides, the Figure 5.8 shows the comparison between the theoretical and empirical false alarm probability as a function of decision threshold. The two proposed GLRTs $\bar{\delta}_{\text{het}}^*$ and $\tilde{\delta}_{\text{het}}^*$ show a capacity of guaranteeing a prescribed false alarm rate, even though there is a slight difference in some cases (typically $\alpha_0 \leq 10^{-3}$) due to the influence of image content, the presence of weak outliers that can not be detected by the above outlier removal process, and the inaccuracy of the CLT for modeling tails.

5.6.3 Detection Performance on a Large Image Database

Experiments are then conducted on a large database to verify the efficiency of the proposed approach. The test set includes cameras from Dresden database [168], BOSS base [171] and our own database. Technical specifications of the cameras are shown in Table 5.1. The test set covers different devices per camera model, different

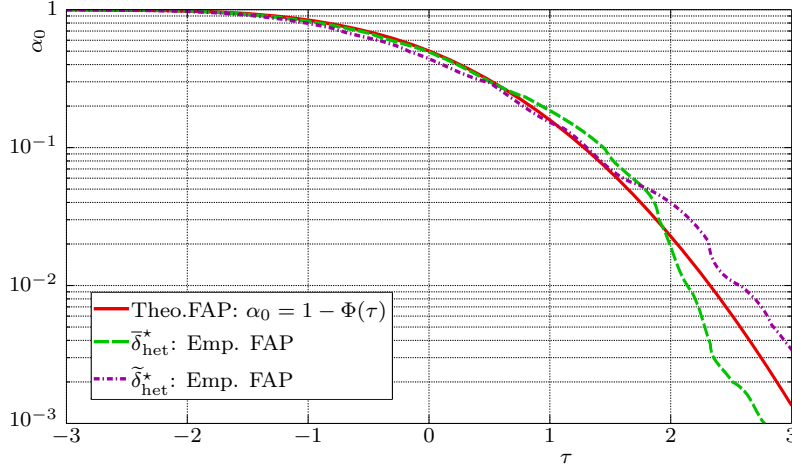


Figure 5.8: Comparison between the theoretical false alarm probability (FAP) and the empirical FAP, from real images of Dresden database, plotted as a function of decision threshold τ .

Camera Model		No. devices	Sensor size	Bit Depth	ISO Sensitivity	No. images
Nikon D70 [168, 171]	N70	3	23.7 × 15.6 mm CCD	12	200-400-800	1300
Nikon D90*	N90	2	23.6 × 15.8 mm CMOS	12	200-400-800	800
Nikon D200 [168]	N200	2	23.6 × 15.8 mm CCD	12	200	750
Canon 7D [171]	C7	1	22.3 × 14.9 mm CMOS	14	100	250
Canon 40D*	C40	2	22.2 × 14.8 mm CMOS	14	200-400-800	800
Canon 400D [171]	C400	1	22.2 × 14.8 mm CMOS	12	100-200-800	1300
Canon 450D*	C450	2	22.2 × 14.8 mm CMOS	14	100-400	800
Pentax K20D [171]	P	1	23.4 × 15.6 mm CMOS	12	100-200-400	1200

Table 5.1: Camera Model Used in Experiments (the symbol * indicates our own camera)

imaged scenes, different camera settings and different environmental conditions. The Dresden database [168] contains two devices per camera model. In case of the *Nikon D200* camera, two SLR-camera bodies were used with interchanging two different lenses for each acquired scene. Note that the BOSS base [171] also contains a *Nikon D70* camera device. Therefore, the test set finally contains 3 different devices of the *Nikon D70* camera model.

Prior to the experiments, a training stage involves using 50 RAW images per ISO sensitivity and per camera model to estimate the reference parameters (a_0, b_0) . Evidently, using more images will get a better estimate but it is also less realistic. The number of 50 is a good trade-off. In the experiments, the test $\tilde{\delta}_{het}^*$ is used to verify whether a given image is acquired by camera model under investigation. The decision threshold $\tilde{\tau}_{het}^*$ is given by Theorem 5.3 corresponding to the false alarm rate $\alpha_0 = 10^{-5}$. If the normalized GLR $\tilde{\Lambda}_{het}^*(\mathbf{Z})$ is smaller than the decision threshold $\tilde{\tau}_{het}^*$, the hypothesis $\mathcal{H}_0$ is accepted, i.e. the given image is declared taken from the camera model under investigation. On the contrary, the hypothesis $\mathcal{H}_1$ is accepted. The detection performance of the proposed test for each ISO 100, 200, 400, 800 is shown in Table 5.2. In these tables, each camera model is considered as hypothesis $\mathcal{H}_0$ (row) and all images (column) are tested against $\mathcal{H}_0$. The values in the tables

Inspected Images					Inspected Images						
	C7	C400	C450	P		N70	N90	N200	C40	C400	P
$\mathcal{H}_0$	C7	97.3	0	0	0	N70	99.6	1.1	0.2	0	0
	C400	0	99.7	0	0	N90	1.1	100	0	0	0
	C450	1.7	0	100	0	N200	0	0	99.6	0	0
	P	0	0	0	99.8	C40	0	0	0	100	0
						C400	0	0	0	0	99.8
(a) ISO 100					P	0.7	0	0	0	0	99.8
Inspected Images					Inspected Images						
	N70	N90	C40	P		N70	N90	C40	C400		
$\mathcal{H}_0$	N70	100	0	0	0	N70	100	0	0	0	
	N90	0	100	0	0	N90	0	100	0	0	
	C40	0	0	100	0	C40	0	0	100	0	
	P	0	0	0	100	C400	0	0	0	99.5	
(c) ISO 400					(b) ISO 200						
Inspected Images					Inspected Images						
	N70	N90	C40	P		N70	N90	C40	C400		
$\mathcal{H}_0$	N70	100	0	0	0	N70	100	0	0	0	
	N90	0	100	0	0	N90	0	100	0	0	
	C40	0	0	100	0	C40	0	0	100	0	
	P	0	0	0	100	C400	0	0	0	99.5	
(d) ISO 800					(a) ISO 100						

Table 5.2: Detection performance of the proposed detector.

Inspected Images						
	N70	N90	N200	C40	C400	P
$\mathcal{H}_0$	N70	97.9	0	0	0	0
	N90	0	85.5	0	0	0
	N200	0	0	100	0	0
	C40	0	0	0	100	0
	C400	0	0	0	0	99.8
	P	0	0	0	0	99.3

Table 5.3: Detection performance of PRNU-based detector for ISO 200.

indicate the percentage of images that are claimed taken from the camera model $\mathcal{H}_0$. It should be noted that these tables are not used in the same way as in the classification in which the sum for each class yields 100%. The inspected image is brought into the binary testing of the known camera model $\mathcal{H}_0$ against the others, thus the sum of a class may not yield 100%. Therefore, it could lead to a scenario that an image is claimed taken by at least two camera models. To deal with this scenario, the GLRT $\bar{\delta}_{\text{het}}^*$ can be performed to decide which camera model acquires the inspected image.

Potentially, there are many detectors in the literature for camera model identification, such as PRNU-based detector [31] and CFA-based detector [24]. However, they are based on the fact that the fingerprint obtained from images in the TIFF or JPEG format contains traces of post-acquisition processes (e.g. demosaicing) that carry information about the camera model. This chapter focuses camera model identification from RAW images, which have not gone through post-acquisition operations yet. Therefore, those detectors are not relevant to compare with the proposed detector. To the best of our knowledge, the proposed detector is the only one that focuses on RAW images to identify camera models. The PRNU-based detector proposed in [30, 32] can also deal with RAW images but it was proposed for camera instance identification, which differs from camera model identification. Nev-

ertheless, this detector is performed on the test set to compare with the proposed detector. The reference PRNU is calculated using 50 images per camera model. Instead of the normalized correlation [30], the Peak to Correlation Energy (PCE) [32] is used as a test statistic for the decision problem. The decision threshold is given by $\tau = (\Phi^{-1}(1 - \alpha_0))^2$ with $\alpha_0 = 10^{-5}$. This detector does not depend on ISO sensitivity. However, to ensure that the experiments are conducted in the same scenario for the two detectors, only images with the same ISO from the same device are considered. Table 5.3 shows the detection performance of the PRNU-based detector [32] for all images from the same device per camera model with ISO 200. It is noted that the proposed detector slightly outperforms the PRNU-based detector.

Remark 5.1. This chapter has shown the discriminability of the parameters (a, b) for different camera models at the same ISO sensitivity and has exploited this fingerprint for camera model identification. Therefore, one can rely on this fingerprint to train a classifier. However, as discussed in Section 2.3, the classification approach can not analytically provide the statistical performance, nor warrant a prescribed false alarm probability. By contrast, these criteria can be met in our works that design a statistical test within hypothesis testing framework based on *a priori* knowledge of image model.

5.7 Conclusion

In the literature, most of existing forensic methods are designed within a supervised classification framework using imaging noise (e.g. PRNU) or post-acquisition processes (e.g. CFA interpolation) as forensic features for camera model identification. This chapter proposes a different approach that designs a statistical test in the framework of hypothesis testing theory. The approach is based on the heteroscedastic noise model that is more relevant to describe a natural RAW image than the conventional AWGN model. Two parameters (a, b) characterizing the heteroscedastic noise model are exploited as camera fingerprint for camera model identification. The main strength of the proposed approach is the designing of the GLRTs with analytic performance and the guaranteeing of a prescribed false alarm probability.

The main limitation is that the proposed approach mainly focuses on RAW images that may not be available in practice. Since the proposed approach shows a almost perfect detection performance, it is worth extending it to other image formats, e.g. TIFF and JPEG. The most challenging part when extending this work is the impact of post-acquisition enhancement and compression processes. Non-linear processes (e.g. gamma correction) modify the heteroscedastic noise model. Moreover, the spatial correlation caused by CFA interpolation can lead to a difficulty of estimating accurately noise model parameters. Another limitation is the dependence of the proposed approach on ISO sensitivity. However, this is not crucial because there are not many ISO sensitivity for a camera and only a small number of images are sufficient to estimate the fingerprint (a, b) . Additionally, we can also exploit the relation between the camera parameters (a, b) and ISO sensitivity, as shown in Sec-

tion 3.2.1, to avoid the dependence of the proposed detectors. Moreover, it would be interesting that the ISO factor can be estimated and taken into account in the test. However, the design of such detectors lies out of the scope of the present works.

In our works, the Gaussian distribution is used to simplify the statistical study of the LR. However, this distribution is not accurate enough for modeling tails, which leads to the inaccuracy for very small false alarm probability. A more solid study for this case is desirable in future works.

In terms of computational complexity, the algorithm depends on the number of segments K and the image size. In this paper, the number of segments K is set to the number of quantization levels, e.g. $K = 2^\nu$. Therefore, the algorithm is of the order of $O(N \cdot 2^\nu)$.

5.8 Appendix

5.8.1 Expectation and Variance of the GLR $\bar{\Lambda}_{\text{het}}(z_{k,i}^{\text{wapp}})$ under Hypothesis $\mathcal{H}_j$

From (4.2) and (4.6), we obtain

$$z_{k,i}^{\text{wapp}} - \hat{\mu}_k \sim \mathcal{N}\left(0, \|\varphi\|_2^2 \sigma_{k,j}^2 \left(1 + \frac{1}{n_k}\right)\right). \quad (5.41)$$

Therefore, the expectation and variance of $\frac{(z_{k,i}^{\text{wapp}} - \hat{\mu}_k)^2}{\|\varphi\|_2^2}$ are given by

$$\mathbb{E}_{\mathcal{H}_j} \left[\frac{(z_{k,i}^{\text{wapp}} - \hat{\mu}_k)^2}{\|\varphi\|_2^2} \right] = \sigma_{k,j}^2 \left(1 + \frac{1}{n_k}\right) \quad (5.42)$$

$$\mathbb{V}\text{ar}_{\mathcal{H}_j} \left[\frac{(z_{k,i}^{\text{wapp}} - \hat{\mu}_k)^2}{\|\varphi\|_2^2} \right] = 2\sigma_{k,j}^4 \left(1 + \frac{1}{n_k}\right)^2. \quad (5.43)$$

Besides, it can be noted that two functions $h_1(\hat{\mu}_k)$ and $h_2(\hat{\mu}_k)$ only depend on the parameter $\hat{\mu}_k$. These functions are continuous and differentiable on $\mathbb{R}^+$. Using the Delta method in Lemma A.2, it follows from (4.6) that

$$h_1(\hat{\mu}_k) \xrightarrow{d} \mathcal{N}\left(h_1(\mu_k), (h_1'(\mu_k))^2 \frac{\|\varphi\|_2^2}{n_k} \sigma_{k,j}^2\right) \quad (5.44)$$

$$h_2(\hat{\mu}_k) \xrightarrow{d} \mathcal{N}\left(h_2(\mu_k), (h_2'(\mu_k))^2 \frac{\|\varphi\|_2^2}{n_k} \sigma_{k,j}^2\right). \quad (5.45)$$

Consequently, the first two moments of $\bar{\Lambda}_{\text{het}}(z_{k,i}^{\text{wapp}})$ under hypothesis $\mathcal{H}_j$ can be

given by

$$\begin{aligned}\mathbb{E}_{\mathcal{H}_j} [\bar{\Lambda}_{\text{het}}(z_{k,i}^{\text{wapp}})] &= \frac{1}{2} \mathbb{E}_{\mathcal{H}_j} [h_1(\hat{\mu}_k)] + \frac{1}{2} \mathbb{E}_{\mathcal{H}_j} [h_2(\hat{\mu}_k)] \mathbb{E}_{\mathcal{H}_j} \left[\frac{(z_{k,i}^{\text{wapp}} - \hat{\mu}_k)^2}{\|\varphi\|_2^2} \right] \\ &= \frac{1}{2} h_1(\mu_k) + \frac{1}{2} h_2(\mu_k) \sigma_{k,j}^2 \left(1 + \frac{1}{n_k} \right)\end{aligned}\quad (5.46)$$

$$\begin{aligned}\text{Var}_{\mathcal{H}_j} [\bar{\Lambda}_{\text{het}}(z_{k,i}^{\text{wapp}})] &= \frac{1}{4} \text{Var}_{\mathcal{H}_j} [h_1(\hat{\mu}_k)] + \frac{1}{4} \text{Var}_{\mathcal{H}_j} \left[h_2(\hat{\mu}_k) \frac{(z_{k,i}^{\text{wapp}} - \hat{\mu}_k)^2}{\|\varphi\|_2^2} \right] \\ &= \frac{1}{4} \text{Var}_{\mathcal{H}_j} [h_1(\hat{\mu}_k)] + \frac{1}{4} \mathbb{E}_{\mathcal{H}_j}^2 [h_2(\hat{\mu}_k)] \text{Var}_{\mathcal{H}_j} \left[\frac{(z_{k,i}^{\text{wapp}} - \hat{\mu}_k)^2}{\|\varphi\|_2^2} \right] \\ &\quad + \frac{1}{4} \text{Var}_{\mathcal{H}_j} [h_2(\hat{\mu}_k)] \mathbb{E}_{\mathcal{H}_j} \left[\frac{(z_{k,i}^{\text{wapp}} - \hat{\mu}_k)^4}{\|\varphi\|_2^4} \right] \\ &= \frac{1}{4} (h_1'(\mu_k))^2 \frac{\|\varphi\|_2^2}{n_k} \sigma_{k,j}^2 + \frac{1}{2} (h_2(\mu_k))^2 \sigma_{k,j}^4 \left(1 + \frac{1}{n_k} \right)^2 \\ &\quad + \frac{3}{4} (h_2'(\mu_k))^2 \frac{\|\varphi\|_2^2}{n_k} \sigma_{k,j}^6 \left(1 + \frac{1}{n_k} \right)^2.\end{aligned}\quad (5.47)$$

5.8.2 Expectation and Variance of the GLR $\tilde{\Lambda}_{\text{het}}(z_{k,i}^{\text{wapp}})$ under Hypothesis $\mathcal{H}_j$

For brevity, let denote

$$\zeta_k = \log(a_0 \hat{\mu}_k + b_0) - \log(\hat{a}_1 \hat{\mu}_k + \hat{b}_1) \quad \text{and} \quad \varrho_k = \frac{1}{a_0 \hat{\mu}_k + b_0} - \frac{1}{\hat{a}_1 \hat{\mu}_k + \hat{b}_1}, \quad (5.48)$$

such that the $\tilde{\Lambda}_{\text{het}}(z_{k,i}^{\text{wapp}})$ can be rewritten as

$$\tilde{\Lambda}_{\text{het}}(z_{k,i}^{\text{wapp}}) = \frac{1}{2} \zeta_k + \frac{1}{2} \varrho_k \frac{(z_{k,i}^{\text{wapp}} - \hat{\mu}_k)^2}{\|\varphi\|_2^2}. \quad (5.49)$$

The first two moments of $\frac{(z_{k,i}^{\text{wapp}} - \hat{\mu}_k)^2}{\|\varphi\|_2^2}$ under hypothesis $\mathcal{H}_j$ have been already given in (5.42) and (5.43). Therefore, it is necessary to calculate the first two moments of ζ_k and ϱ_k .

5.8.2.1 Statistical Properties of ζ_k

Using the Delta method in Lemma A.2 with the function $h(x) = \log(ax + b)$, from (4.6) we obtain

$$\log(a_0 \hat{\mu}_k + b_0) \xrightarrow{d} \mathcal{N} \left(\log(\sigma_{k,0}^2), \frac{a_0^2}{\sigma_{k,0}^4} \frac{\|\varphi\|_2^2}{n_k} \sigma_{k,j}^2 \right). \quad (5.50)$$

Besides, based on the definitions of mathematical expectation and variance, a little calculation shows that

$$\mathbb{E}_{\mathcal{H}_j}[\hat{a}_1\hat{\mu}_k + \hat{b}_1] = a_1\mu_k + b_1 = \sigma_{k,1}^2 \quad (5.51)$$

$$\begin{aligned} \mathbb{V}\text{ar}_{\mathcal{H}_j}[\hat{a}_1\hat{\mu}_k + \hat{b}_1] &= \mathbb{V}\text{ar}_{\mathcal{H}_j}[\hat{a}_1\hat{\mu}_k] + \mathbb{V}\text{ar}_{\mathcal{H}_j}[\hat{b}_1] + 2\mathbb{C}\text{ov}_{\mathcal{H}_j}[\hat{a}_1\hat{\mu}_k, \hat{b}_1] \\ &= \mathbb{E}_{\mathcal{H}_j}^2[\hat{a}_1]\mathbb{V}\text{ar}_{\mathcal{H}_j}[\hat{\mu}_k] + \mathbb{E}_{\mathcal{H}_j}[\hat{\mu}_k^2]\mathbb{V}\text{ar}_{\mathcal{H}_j}[\hat{a}_1] + \mathbb{V}\text{ar}_{\mathcal{H}_j}[\hat{b}_1] \\ &\quad + 2\mathbb{C}\text{ov}_{\mathcal{H}_j}[\hat{a}_1\hat{\mu}_k, \hat{b}_1] \\ &= a_1^2 \frac{\|\varphi\|_2^2}{n_k} \sigma_{k,j}^2 + (\mu_k^2 + \frac{\|\varphi\|_2^2}{n_k} \sigma_{k,j}^2) \sigma_{a_1}^2 + 2\mu_k \sigma_{a_1 b_1} + \sigma_{b_1}^2. \end{aligned} \quad (5.52)$$

Consequently, it follows from Corollary A.1 that

$$\mathbb{E}_{\mathcal{H}_j}[\log(\hat{a}_1\hat{\mu}_k + \hat{b}_1)] = \log(a_1\mu_k + b_1) = \log(\sigma_{k,1}^2) \quad (5.53)$$

$$\mathbb{V}\text{ar}_{\mathcal{H}_j}[\log(\hat{a}_1\hat{\mu}_k + \hat{b}_1)] = \frac{\mathbb{V}\text{ar}_{\mathcal{H}_j}[\hat{a}_1\hat{\mu}_k + \hat{b}_1]}{\sigma_{k,1}^4}, \quad (5.54)$$

where $\mathbb{V}\text{ar}_{\mathcal{H}_j}[\hat{a}_1\hat{\mu}_k + \hat{b}_1]$ is given in (5.52). Combining (5.50), (5.53) and (5.54), the first two moments of ζ_k can be expressed as

$$\begin{aligned} \mathbb{E}_{\mathcal{H}_j}[\zeta_k] &= \mathbb{E}_{\mathcal{H}_j}[\log(a_0\hat{\mu}_k + b_0)] - \mathbb{E}_{\mathcal{H}_j}[\log(\hat{a}_1\hat{\mu}_k + \hat{b}_1)] \\ &= \log(\sigma_{k,0}^2) - \log(\sigma_{k,1}^2) = h_1(\mu_k) \end{aligned} \quad (5.55)$$

$$\begin{aligned} \mathbb{V}\text{ar}_{\mathcal{H}_j}[\zeta_k] &= \mathbb{V}\text{ar}_{\mathcal{H}_j}[\log(a_0\hat{\mu}_k + b_0)] + \mathbb{V}\text{ar}_{\mathcal{H}_j}[\log(\hat{a}_1\hat{\mu}_k + \hat{b}_1)] \\ &= \frac{a_0^2}{\sigma_{k,0}^4} \frac{\|\varphi\|_2^2}{n_k} \sigma_{k,j}^2 + \frac{\mathbb{V}\text{ar}_{\mathcal{H}_j}[\hat{a}_1\hat{\mu}_k + \hat{b}_1]}{\sigma_{k,1}^4}. \end{aligned} \quad (5.56)$$

5.8.2.2 Statistical Properties of ϱ_k

Using the Delta method in Lemma A.2 with the function $h(x) = \frac{1}{ax+b}$, from (4.6) we obtain

$$\frac{1}{a_0\hat{\mu}_k + b_0} \xrightarrow{d} \mathcal{N}\left(\frac{1}{\sigma_{k,0}^2}, \frac{a_0^2}{\sigma_{k,0}^8} c_k \sigma_{k,j}^2\right). \quad (5.57)$$

Using Corollary A.1, it follows from (5.51) and (5.52) that

$$\mathbb{E}_{\mathcal{H}_j}\left[\frac{1}{\hat{a}_1\hat{\mu}_k + \hat{b}_1}\right] = \frac{1}{a_1\mu_k + b_1} = \frac{1}{\sigma_{k,1}^2} \quad (5.58)$$

$$\mathbb{V}\text{ar}_{\mathcal{H}_j}\left[\frac{1}{\hat{a}_1\hat{\mu}_k + \hat{b}_1}\right] = \frac{\mathbb{V}\text{ar}_{\mathcal{H}_j}[\hat{a}_1\hat{\mu}_k + \hat{b}_1]}{\sigma_{k,1}^8}. \quad (5.59)$$

Combining (5.57), (5.58) and (5.59), the first two moments of ϱ_k can be expressed as

$$\begin{aligned}\mathbb{E}_{\mathcal{H}_j}[\varrho_k] &= \mathbb{E}_{\mathcal{H}_j}\left[\frac{1}{a_0\hat{\mu}_k + b_0}\right] - \mathbb{E}_{\mathcal{H}_j}\left[\frac{1}{\hat{a}_1\hat{\mu}_k + \hat{b}_1}\right] \\ &= \frac{1}{\sigma_{k,0}^2} - \frac{1}{\sigma_{k,1}^2} = h_2(\mu_k)\end{aligned}\quad (5.60)$$

$$\begin{aligned}\mathbb{V}\text{ar}_{\mathcal{H}_j}[\varrho_k] &= \mathbb{V}\text{ar}_{\mathcal{H}_j}\left[\frac{1}{a_0\hat{\mu}_k + b_0}\right] + \mathbb{V}\text{ar}_{\mathcal{H}_j}\left[\frac{1}{\hat{a}_1\hat{\mu}_k + \hat{b}_1}\right] \\ &= \frac{a_0^2}{\sigma_{k,0}^8} \frac{\|\varphi\|_2^2}{n_k} \sigma_{k,j}^2 + \frac{\mathbb{V}\text{ar}_{\mathcal{H}_j}[\hat{a}_1\hat{\mu}_k + \hat{b}_1]}{\sigma_{k,1}^8}.\end{aligned}\quad (5.61)$$

5.8.2.3 Statistical Properties of $\tilde{\Lambda}_{\text{het}}(z_{k,i}^{\text{wapp}})$

Combining (5.42), (5.43), (5.55), (5.56), (5.60), (5.61), we derive the first two moments of the GLR $\tilde{\Lambda}_{\text{het}}(z_{k,i}^{\text{wapp}})$ under hypothesis $\mathcal{H}_j$

$$\begin{aligned}\mathbb{E}_{\mathcal{H}_j}[\tilde{\Lambda}_{\text{het}}(z_{k,i}^{\text{wapp}})] &= \frac{1}{2}\mathbb{E}_{\mathcal{H}_j}[\zeta_k] + \frac{1}{2}\mathbb{E}_{\mathcal{H}_j}[\varrho_k]\mathbb{E}_{\mathcal{H}_j}\left[\frac{(z_{k,i}^{\text{wapp}} - \hat{\mu}_k)^2}{\|\varphi\|_2^2}\right] \\ &= \frac{1}{2}h_1(\mu_k) + \frac{1}{2}h_2(\mu_k)\sigma_{k,j}^2\left(1 + \frac{1}{n_{k,j}}\right)\end{aligned}\quad (5.62)$$

$$\begin{aligned}\mathbb{V}\text{ar}_{\mathcal{H}_j}[\tilde{\Lambda}_{\text{het}}(z_{k,i}^{\text{wapp}})] &= \frac{1}{4}\mathbb{V}\text{ar}_{\mathcal{H}_j}[\zeta_k] + \frac{1}{4}\mathbb{E}_{\mathcal{H}_j}^2[\varrho_k]\mathbb{V}\text{ar}_{\mathcal{H}_j}\left[\frac{(z_{k,i}^{\text{wapp}} - \hat{\mu}_k)^2}{\|\varphi\|_2^2}\right] \\ &\quad + \frac{1}{4}\mathbb{V}\text{ar}_{\mathcal{H}_j}[\varrho_k]\mathbb{E}_{\mathcal{H}_j}\left[\frac{(z_{k,i}^{\text{wapp}} - \hat{\mu}_k)^4}{\|\varphi\|_2^4}\right] \\ &= \frac{1}{4}\left[\frac{a_0^2}{\sigma_{k,0}^4} \frac{\|\varphi\|_2^2}{n_k} \sigma_{k,j}^2 + \frac{\mathbb{V}\text{ar}_{\mathcal{H}_j}[\hat{a}_1\hat{\mu}_k + \hat{b}_1]}{\sigma_{k,1}^4}\right] \\ &\quad + \frac{1}{2}(h_2(\mu_k))^2 \sigma_{k,j}^4 \left(1 + \frac{1}{n_k}\right)^2 \\ &\quad + \frac{3}{4}\sigma_{k,j}^4 \left(1 + \frac{1}{n_k}\right)^2 \left[\frac{a_0^2}{\sigma_{k,0}^8} \frac{\|\varphi\|_2^2}{n_k} \sigma_{k,j}^2 + \frac{\mathbb{V}\text{ar}_{\mathcal{H}_j}[\hat{a}_1\hat{\mu}_k + \hat{b}_1]}{\sigma_{k,1}^8}\right].\end{aligned}\quad (5.63)$$

Camera Model Identification Based on the Generalized Noise Model

Contents

6.1	Introduction	99
6.2	Camera Fingerprint	100
6.3	Optimal Detector for Camera Model Identification Problem	102
6.3.1	Hypothesis Testing Formulation	102
6.3.2	LRT for Two Simple Hypotheses	103
6.4	Practical Context: GLRT	105
6.4.1	GLRT with Unknown Image Parameters	105
6.4.2	GLRT with Unknown Image and Camera Parameters	106
6.5	Numerical Experiments	108
6.5.1	Detection Performance on Simulated Database	108
6.5.2	Detection Performance on Two <i>Nikon D70</i> and <i>Nikon D200</i> Camera Models	109
6.5.3	Detection Performance on a Large Image Database	110
6.6	Conclusion	111
6.7	Appendix	112

6.1 Introduction

Chapter 5 has proposed a new approach for camera model identification problem by designing a statistical test within the framework of hypothesis testing theory based on the heteroscedastic noise model. Even though that approach shows a nearly perfect detection performance, it has two main limitations. Firstly, it mainly focuses on RAW images that may not be available in practice. Secondly, the proposed fingerprint (a, b) depends on ISO sensitivity [134]. Although this is not crucial because there are not many ISO sensitivity for a camera and only a small number of images are sufficient to estimate the reference camera parameters (a, b) for each ISO sensitivity, it is more desirable to rely on a fingerprint that is invariant to image

content and camera settings, and robust to non-linear processing operations (e.g. gamma correction).

The goal of this chapter is to design a statistical test within hypothesis testing framework for camera model identification from JPEG images based on the generalized noise model (4.42). The main contributions are the following:

- This approach is based on the generalized noise model that characterizes a natural image in TIFF or JPEG format accurately. The parameters $(\tilde{a}, \tilde{b}, \gamma)$ characterizing the generalized noise model are proposed to be exploited as camera fingerprint to identify camera models.
- Stating the camera model identification problem into hypothesis testing framework, the most powerful test given by the LRT is studied in the ideal context where all model parameters are known. In practice, the model parameters are unknown. Two GLRTs are designed to deal with the difficulty of unknown parameters. The statistical performance of the proposed tests is analytically established. Moreover, they allow the guaranteeing of a prescribed false-alarm rate and the setting of decision threshold independently of the image content.

The work in this chapter has been presented in [176] and valorized in the patent [177]. The chapter is organized as follows. Section 6.2 studies properties of the parameters $(\tilde{a}, \tilde{b}, \gamma)$ that are further exploited as camera fingerprint for camera model identification. Section 6.3 states the camera model identification problem in the framework of hypothesis testing theory and studies an optimal detector assuming that all model parameters are known. Section 6.4 designs two GLRTs to address the difficulty of unknown parameters. Section 6.5 presents numerical results of two proposed GLRTs on simulated and real natural JPEG images. Finally, Section 6.6 concludes the chapter.

6.2 Camera Fingerprint

The study of noise statistics in a natural image from RAW format to TIFF format has been performed in Section 4.3. The main steps of image processing pipeline are modeled. By starting from the heteroscedastic noise model and taking into account the non-linear effect of gamma correction, the generalized noise model (4.42) is developed. This noise model gives the pixel's variance as a non-linear function of pixel's expectation. The generalized noise model is relevant to characterize a natural image in TIFF or JPEG format acquired by a digital camera. This thesis mainly focuses on JPEG format that is the most popular format for images produced by a digital camera. TIFF images could be sensitive to ISO and they are less available in practice.

The application of hypothesis testing theory requires to know the statistical distribution of a JPEG pixel. To this end, it is necessary to model the JPEG compression chain rigorously. JPEG compression mainly involves the DCT operation

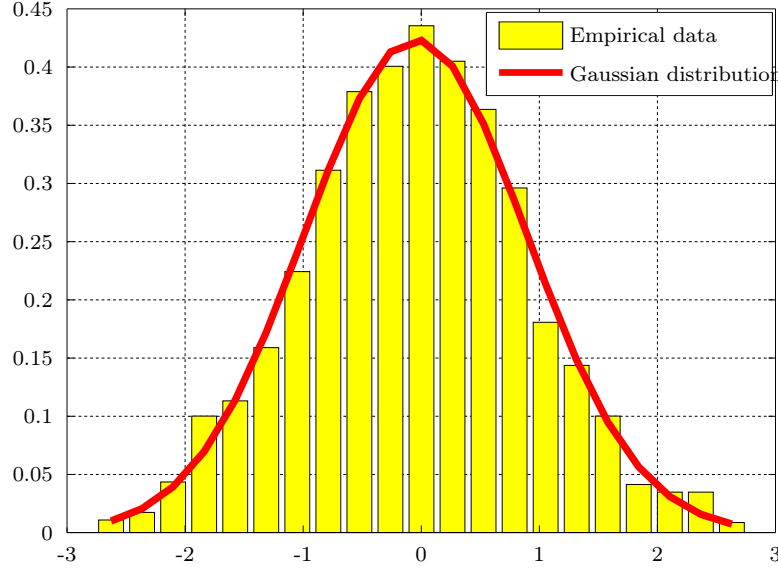


Figure 6.1: Empirical distribution of noise residuals $\tilde{z}_{k,i}^{\text{res}}$ in a segment compared with theoretical Gaussian distribution.

and the quantization in the DCT domain, while the JPEG decompression performs dequantization and inverse DCT operation to return to spatial domain, see Section 2.2. In general, the reconstructed image from a compressed file usually differs from the original image. There are two fundamental factors involved in this spatial-domain error [166]: the DCT basis vectors, and quantization error introduced in the DCT domain. The spatial-domain error at a pixel location is a weighted sum of 64 DCT-domain quantization errors within a 8×8 block. Providing an exact statistical distribution of a JPEG pixel is a challenging task due to the difficulty of establishing mathematically the model of DCT coefficients, characterizing the effect of quantization in the DCT domain, and deriving the distribution of the sum of those random variables.

To overcome this difficulty, it is proposed to invoke the Lindeberg CLT. In the homogeneous segment $\tilde{S}_k$ (see Section 4.3.2), a JPEG pixel $\tilde{z}_{k,i}$ can be decomposed as

$$\tilde{z}_{k,i} = \tilde{\mu}_k + \eta_{\tilde{z}_{k,i}}, \quad (6.1)$$

where $\tilde{\mu}_k$ denotes the expectation of all pixels in the segment $\tilde{S}_k$ and $\eta_{\tilde{z}_{k,i}}$ accounts for spatial-domain noise after JPEG compression. The pixels in the segment $\tilde{S}_k$ are assumed to be i.i.d, thus they have the same expectation and variance. Since the DCT can approximately decorrelate the input image [166], the spatial-domain noise $\eta_{\tilde{z}_{k,i}}$ in the decompressed JPEG image can be seen as a linear combination of independent random variables because of IDCT operation. In virtue of the Lindeberg CLT, the random variable $\eta_{\tilde{z}_{k,i}}$ can be approximately modeled by the Gaussian distribution with zero-mean [166]. Meanwhile, the variance of noise $\eta_{\tilde{z}_{k,i}}$ depends on pixel's expectation $\tilde{\mu}_k$ according to the generalized noise model (4.42). Figure 6.1

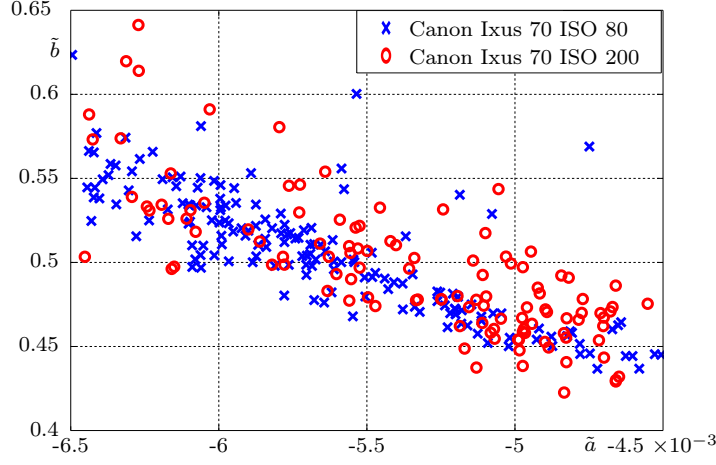


Figure 6.2: Estimated parameters $(\tilde{a}, \tilde{b})$ on JPEG images issued from *Canon Ixus 70* camera with different camera settings.

shows the empirical distribution of residuals $\tilde{z}_{k,i}^{\text{res}}$ in a segment extracted from a natural JPEG image, compared with theoretical Gaussian distribution. Finally, the JPEG pixel $\tilde{z}_{k,i}$ can be modeled as

$$\tilde{z}_{k,i} \sim \mathcal{N}(\tilde{\mu}_k, \tilde{\sigma}_k^2), \quad \text{with} \quad \tilde{\sigma}_k^2 = f_{\text{gen}}(\tilde{\mu}_k; \tilde{a}, \tilde{b}, \gamma). \quad (6.2)$$

For camera model identification problem, it is necessary to evaluate the variability of the camera parameters $(\tilde{a}, \tilde{b}, \gamma)$ for different camera settings and different devices per camera model, and to verify their discriminability for different camera models. The Figure 6.2 shows the parameters $(\tilde{a}, \tilde{b})$ estimated from JPEG images of *Canon Ixus 70* camera with different camera settings. Furthermore, the Figure 6.3 shows the parameters $(\tilde{a}, \tilde{b})$ estimated from JPEG images that are acquired by different devices of *Canon Ixus 70* camera model and the Figure 4.3 illustrates the discriminability of the parameters for different camera models. It is worth noting that in contrast to the parameters (a, b) , the parameters $(\tilde{a}, \tilde{b})$ are invariant to imaged scenes and camera settings, robust to non-linear post-acquisition processes and discriminative for different camera models. Since the difference between estimated gamma factors of different camera models is small, we do not report them in this chapter. The parameters $(\tilde{a}, \tilde{b}, \gamma)$ are proposed to be exploited as camera fingerprint to identify camera models.

6.3 Optimal Detector for Camera Model Identification Problem

6.3.1 Hypothesis Testing Formulation

The camera model $\mathcal{S}_j$, $j \in \{0, 1\}$ is now characterized by three parameters $(\tilde{a}_j, \tilde{b}_j, \gamma_j)$. It is assumed that $(\tilde{a}_0, \tilde{b}_0, \gamma_0) \neq (\tilde{a}_1, \tilde{b}_1, \gamma_1)$. In a binary hypothesis

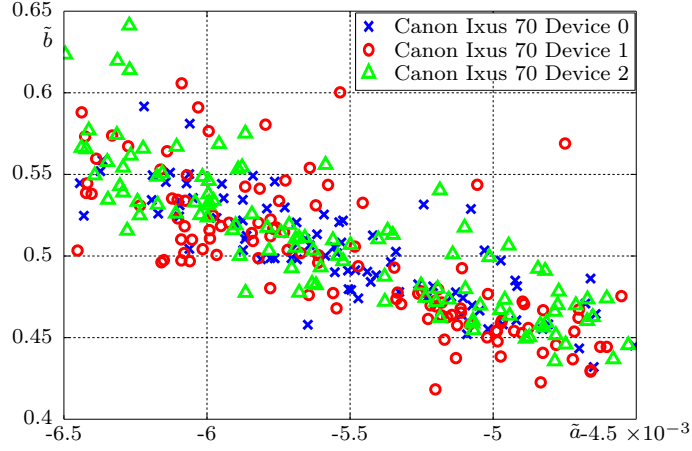


Figure 6.3: Estimated parameters $(\tilde{a}, \tilde{b})$ on JPEG images issued from different devices of *Canon Ixus 70* model.

testing, the inspected JPEG image $\tilde{\mathbf{Z}}$ is either acquired by camera model $\mathcal{S}_0$ or camera model $\mathcal{S}_1$. The goal of the test is to decide between two following hypotheses

$$\begin{cases} \mathcal{H}_0 = \left\{ \tilde{z}_{k,i} \sim \mathcal{N}(\tilde{\mu}_k, \tilde{\sigma}_{k,0}^2), \forall k \in \{1, \dots, K\}, \forall i \in \{1, \dots, \tilde{n}_k\} \right\} \\ \mathcal{H}_1 = \left\{ \tilde{z}_{k,i} \sim \mathcal{N}(\tilde{\mu}_k, \tilde{\sigma}_{k,1}^2), \forall k \in \{1, \dots, K\}, \forall i \in \{1, \dots, \tilde{n}_k\} \right\}, \end{cases} \quad (6.3)$$

where $\tilde{\sigma}_{k,j}^2 = f_{\text{gen}}(\tilde{\mu}_k; \tilde{a}_j, \tilde{b}_j, \gamma_j)$ is the noise variance with respect to the expectation $\tilde{\mu}_k$ under hypothesis $\mathcal{H}_j$. Let

$$\mathcal{K}_{\alpha_0} = \left\{ \delta : \sup_{(\tilde{\mu}, \tilde{a}_0, \tilde{b}_0, \gamma_0)} \mathbb{P}_{\mathcal{H}_0} [\delta(\tilde{\mathbf{Z}}) = \mathcal{H}_1] \leq \alpha_0 \right\}$$

be the class of tests whose false alarm probability is upper-bounded by the rate α_0 where $\tilde{\mu} = (\tilde{\mu}_1, \dots, \tilde{\mu}_K)$ is the mean vector. Among all the tests in the class $\mathcal{K}_{\alpha_0}$, it is aimed at finding a test δ which maximizes the power function $\beta(\delta)$. It can be noted that several steps in this chapter are similar as in Chapter 5. It is assumed that the camera parameters $(\tilde{a}_0, \tilde{b}_0, \gamma_0)$ are known in advance, i.e. the image under investigation $\tilde{\mathbf{Z}}$ will be tested against the known camera model $\mathcal{S}_0$. The main goal of this chapter is to study the optimal test for the problem (6.3) and design the GLRT to deal with the difficulty of unknown parameters $(\tilde{\mu}, \tilde{a}_1, \tilde{b}_1, \gamma_1)$.

6.3.2 LRT for Two Simple Hypotheses

When all model parameters are known, in virtue of the Neyman-Pearson lemma, the MP test δ_{gen}^* solving the problem (6.3) is the LRT given by the following decision

rule

$$\delta_{\text{gen}}^*(\tilde{\mathbf{Z}}) = \begin{cases} \mathcal{H}_0 & \text{if } \Lambda_{\text{gen}}(\tilde{\mathbf{Z}}) = \sum_{k=1}^K \sum_{i=1}^{\tilde{n}_k} \Lambda_{\text{gen}}(\tilde{z}_{k,i}) < \tau_{\text{gen}}^* \\ \mathcal{H}_1 & \text{if } \Lambda_{\text{gen}}(\tilde{\mathbf{Z}}) = \sum_{k=1}^K \sum_{i=1}^{\tilde{n}_k} \Lambda_{\text{gen}}(\tilde{z}_{k,i}) \geq \tau_{\text{gen}}^*, \end{cases} \quad (6.4)$$

where the LR $\Lambda_{\text{gen}}(\tilde{z}_{k,i})$ is defined by

$$\Lambda_{\text{gen}}(\tilde{z}_{k,i}) = \frac{1}{2} \left(\log(\tilde{\sigma}_{k,0}^2) - \log(\tilde{\sigma}_{k,1}^2) \right) + \frac{1}{2} \left(\frac{1}{\tilde{\sigma}_{k,0}^2} - \frac{1}{\tilde{\sigma}_{k,1}^2} \right) (\tilde{z}_{k,i} - \tilde{\mu}_k)^2. \quad (6.5)$$

Proposition 6.1. *Under hypothesis $\mathcal{H}_j$, the first two moments of the LR $\Lambda_{\text{gen}}(\tilde{z}_{k,i})$ are given by*

$$\mathbb{E}_{\mathcal{H}_j} [\Lambda_{\text{gen}}(\tilde{z}_{k,i})] = \frac{1}{2} \left(\log(\tilde{\sigma}_{k,0}^2) - \log(\tilde{\sigma}_{k,1}^2) \right) + \frac{1}{2} \left(\frac{1}{\tilde{\sigma}_{k,0}^2} - \frac{1}{\tilde{\sigma}_{k,1}^2} \right) \tilde{\sigma}_{k,j}^2 \quad (6.6)$$

$$\text{Var}_{\mathcal{H}_j} [\Lambda_{\text{gen}}(\tilde{z}_{k,i})] = \frac{1}{2} \left(\frac{1}{\tilde{\sigma}_{k,0}^2} - \frac{1}{\tilde{\sigma}_{k,1}^2} \right)^2 \tilde{\sigma}_{k,j}^4. \quad (6.7)$$

Proof. It follows from $\tilde{z}_{k,i} \sim \mathcal{N}(\tilde{\mu}_k, \tilde{\sigma}_{k,j}^2)$ under each hypothesis $\mathcal{H}_j$, $j \in \{0, 1\}$, that

$$\frac{(\tilde{z}_{k,i} - \tilde{\mu}_k)^2}{\tilde{\sigma}_{k,j}^2} \sim \chi_1^2.$$

Therefore, the first two moments of $(\tilde{z}_{k,i} - \tilde{\mu}_k)^2$ under $\mathcal{H}_j$ can be given by

$$\begin{aligned} \mathbb{E}_{\mathcal{H}_j} [(\tilde{z}_{k,i} - \tilde{\mu}_k)^2] &= \tilde{\sigma}_{k,j}^2 \\ \text{Var}_{\mathcal{H}_j} [(\tilde{z}_{k,i} - \tilde{\mu}_k)^2] &= 2\tilde{\sigma}_{k,j}^4. \end{aligned}$$

The proof follows immediately. $\square$

Under hypothesis $\mathcal{H}_j$, in virtue of Lindeberg CLT, the LR $\Lambda_{\text{gen}}(\tilde{\mathbf{Z}})$ follows the Gaussian distribution with expectation $m_{\text{gen}}^{(j)}$ and variance $v_{\text{gen}}^{(j)}$ as

$$\Lambda_{\text{gen}}(\tilde{\mathbf{Z}}) \xrightarrow{d} \mathcal{N}(m_{\text{gen}}^{(j)}, v_{\text{gen}}^{(j)}) \quad (6.8)$$

with

$$m_{\text{gen}}^{(j)} = \sum_{k=1}^K \sum_{i=1}^{\tilde{n}_k} \left[\frac{1}{2} \left(\log(\tilde{\sigma}_{k,0}^2) - \log(\tilde{\sigma}_{k,1}^2) \right) + \frac{1}{2} \left(\frac{1}{\tilde{\sigma}_{k,0}^2} - \frac{1}{\tilde{\sigma}_{k,1}^2} \right) \tilde{\sigma}_{k,j}^2 \right] \quad (6.9)$$

$$v_{\text{gen}}^{(j)} = \sum_{k=1}^K \sum_{i=1}^{\tilde{n}_k} \frac{1}{2} \left(\frac{1}{\tilde{\sigma}_{k,0}^2} - \frac{1}{\tilde{\sigma}_{k,1}^2} \right)^2 \tilde{\sigma}_{k,j}^4. \quad (6.10)$$

Finally, the LRT δ_{gen}^* based on the normalized LR $\Lambda_{\text{gen}}^*(\tilde{\mathbf{Z}}) = \frac{\Lambda_{\text{gen}}(\tilde{\mathbf{Z}}) - m_{\text{gen}}^{(0)}}{\sqrt{v_{\text{gen}}^{(0)}}}$ is rewritten as follows

$$\delta_{\text{gen}}^*(\tilde{\mathbf{Z}}) = \begin{cases} \mathcal{H}_0 & \text{if } \Lambda_{\text{gen}}^*(\tilde{\mathbf{Z}}) < \tau_{\text{gen}}^* \\ \mathcal{H}_1 & \text{if } \Lambda_{\text{gen}}^*(\tilde{\mathbf{Z}}) \geq \tau_{\text{gen}}^* \end{cases} \quad (6.11)$$

where the decision threshold τ_{gen}^* is the solution of the equation $\mathbb{P}_{\mathcal{H}_0}[\Lambda_{\text{gen}}^*(\tilde{\mathbf{Z}}) \geq \tau_{\text{gen}}^*] = \alpha_0$ to ensure the LRT δ_{gen}^* to be in the class $\mathcal{K}_{\alpha_0}$.

Theorem 6.1. *In an ideal context where all the model parameters $(\tilde{\mu}, \tilde{a}_j, \tilde{b}_j, \gamma)$ are known in advance, the decision threshold and the power function of the LRT δ_{gen}^* are given by*

$$\tau_{\text{gen}}^* = \Phi^{-1}(1 - \alpha_0) \quad (6.12)$$

$$\beta(\delta_{\text{gen}}^*) = 1 - \Phi\left(\frac{m_{\text{gen}}^{(0)} - m_{\text{gen}}^{(1)} + \tau_{\text{gen}}^* \sqrt{v_{\text{gen}}^{(0)}}}{\sqrt{v_{\text{gen}}^{(1)}}}\right). \quad (6.13)$$

Proof. From (6.8) and the definitions of the decision threshold and the test power, the proof follows immediately. $\square$

6.4 Practical Context: GLRT

6.4.1 GLRT with Unknown Image Parameters

The GLRT designed in this subsection deals with the difficulty of unknown image parameters $\tilde{\mu}_k$ assuming that the camera parameters $(\tilde{a}_0, \tilde{b}_0, \gamma_0)$ and $(\tilde{a}_1, \tilde{b}_1, \gamma_1)$ are known. By replacing $\tilde{\mu}_k$ by $\hat{\mu}_k$, which is given in (4.47), in the LR $\Lambda_{\text{gen}}(\tilde{z}_{k,i})$, the GLR $\bar{\Lambda}_{\text{gen}}(\tilde{z}_{k,i})$ is now given by

$$\begin{aligned} \bar{\Lambda}_{\text{gen}}(\tilde{z}_{k,i}) &= \frac{1}{2} \left(\log(f_{\text{gen}}(\hat{\mu}_k; \tilde{a}_0, \tilde{b}_0, \gamma_0)) - \log(f_{\text{gen}}(\hat{\mu}_k; \tilde{a}_1, \tilde{b}_1, \gamma_1)) \right) \\ &+ \frac{1}{2} \left(\frac{1}{f_{\text{gen}}(\hat{\mu}_k; \tilde{a}_0, \tilde{b}_0, \gamma_0)} - \frac{1}{f_{\text{gen}}(\hat{\mu}_k; \tilde{a}_1, \tilde{b}_1, \gamma_1)} \right) (\tilde{z}_{k,i} - \hat{\mu}_k)^2. \end{aligned} \quad (6.14)$$

Since the variance of $\hat{\mu}_k$ is assumed to be negligible, the estimate $\hat{\mu}_k$ can be treated as constant $\tilde{\mu}_k$. Therefore, the expectation and variance of the GLR $\bar{\Lambda}_{\text{gen}}(\tilde{z}_{k,i})$ under each hypothesis do not change

$$\begin{aligned} \mathbb{E}_{\mathcal{H}_j}[\bar{\Lambda}_{\text{gen}}(\tilde{z}_{k,i})] &= \mathbb{E}_{\mathcal{H}_j}[\Lambda_{\text{gen}}(\tilde{z}_{k,i})] \\ &= \frac{1}{2} \left(\log(\tilde{\sigma}_{k,0}^2) - \log(\tilde{\sigma}_{k,1}^2) \right) + \frac{1}{2} \left(\frac{1}{\tilde{\sigma}_{k,0}^2} - \frac{1}{\tilde{\sigma}_{k,1}^2} \right) \tilde{\sigma}_{k,j}^2 \end{aligned} \quad (6.15)$$

$$\begin{aligned} \text{Var}_{\mathcal{H}_j}[\bar{\Lambda}_{\text{gen}}] &= \text{Var}_{\mathcal{H}_j}[\Lambda_{\text{gen}}(\tilde{z}_{k,i})] \\ &= \frac{1}{2} \left(\frac{1}{\tilde{\sigma}_{k,0}^2} - \frac{1}{\tilde{\sigma}_{k,1}^2} \right)^2 \tilde{\sigma}_{k,j}^4. \end{aligned} \quad (6.16)$$

Consequently, the GLR $\bar{\Lambda}_{\text{gen}}(\tilde{\mathbf{Z}}) = \sum_{k=1}^K \sum_{i=1}^{\tilde{n}_k} \hat{\Lambda}_1(z_{k,i})$ also follows the Gaussian distribution with expectation $m_{\text{gen}}^{(j)}$ and variance $v_{\text{gen}}^{(j)}$ under hypothesis $\mathcal{H}_j$

$$\bar{\Lambda}_{\text{gen}}(\tilde{\mathbf{Z}}) \xrightarrow{d} \mathcal{N}(m_{\text{gen}}^{(j)}, v_{\text{gen}}^{(j)}) \quad (6.17)$$

where $m_{\text{gen}}^{(j)}$ and $v_{\text{gen}}^{(j)}$ are given in (6.9) and (6.10) respectively.

Finally, the GLRT $\bar{\delta}_{\text{gen}}^*$ based on the normalized GLR $\bar{\Lambda}_{\text{gen}}^*(\tilde{\mathbf{Z}}) = \frac{\bar{\Lambda}_{\text{gen}}(\tilde{\mathbf{Z}}) - \hat{m}_{\text{gen}}^{(0)}}{\sqrt{\hat{v}_{\text{gen}}^{(0)}}}$ is given by

$$\bar{\delta}_{\text{gen}}^*(\tilde{\mathbf{Z}}) = \begin{cases} \mathcal{H}_0 & \text{if } \bar{\Lambda}_{\text{gen}}^*(\tilde{\mathbf{Z}}) < \bar{\tau}_{\text{gen}}^* \\ \mathcal{H}_1 & \text{if } \bar{\Lambda}_{\text{gen}}^*(\tilde{\mathbf{Z}}) \geq \bar{\tau}_{\text{gen}}^* \end{cases} \quad (6.18)$$

where $\hat{m}_{\text{gen}}^{(0)}$ and $\hat{v}_{\text{gen}}^{(0)}$ are the estimates of the expectation $m_{\text{gen}}^{(0)}$ and variance $v_{\text{gen}}^{(0)}$ by replacing $\tilde{\mu}_k$ by $\hat{\mu}_k$ in (6.9) and (6.10), and the decision threshold $\bar{\tau}_{\text{gen}}^*$ is the solution of the equation $\mathbb{P}_{\mathcal{H}_0}[\bar{\Lambda}_{\text{gen}}^*(\tilde{\mathbf{Z}}) \geq \bar{\tau}_{\text{gen}}^*] = \alpha_0$. From the Slutsky's theorem, the decision threshold and the power of the GLRT $\bar{\delta}_{\text{gen}}^*$ can be accordingly defined as in the Theorem 6.1.

6.4.2 GLRT with Unknown Image and Camera Parameters

Before designing the GLRT, the ML estimation of camera parameters $(\tilde{a}_1, \tilde{b}_1, \gamma_1)$ is performed on the inspected image $\tilde{\mathbf{Z}}$; see Section 4.3.2. Instead of estimating three parameters simultaneously, it is proposed to set $\gamma_1 = \gamma_0$ and reduce the maximization problem (4.51) with three parameters back to the one with two parameters. By setting $\gamma_1 = \gamma_0$, the inspected image $\tilde{\mathbf{Z}}$ is expected to be taken by the camera model $\mathcal{S}_0$. Let $(\hat{a}_1, \hat{b}_1)$ be the solution of that maximization problem. The ML estimates $(\hat{a}_1, \hat{b}_1)$ are asymptotically consistent, i.e. they asymptotically converge in probability to their true value: $\hat{a}_1 \xrightarrow{p} \tilde{a}_1$ and $\hat{b}_1 \xrightarrow{p} \tilde{b}_1$. The parameters $(\tilde{a}_1, \tilde{b}_1, \gamma_1)$ would characterize an unknown camera model. Furthermore, the ML estimates $(\hat{a}_1, \hat{b}_1)$ exhibit a certain variability. Let $\sigma_{\hat{a}_1}^2$, $\sigma_{\hat{b}_1}^2$, $\sigma_{\hat{a}_1 \hat{b}_1}$ denote respectively the variance of $\hat{a}_1$, the variance of $\hat{b}_1$ and the covariance between $\hat{a}_1$ and $\hat{b}_1$, see also discussions in Section 6.5.1 for this covariance matrix.

The GLRT is designed by replacing unknown parameters $(\tilde{\mu}_k, \tilde{a}_1, \tilde{b}_1)$ by $(\hat{\mu}_k, \hat{a}_1, \hat{b}_1)$ in the LR $\Lambda_{\text{gen}}(\tilde{z}_{k,i})$. The GLR $\tilde{\Lambda}_{\text{gen}}(\tilde{z}_{k,i})$ is now given by

$$\begin{aligned} \tilde{\Lambda}_{\text{gen}}(\tilde{z}_{k,i}) = & \frac{1}{2} \left(\log(f_{\text{gen}}(\hat{\mu}_k; \tilde{a}_0, \tilde{b}_0, \gamma_0)) - \log(f_{\text{gen}}(\hat{\mu}_k; \hat{a}_1, \hat{b}_1, \gamma_1)) \right) \\ & + \frac{1}{2} \left(\frac{1}{f_{\text{gen}}(\hat{\mu}_k; \tilde{a}_0, \tilde{b}_0, \gamma_0)} - \frac{1}{f_{\text{gen}}(\hat{\mu}_k; \hat{a}_1, \hat{b}_1, \gamma_1)} \right) (\tilde{z}_{k,i} - \hat{\mu}_k)^2. \end{aligned} \quad (6.19)$$

Proposition 6.2. *Under hypothesis $\mathcal{H}_j$, the first two moments of the GLR $\tilde{\Lambda}_{\text{gen}}(\tilde{z}_{k,i})$ are given by*

$$\mathbb{E}_{\mathcal{H}_j} [\tilde{\Lambda}_{\text{gen}}(\tilde{z}_{k,i})] = \frac{1}{2} \left(\log(\tilde{\sigma}_{k,0}^2) - \log(\tilde{\sigma}_{k,1}^2) \right) + \frac{1}{2} \left(\frac{1}{\tilde{\sigma}_{k,0}^2} - \frac{1}{\tilde{\sigma}_{k,1}^2} \right) \tilde{\sigma}_{k,j}^2 \quad (6.20)$$

$$\begin{aligned} \text{Var}_{\mathcal{H}_j} [\tilde{\Lambda}_{\text{gen}}(\tilde{z}_{k,i})] &= \frac{1}{2} \left(\frac{1}{\tilde{\sigma}_{k,0}^2} - \frac{1}{\tilde{\sigma}_{k,1}^2} \right)^2 \tilde{\sigma}_{k,j}^4 \\ &\quad + \frac{\frac{\tilde{\mu}_k^{4-2\gamma_1}}{\gamma_1^4} \sigma_{\tilde{a}_1}^2 + \frac{\tilde{\mu}_k^{4-4\gamma_1}}{\gamma_1^4} \sigma_{\tilde{b}_1}^2 + 2 \frac{\tilde{\mu}_k^{4-3\gamma_1}}{\gamma_1^4} \sigma_{\tilde{a}_1 \tilde{b}_1}}{\tilde{\sigma}_{k,1}^4} \left(\frac{1}{4} + \frac{3}{4} \frac{\tilde{\sigma}_{k,j}^4}{\tilde{\sigma}_{k,1}^4} \right). \end{aligned} \quad (6.21)$$

Proof. of Proposition 6.2 is given in Appendix 6.7. $\square$

It can be noted that the second term in (6.21) aims to take into account the variability of the ML estimates $(\hat{\tilde{a}}_1, \hat{\tilde{b}}_1)$. In virtue of Lindeberg CLT, under hypothesis $\mathcal{H}_j$, the GLR $\tilde{\Lambda}_{\text{gen}}(\tilde{\mathbf{Z}}) = \sum_{k=1}^K \sum_{i=1}^{n_k} \tilde{\Lambda}_{\text{gen}}(\tilde{z}_{k,i})$ follows the Gaussian distribution as

$$\tilde{\Lambda}_{\text{gen}}(\tilde{\mathbf{Z}}) \xrightarrow{d} \mathcal{N} \left(m_{\text{gen}}^{(j)}, \tilde{v}_{\text{gen}}^{(j)} \right), \quad (6.22)$$

where the expectation $m_{\text{gen}}^{(j)}$ is given in (6.9) and the variance $\tilde{v}_{\text{gen}}^{(j)}$ is given by

$$\begin{aligned} \tilde{v}_{\text{gen}}^{(j)} &= \sum_{k=1}^K \tilde{n}_k \left[\frac{1}{2} \left(\frac{1}{\tilde{\sigma}_{k,0}^2} - \frac{1}{\tilde{\sigma}_{k,1}^2} \right)^2 \tilde{\sigma}_{k,j}^4 \right. \\ &\quad \left. + \frac{\frac{\tilde{\mu}_k^{4-2\gamma_1}}{\gamma_1^4} \sigma_{\tilde{a}_1}^2 + \frac{\tilde{\mu}_k^{4-4\gamma_1}}{\gamma_1^4} \sigma_{\tilde{b}_1}^2 + 2 \frac{\tilde{\mu}_k^{4-3\gamma_1}}{\gamma_1^4} \sigma_{\tilde{a}_1 \tilde{b}_1}}{\tilde{\sigma}_{k,1}^4} \left(\frac{1}{4} + \frac{3}{4} \frac{\tilde{\sigma}_{k,j}^4}{\tilde{\sigma}_{k,1}^4} \right) \right]. \end{aligned} \quad (6.23)$$

The GLRT $\tilde{\delta}_{\text{gen}}^*$ based on the normalized GLR $\tilde{\Lambda}_{\text{gen}}^*(\tilde{\mathbf{Z}}) = \frac{\tilde{\Lambda}_{\text{gen}}(\tilde{\mathbf{Z}}) - \hat{m}_{\text{gen}}^{(0)}}{\sqrt{\hat{v}_{\text{gen}}^{(0)}}}$ is given as follows

$$\tilde{\delta}_{\text{gen}}^*(\tilde{\mathbf{Z}}) = \begin{cases} \mathcal{H}_0 & \text{if } \tilde{\Lambda}_{\text{gen}}^*(\tilde{\mathbf{Z}}) < \tilde{\tau}_{\text{gen}}^* \\ \mathcal{H}_1 & \text{if } \tilde{\Lambda}_{\text{gen}}^*(\tilde{\mathbf{Z}}) \geq \tilde{\tau}_{\text{gen}}^*, \end{cases} \quad (6.24)$$

where $\hat{m}_{\text{gen}}^{(0)}$ and $\hat{v}_{\text{gen}}^{(0)}$ are obtained by replacing unknown parameters $(\tilde{\mu}_k, \tilde{a}_1, \tilde{b}_1)$ by estimates $(\hat{\tilde{\mu}}_k, \hat{\tilde{a}}_1, \hat{\tilde{b}}_1)$ in (6.9) and (6.23). The decision threshold and the power function of the GLRT $\tilde{\delta}_{\text{gen}}^*$ are given in the following theorem.

Theorem 6.2. *For testing against a known camera model $\mathcal{S}_0$, i.e. the parameters $(\tilde{a}_0, \tilde{b}_0, \gamma_0)$ are known, the decision threshold and the power of the GLRT $\tilde{\delta}_{\text{gen}}^*$ are given by*

$$\tilde{\tau}_{\text{gen}}^* = \Phi^{-1}(1 - \alpha_0) \quad (6.25)$$

$$\beta(\tilde{\delta}_{\text{gen}}^*) = 1 - \Phi \left(\frac{m_{\text{gen}}^{(0)} - m_{\text{gen}}^{(1)} + \tilde{\tau}_{\text{gen}}^* \sqrt{\hat{v}_{\text{gen}}^{(0)}}}{\sqrt{\hat{v}_{\text{gen}}^{(1)}}} \right). \quad (6.26)$$

Proof. Based on the Slutsky's theorem, from (6.22), the proof follows immediately. $\square$

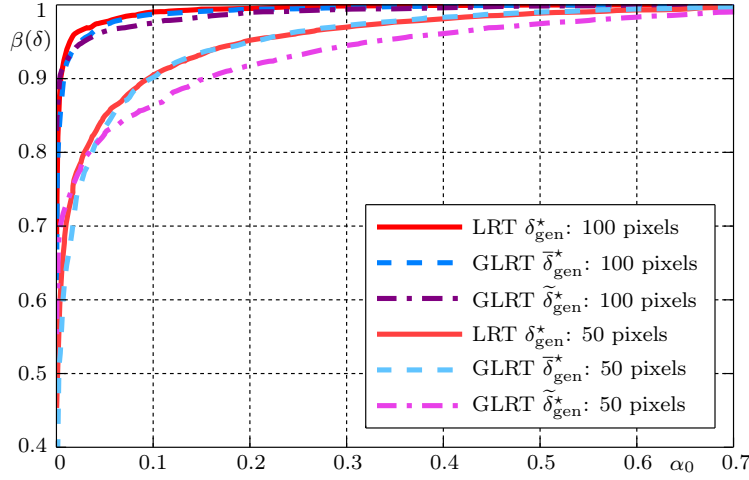


Figure 6.4: Detection performance of the proposed tests for 50 and 100 pixels extracted randomly from simulated JPEG images with quality factor 100.

6.5 Numerical Experiments

6.5.1 Detection Performance on Simulated Database

The detection performance of the proposed tests is first theoretically studied on a simulated database. The camera model $\mathcal{S}_0$ and $\mathcal{S}_1$ are respectively characterized by $(\tilde{a}_0, \tilde{b}_0, \gamma_0) = (-0.0012, 0.11, 0.8)$ and $(\tilde{a}_1, \tilde{b}_1, \gamma_1) = (-0.0025, 0.20, 0.85)$. These parameters respectively correspond to *Nikon D70* and *Nikon D200* camera models in the Dresden image database [168], see Figure 4.2. The camera parameters are used to generate randomly 5000 images for camera model $\mathcal{S}_0$ and 5000 images for camera model $\mathcal{S}_1$ that further are compressed with different quality factors .

The implementation of the GLRT $\tilde{\delta}_{\text{gen}}^*$ requires to know the covariance matrix of ML estimates $(\hat{a}_1, \hat{b}_1)$. However, the ML estimates $(\hat{a}_1, \hat{b}_1)$ are given numerically, which causes a difficulty of establishing the covariance matrix analytically. To overcome this difficulty, it is proposed to estimate $(\hat{a}, \hat{b})$ on each image from 50 images that are taken by the camera model $\mathcal{S}_0$. Then the empirical covariance matrix can be calculated from previous couples $(\hat{a}, \hat{b})$. Speaking rigorously, this is the covariance matrix characterizing the variability of $(\tilde{a}_0, \tilde{b}_0)$. By doing so, it is expected that the parameters $(\hat{a}_1, \hat{b}_1)$ fall into the neighborhood of $(\tilde{a}_0, \tilde{b}_0)$, namely that the inspected image $\tilde{\mathbf{Z}}$ is acquired by the camera model $\mathcal{S}_0$. This leads us to exploit this covariance matrix in the implementation of the GLRT $\tilde{\delta}_{\text{gen}}^*$. This step is also performed in the test with real images.

The detection performance of the proposed tests for different number of pixels is illustrated in Figure 6.4. The very small loss of power between the LRT δ_{gen}^* and the GLRT $\tilde{\delta}_{\text{gen}}^*$ highlights the accuracy of the proposed estimation algorithm. Moreover, from Figure 6.4, it can be noted that the loss of power between the GLRT $\tilde{\delta}_{\text{gen}}^*$ and

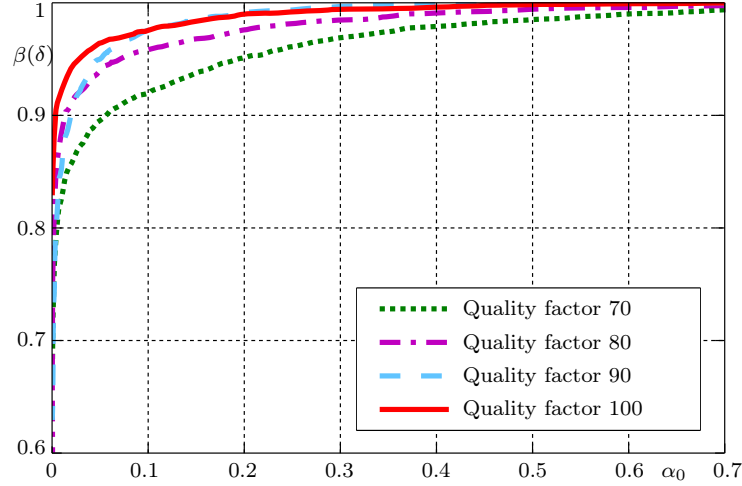


Figure 6.5: Detection performance of the GLRT $\tilde{\delta}_{\text{gen}}^*$ for 100 pixels extracted randomly from simulated JPEG images with different quality factors.

the GLRT $\tilde{\delta}_{\text{gen}}^*$ decreases when the number of pixels increases. For only 50 pixels, there is no sufficient statistics for the convergence of the CLT. Furthermore, the detection performance of the GLRT $\tilde{\delta}_{\text{gen}}^*$ for different quality factors is illustrated in Figure 6.5. As expected, the correct detection probability $\beta(\delta)$ decreases with the decline of the quality factor. Like the GLRTs proposed in Chapter 5, the GLRTs proposed here only require a small number of pixels to achieve a perfect detection performance (e.g. from 500 pixels).

6.5.2 Detection Performance on Two *Nikon D70* and *Nikon D200* Camera Models

To improve the performance of the GLRT when dealing with real images, it is proposed to remove outlier pixels $\tilde{z}_{k,i}$ in each segment $\tilde{S}_k$ using the classical three-sigma rule [175]. The pixel $\tilde{z}_{k,i}$ is considered as non-outlier if the following condition is satisfied: $|\tilde{z}_{k,i}^{\text{res}}| \leq 3\hat{\sigma}_k$. This step is repeated iteratively to obtain better estimates $\hat{\mu}_k$ and $\hat{\sigma}_k^2$. After outlier removal, all remaining pixels are used for the proposed tests.

It is important to remind that the proposed GLRTs are designed in the framework of hypothesis testing theory where the reference camera parameters $(\tilde{a}_0, \tilde{b}_0, \gamma_0)$ under hypothesis $\mathcal{H}_0$ are assumed to be known in advance. Therefore, those parameters need to be defined accurately in practice. To this end, the parameters $(\tilde{a}, \tilde{b}, \gamma)$ are estimated on each image of 50 images. The reference parameter γ_0 is calculated as the average of previous gamma values. Then the parameters $(\tilde{a}, \tilde{b})$ are re-estimated on each image by setting the parameter γ to the averaged γ_0 . The reference parameters $(\tilde{a}_0, \tilde{b}_0)$ are finally obtained by averaging the previous estimates $(\tilde{a}, \tilde{b})$.

To highlight the relevance of the proposed GLRTs, two *Nikon D70* and *Nikon*

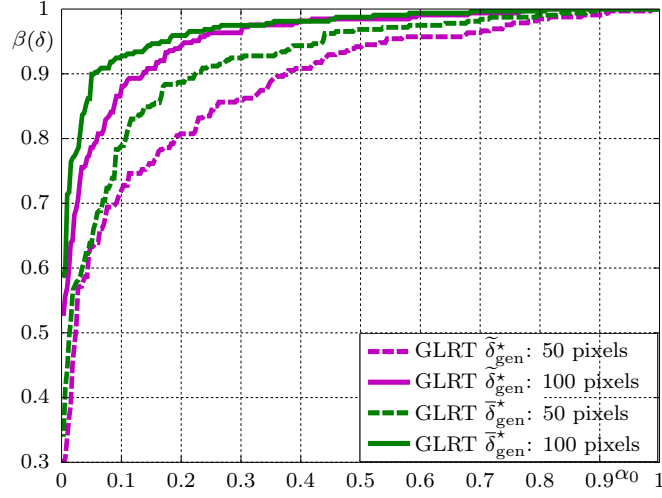


Figure 6.6: Detection performance of the GLRT $\bar{\delta}_{gen}^*$ and $\tilde{\delta}_{gen}^*$ for 50 and 100 pixels extracted randomly from JPEG images of *Nikon D70* and *Nikon D200* cameras.

D200 camera models of the Dresden image database [168] are chosen to conduct experiments. Only the red color channel is used in this experiment. The *Nikon D70* and *Nikon D200* cameras are respectively set at $\mathcal{H}_0$ and $\mathcal{H}_1$. The Figure 6.6 shows the detection performance of the GLRT $\bar{\delta}_{gen}^*$ and $\tilde{\delta}_{gen}^*$ for different numbers of pixels. We can note a similar behavior to the ROC on simulated database. There is a small loss of power between the two power functions since the GLRT $\tilde{\delta}_{gen}^*$ takes into account different estimates $(\hat{a}_1, \hat{b}_1)$ that are influenced by the image content. Nevertheless, two proposed GLRTs are almost perfect from 500 pixels. Besides, the Figure 6.7 shows the comparison between the theoretical and empirical false alarm probability, which are plotted as a function of decision threshold τ . The two proposed GLRTs $\bar{\delta}_{gen}^*$ and $\tilde{\delta}_{gen}^*$ also show an ability of guaranteeing a prescribed false alarm rate.

6.5.3 Detection Performance on a Large Image Database

Experiments are then conducted on the Dresden image database [168] to verify the efficiency of the proposed approach. Technical specifications of the cameras are shown in Table 6.1. In these experiments, the GLRT $\tilde{\delta}_{gen}^*$ is used to verify whether a given image is acquired by the camera model of interest. The decision threshold $\tilde{\tau}_{gen}^*$ is given by the Theorem 6.2 corresponding to the false alarm rate $\alpha_0 = 10^{-5}$. The detection performance of the proposed detector $\tilde{\delta}_{gen}^*$ is shown in Table 6.2. A SVM-based detector has been already performed on the Dresden database [168] using 46 different features to capture characteristics of different camera components of a digital camera. This SVM-based detector has used 60% of the images of one device per model for training and all images of the remaining devices for testing. The detection performance of the SVM-based detector is shown in Table 6.3. The

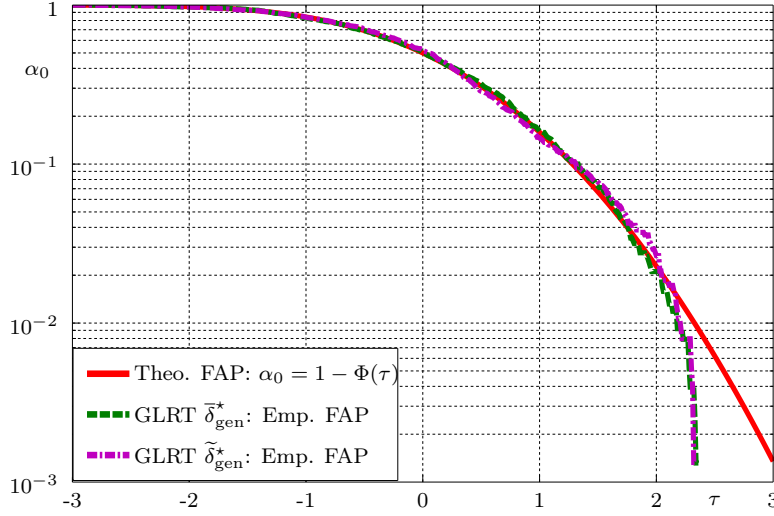


Figure 6.7: Comparison between the theoretical false alarm probability (FAP) and the empirical FAP, plotted as a function of decision threshold τ .

PRNU-based detector [32] is also performed in these experiments. This detector [32] is only performed on one device per model. Its detection performance is shown in Table 6.4. It can be noted that the proposed detector slightly outperforms prior-art ones.

6.6 Conclusion

The goal of this chapter is to design a statistical test within the framework of hypothesis testing theory for camera model identification from JPEG images. The approach is based on the generalized noise model that accurately characterizes a natural image in TIFF or JPEG format acquired by a digital camera. The parameters $(\tilde{a}, \tilde{b}, \gamma)$ characterizing the generalized noise model are exploited as camera fingerprint for camera model identification. Based on this accurate image model, the most powerful LRT are studied in the theoretical context and two GLRTs are designed to deal with the difficulty of unknown parameters. The strength of the proposed approach is that statistical performance of the tests can be analytically established as well as they can warrant a prescribed false alarm rate while ensuring a high detection performance. This chapter also completes the limitations of the GLRTs proposed in Chapter 5.

A limitation in this approach is that the demosaicing and white balancing are not completely modeled. The proposed approach only assumes that those operations are linear, which simplifies the statistical study and makes the resulting model more exploitable. The demosaicing always involves a convolution operation, which requires us to consider a multivariate distribution. That distribution could be more relevant but inexploitable. Overall, the common philosophy in this thesis is a trade-

Camera Model		No. devices	Resolution	No. images
Canon Ixus 55	Cn1	1	2592×1944	200
Canon Ixus 70	Cn2	3	3072×2304	500
Casio EX-Z150	Cs	5	3264×2448	600
Fujifilm J50	F	3	3264×2448	500
Nikon D200	N2	2	3872×2592	500
Nikon D70	N3	2	3008×2000	300
Olympus 1050SW	O	5	3648×2736	700
Panasonic FZ50	Pa	3	3648×2736	600
Pentax A40	Pe	4	4000×3000	600
Praktica DCZ	Pr	5	2560×1920	700
Ricoh GX100	Ri	5	3648×2736	700
Σ	11	38		5900

Table 6.1: Camera Model Used in Experiments

		Inspected images											Avg
		Cn1	Cn2	Cs	F	N2	N3	O	Pa	Pe	Pr	Ri	
$\mathcal{H}_0$	Cn1	98.41	0	0	0	0	0	0	0	0	0	0	
	Cn2	0	100	0	0	0	0	0	0	0	0	0	
	Cs	0	0	99.82	0	0	0	0	0	0	0	0	
	F	0	0	0	91.55	0	0	0	0	0	0	0	
	N2	2.26	0	0	0	100	0	0	0	0	0	0	
	N3	0	0	0	0	0	100	0	0	0	0	0	
	O	0	0	0	0	0	0	97.25	0	0	0	0	
	Pa	0	0	0	0	0	0	0	98.57	0	0	0	
	Pe	0	0	0	0	0	0	0	0	92.21	0	0	
	Pr	0	0	0	0	0	0	0	0	0	100	0	
	Ri	0	0	0	0	0	0	0	0	0	0	98.33	
													97.83

Table 6.2: Performance of proposed detector

off between the reality in image acquisition and exploitability in practice.

6.7 Appendix

Since the variance of $\hat{\mu}_k$ is negligible, the estimate $\hat{\mu}_k$ can be treated as a constant $\tilde{\mu}_k$. Moreover, the estimates $(\hat{a}_1, \hat{b}_1)$ are asymptotically consistent. Therefore, the mathematical expectation of the GLR $\tilde{\Lambda}_{\text{gen}}(\tilde{z}_{k,i})$ under hypothesis $\mathcal{H}_j$ does not change

$$\mathbb{E}_{\mathcal{H}_j}[\tilde{\Lambda}_{\text{gen}}(\tilde{z}_{k,i})] = \frac{1}{2} \left(\log(\tilde{\sigma}_{k,0}^2) - \log(\tilde{\sigma}_{k,1}^2) \right) + \frac{1}{2} \left(\frac{1}{\tilde{\sigma}_{k,0}^2} - \frac{1}{\tilde{\sigma}_{k,1}^2} \right) \tilde{\sigma}_{k,j}^2. \quad (6.27)$$

		Inspected images										Avg
		Cn2	Cs	F	N2	N3	O	Pa	Pe	Pr	Ri	
$\mathcal{H}_0$	Cn2	*	98.26	*	2.61	*	*	*	*	*	*	96.17
	Cs	*	*	99.71	2.77	*	*	*	*	*	*	
	F	*	*	*	99.94	*	*	*	*	*	*	
	N2	*	*	*	93.73	*	*	*	*	*	*	
	N3	*	*	*	*	84.57	*	*	*	*	*	
	O	*	*	*	*	*	98.95	*	*	*	*	
	Pa	*	*	*	*	*	*	97.31	*	*	*	
	Pe	*	*	*	*	*	*	*	90.86	*	*	
	Pr	*	*	*	*	*	*	*	*	98.42	*	
	Ri	*	*	*	*	*	*	*	*	*	100	
												96.17

Table 6.3: Performance of SVM-based detector (the symbol * represents values smaller than 2%)

		Inspected images											
		Cn1	Cn2	Cs	F	N2	N3	O	Pa	Pe	Pr	Ri	Avg
$\mathcal{H}_0$	Cn1	100	0	0	0	0	0	0	0	0	0	0	
	Cn2	0	100	0	0	0	0	0	0	0	0	0	
	Cs	0	0	99.44	0	0	0	0	0	0	0	0	
	F	0	0	0	73.81	0	0	0	0	0	0	0	
	N2	0	0	0	0	98.92	0.33	0.28	0	0	0	0	
	N3	0	0	0	0	0	100	0	0	0	0	0	
	O	0	0	0	0	0	0.16	87.25	0	0	0	0	
	Pa	0	0	0	0	0	0	0	94.71	0	0	0	
	Pe	0	0	0	0	0	0	0	0	100	0	0	
	Pr	0	0	0	0	0	0	0	0	0	96.65	0	
Ri	0	0	0	0	0	0	0	0	0	0	100		
													95.52

Table 6.4: Performance of PRNU-based detector

Meanwhile, the variance of $f(\hat{\mu}_k; \hat{a}_1, \hat{b}_1, \gamma_1)$ is given by

$$\begin{aligned}
\mathbb{V}\text{ar}_{\mathcal{H}_j} \left[f(\hat{\mu}_k; \hat{a}_1, \hat{b}_1, \gamma_1) \right] &= \mathbb{V}\text{ar}_{\mathcal{H}_j} \left[f(\tilde{\mu}_k; \hat{a}_1, \hat{b}_1, \gamma_1) \right] = \mathbb{V}\text{ar}_{\mathcal{H}_j} \left[\frac{1}{\gamma_1^2} \hat{a}_1 \tilde{\mu}_k^{2-\gamma_1} + \frac{1}{\gamma_1^2} \hat{b}_1 \tilde{\mu}_k^{2-2\gamma_1} \right] \\
&= \frac{\tilde{\mu}_k^{4-2\gamma_1}}{\gamma_1^4} \mathbb{V}\text{ar}_{\mathcal{H}_j} \left[\hat{a}_1 \right] + \frac{\tilde{\mu}_k^{4-2\gamma_1}}{\gamma_1^4} \mathbb{V}\text{ar}_{\mathcal{H}_j} \left[\hat{b}_1 \right] + \frac{2\tilde{\mu}_k^{4-3\gamma_1}}{\gamma_1^4} \text{Cov}_{\mathcal{H}_j} \left[\hat{a}_1; \hat{b}_1 \right] \\
&= \frac{\tilde{\mu}_k^{4-2\gamma_1}}{\gamma_1^4} \sigma_{\hat{a}_1}^2 + \frac{\tilde{\mu}_k^{4-4\gamma_1}}{\gamma_1^4} \sigma_{\hat{b}_1}^2 + 2 \frac{\tilde{\mu}_k^{4-3\gamma_1}}{\gamma_1^4} \sigma_{\hat{a}_1 \hat{b}_1}. \tag{6.28}
\end{aligned}$$

Therefore, we derive from Corollary A.1 that

$$\begin{aligned}
&\mathbb{V}\text{ar}_{\mathcal{H}_j} \left[\log (f(\hat{\mu}_k; \tilde{a}_0, \tilde{b}_0, \gamma_0)) - \log (f(\hat{\mu}_k; \tilde{a}_0, \tilde{b}_0, \gamma_0)) \right] \\
&= \mathbb{V}\text{ar}_{\mathcal{H}_j} \left[\log (f(\tilde{\mu}_k; \hat{a}_1, \hat{b}_1, \gamma_1)) \right] \approx \frac{\mathbb{V}\text{ar}_{\mathcal{H}_j} \left[f(\tilde{\mu}_k; \hat{a}_1, \hat{b}_1, \gamma_1) \right]}{f^2(\tilde{\mu}_k; \tilde{a}_1, \tilde{b}_1, \gamma_1)} \\
&= \frac{1}{\tilde{\sigma}_{k,1}^4} \left(\frac{\tilde{\mu}_k^{4-2\gamma_1}}{\gamma_1^4} \sigma_{\hat{a}_1}^2 + \frac{\tilde{\mu}_k^{4-4\gamma_1}}{\gamma_1^4} \sigma_{\hat{b}_1}^2 + 2 \frac{\tilde{\mu}_k^{4-3\gamma_1}}{\gamma_1^4} \sigma_{\hat{a}_1 \hat{b}_1} \right). \tag{6.29}
\end{aligned}$$

and

$$\begin{aligned}
 & \mathbb{V}\text{ar}_{\mathcal{H}_j} \left[\frac{1}{f(\hat{\mu}_k; \tilde{a}_0, \tilde{b}_0, \gamma_0)} - \frac{1}{f(\hat{\mu}_k; \hat{a}_1, \hat{b}_1, \gamma_1)} \right] \\
 &= \mathbb{V}\text{ar}_{\mathcal{H}_j} \left[\frac{1}{f(\tilde{\mu}_k; \hat{a}_1, \hat{b}_1, \gamma_1)} \right] \approx \frac{\mathbb{V}\text{ar}_{\mathcal{H}_j} [f(\tilde{\mu}_k; \hat{a}_1, \hat{b}_1, \gamma_1)]}{f^4(\tilde{\mu}_k; \hat{a}_1, \hat{b}_1, \gamma_1)} \\
 &= \frac{1}{\tilde{\sigma}_{k,1}^8} \left(\frac{\tilde{\mu}_k^{4-2\gamma_1}}{\gamma_1^4} \sigma_{\tilde{a}_1}^2 + \frac{\tilde{\mu}_k^{4-4\gamma_1}}{\gamma_1^4} \sigma_{\tilde{b}_1}^2 + 2 \frac{\tilde{\mu}_k^{4-3\gamma_1}}{\gamma_1^4} \sigma_{\tilde{a}_1 \tilde{b}_1} \right). \quad (6.30)
 \end{aligned}$$

Consequently, the asymptotic variance of the GLR $\tilde{\Lambda}_{\text{gen}}(\tilde{z}_{k,i})$ is given by

$$\begin{aligned}
 \mathbb{V}\text{ar}_{\mathcal{H}_j} [\tilde{\Lambda}_{\text{gen}}(\tilde{z}_{k,i})] &= \frac{1}{4} \mathbb{V}\text{ar}_{\mathcal{H}_j} \left[\log(f(\hat{\mu}_k; \tilde{a}_0, \tilde{b}_0, \gamma_0)) - \log(f(\hat{\mu}_k; \hat{a}_1, \hat{b}_1, \gamma_1)) \right] \\
 &+ \frac{1}{4} \mathbb{E}_{\mathcal{H}_j}^2 \left[\frac{1}{f(\hat{\mu}_k; \tilde{a}_0, \tilde{b}_0, \gamma_0)} - \frac{1}{f(\hat{\mu}_k; \hat{a}_1, \hat{b}_1, \gamma_1)} \right] \mathbb{V}\text{ar}_{\mathcal{H}_j} [(\tilde{z}_{k,i} - \hat{\mu}_k)^2] \\
 &+ \frac{1}{4} \mathbb{V}\text{ar}_{\mathcal{H}_j} \left[\frac{1}{f(\hat{\mu}_k; \tilde{a}_0, \tilde{b}_0, \gamma_0)} - \frac{1}{f(\hat{\mu}_k; \hat{a}_1, \hat{b}_1, \gamma_1)} \right] \mathbb{E}_{\mathcal{H}_j} [(\tilde{z}_{k,i} - \hat{\mu}_k)^4] \\
 &= \frac{1}{2} \left(\frac{1}{\tilde{\sigma}_{k,0}^2} - \frac{1}{\tilde{\sigma}_{k,1}^2} \right)^2 \tilde{\sigma}_{k,j}^4 \\
 &+ \frac{\frac{\tilde{\mu}_k^{4-2\gamma_1}}{\gamma_1^4} \sigma_{\tilde{a}_1}^2 + \frac{\tilde{\mu}_k^{4-4\gamma_1}}{\gamma_1^4} \sigma_{\tilde{b}_1}^2 + 2 \frac{\tilde{\mu}_k^{4-3\gamma_1}}{\gamma_1^4} \sigma_{\tilde{a}_1 \tilde{b}_1}}{\tilde{\sigma}_{k,1}^4} \left(\frac{1}{4} + \frac{3}{4} \frac{\tilde{\sigma}_{k,j}^4}{\tilde{\sigma}_{k,1}^4} \right). \quad (6.31)
 \end{aligned}$$

Camera Model Identification Based on DCT Coefficient Statistics

Contents

7.1	Introduction	116
7.2	Camera Fingerprint	116
7.2.1	Design of Camera Fingerprint	116
7.2.2	Extraction of Camera Fingerprint	117
7.2.3	Property of Camera Fingerprint	119
7.3	Optimal Detector for Camera Model Identification Problem	120
7.3.1	Hypothesis Testing Formulation	120
7.3.2	LRT for Two Simple Hypotheses	121
7.4	Practical Context: GLRT	123
7.4.1	GLRT with Unknown Parameters α_k	123
7.4.2	GLRT with Unknown Parameters $(\alpha_k, \tilde{c}_{k,1}, \tilde{d}_{k,1})$	124
7.5	Numerical Experiments	126
7.5.1	Detection Performance on Simulated Database	127
7.5.2	Detection Performance on Two <i>Canon Ixus 70</i> and <i>Nikon D200</i> Camera Models	128
7.5.3	Detection Performance on a Large Image Database	129
7.6	Conclusion	132
7.7	Appendix	132
7.7.1	Relation between the Parameters $(\tilde{a}, \tilde{b}, \gamma)$ and $(\alpha_{u,v}, \beta_{u,v})$	132
7.7.2	Laplace's Approximation of DCT Coefficient Model	133
7.7.3	Expectation and Variance of the LR $\Lambda_{\text{dct}}(I_{k,i})$ under Hypoth- esis $\mathcal{H}_j$	134
7.7.4	Asymptotic Expectation and Variance of the GLR $\tilde{\Lambda}_{\text{dct}}(I_{k,i})$ under Hypothesis $\mathcal{H}_j$	135

7.1 Introduction

This chapter proposes a novel approach for camera model identification. Instead of extracting the camera fingerprint in the spatial domain as presented in Chapter 5 and 6, the approach proposed in this chapter relies on the state-of-the-art model of DCT coefficients in order to extract the fingerprint in the DCT domain. The main motivation behind fingerprint extraction in the DCT domain is that the statistics of DCT coefficients change with different sensor noises combining with various in-camera processing algorithms. Relying on the highly accurate model of DCT coefficients allows us to capture this statistical difference in natural images taken by different camera models. The goal of this chapter is to design a statistical test for camera model identification from JPEG images based on the proposed model of DCT coefficients. The main contributions are the following:

- This chapter presents a novel camera fingerprint that is extracted in the DCT domain based on the state-of-the-art statistical model of DCT coefficients (see Section 4.4). The proposed camera fingerprint is invariant to image content and camera settings, and robust to non-linear processing operations (e.g. gamma correction).
- Still stating the camera model identification problem into hypothesis testing framework, this chapter studies the most powerful test given by the LRT in the ideal context where all model parameters are known and designs two GLRTs to deal with the difficulty of unknown parameters. The statistical performance of the proposed tests is analytically established. The proposed tests allow the guaranteeing of a prescribed false-alarm rate and the setting of decision threshold independently of the image content.

The work in this chapter has been presented in [178] and valorized in the patent [179]. The chapter is organized as follows. Section 7.2 designs the camera fingerprint that is further exploited for camera model identification, and proposes an algorithm for camera fingerprint extraction in the DCT domain. Section 7.3 states the camera model identification problem in the framework of hypothesis testing theory and studies the LRT assuming that all model parameters are known in advance. In practice, those parameters are unknown when inspecting a digital image. Section 7.4 designs two GLRTs to address the difficulty of unknown parameters. Section 7.5 presents numerical results of two proposed GLRTs on simulated and real JPEG images. Finally, Section 7.6 concludes the chapter.

7.2 Camera Fingerprint

7.2.1 Design of Camera Fingerprint

The study of image statistics in the spatial domain and DCT domain has been performed in Chapter 4. The parameters (a, b) and $(\tilde{a}, \tilde{b}, \gamma)$, which are extracted in the

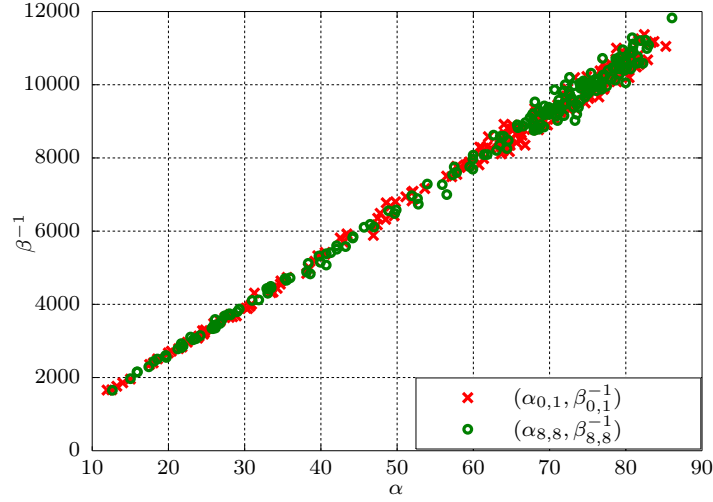


Figure 7.1: Estimated parameters (α, β) at frequency $(0, 1)$ and $(8, 8)$ of uniform images generated using $\tilde{a} = 0.1$, $\tilde{b} = 2$, $\gamma = 2.2$.

spatial domain, are exploited as camera fingerprint for camera model identification. In other words, the parameters $(\tilde{a}, \tilde{b}, \gamma)$ contain information about camera model. After transforming into the DCT domain, this information is expanded over different frequencies. Moreover, the statistics of DCT coefficients can be accurately captured by the model proposed in Section 4.4, which is characterized by the parameters (α, β) . Therefore, it is proposed to establish the relation between the parameters $(\tilde{a}, \tilde{b}, \gamma)$ and (α, β) to capture such information in the DCT domain. For the sake of simplification, this relation is given by

$$\beta^{-1} = \tilde{c}\alpha + \tilde{d}, \quad (7.1)$$

where the parameters $(\tilde{c}, \tilde{d})$ depend on $(\tilde{a}, \tilde{b}, \gamma)$ (see more details in Appendix 7.7.1). This suggests that the parameters $(\tilde{c}, \tilde{d})$ can be also used for camera model identification. It can be said that while the relations (4.1) and (4.42) characterize the non-stationarity of noise in the spatial domain, the relation (7.1) characterizes this property in the DCT domain. The relation (7.1) can capture the difference of noise statistics in natural images taken by different camera models. Moreover, the linearity in relation (7.1) can facilitate the estimation of the parameters $(\tilde{c}, \tilde{d})$. It should be noted that in an image whose each 8×8 block is uniform, the same parameters (α, β) and $(\tilde{c}, \tilde{d})$ are shared among DCT coefficients at different frequencies. The relation (7.1) on such images is illustrated in Figure 7.1.

7.2.2 Extraction of Camera Fingerprint

An important requirement when using the parameters $(\tilde{c}, \tilde{d})$ as camera fingerprint is that they should be invariant to image content. Furthermore, to guarantee the whole statistical analysis from RAW format to JPEG format, it is necessary to work on homogeneous blocks. These considerations are addressed in this section.

Because of heterogeneity and noise non-stationarity in a natural image, the energy tends to be more located in lower frequencies. Consequently, DCT coefficients at different frequencies do not share the same parameters (α, β) and $(\tilde{c}, \tilde{d})$. Therefore, the estimation of parameters (α, β) and $(\tilde{c}, \tilde{d})$ should be performed on each frequency separately. The DCT coefficients are arranged into 64 vectors of coefficients according to the zig-zag order. Let $\mathbf{I}_k = (I_{k,1}, \dots, I_{k,N_{\text{blk}}})$ with $k \in \{1, \dots, 64\}$, be the vector of length N_{blk} that contains coefficients at the frequency k . Analogously, let denote (α_k, β_k) and $(\tilde{c}_k, \tilde{d}_k)$ the parameters of the AC coefficients $\mathbf{I}_k$.

The proposed algorithm for estimation of parameters $(\tilde{c}_k, \tilde{d}_k)$ consists of three fundamental steps: image denoising, homogeneous block detection, and Least-Squares (LS) estimation [161]. Image denoising step aims to attenuate the impact of image content. The detection of homogeneous blocks is performed subsequently to provide appropriate sample data for parameter estimation. The LS approach is applied straightforwardly as the relation (7.1) is linear.

Let $\tilde{\mathbf{Z}}$ be a two-dimensional matrix representing a natural image in TIFF or JPEG format in the spatial domain. To remove the image content, a denoising filter $\mathfrak{D}$ is employed so that the residual image $\tilde{\mathbf{Z}}^{\text{res}}$ is given by

$$\tilde{\mathbf{Z}}^{\text{res}} = \tilde{\mathbf{Z}} - \mathfrak{D}(\tilde{\mathbf{Z}}). \quad (7.2)$$

If $\tilde{\mathbf{Z}}$ is a color image, the denoising filter $\mathfrak{D}$ is performed on each color component, then three residual components are combined into one residual image using the usual conversion from RGB to grayscale

$$\tilde{\mathbf{Z}}^{\text{res}} = 0.2989 \cdot \tilde{\mathbf{Z}}_{\text{R}}^{\text{res}} + 0.587 \cdot \tilde{\mathbf{Z}}_{\text{G}}^{\text{res}} + 0.114 \cdot \tilde{\mathbf{Z}}_{\text{B}}^{\text{res}}, \quad (7.3)$$

where $\tilde{\mathbf{Z}}_{\text{R}}^{\text{res}}$, $\tilde{\mathbf{Z}}_{\text{G}}^{\text{res}}$, and $\tilde{\mathbf{Z}}_{\text{B}}^{\text{res}}$ denote respectively residuals of red, green, and blue component. The residual image $\tilde{\mathbf{Z}}^{\text{res}}$ is then transformed into the DCT domain

$$\mathbf{I} = \text{DCT}(\tilde{\mathbf{Z}}^{\text{res}}), \quad (7.4)$$

where $\mathbf{I}$ is the image of DCT coefficients of the residual image $\tilde{\mathbf{Z}}^{\text{res}}$.

For homogeneous block detection, it is proposed to calculate the standard deviation of each block and compare it with a threshold ϑ . The empirical standard deviation of the block u is calculated in the DCT domain as

$$\varsigma_u = \text{MAD}(I_{2,u}, \dots, I_{64,u}). \quad (7.5)$$

The DC coefficient $I_{1,u}$ is excluded in the calculation of ς_u . The block u is selected if the standard deviation ς_u is smaller than the threshold ϑ . Let denote N_b as the number of selected homogeneous blocks.

Suppose that M couples $(\hat{\alpha}_{k,m}, \hat{\beta}_{k,m})$, $m \in \{1, \dots, M\}$, are available, the LS estimates of the parameters $(\tilde{c}_k, \tilde{d}_k)$ are given by

$$\begin{pmatrix} \hat{\tilde{c}}_k \\ \hat{\tilde{d}}_k \end{pmatrix} = (\mathbf{H}_k^T \mathbf{H}_k)^{-1} \mathbf{H}_k^T \mathbf{v}_k \quad (7.6)$$

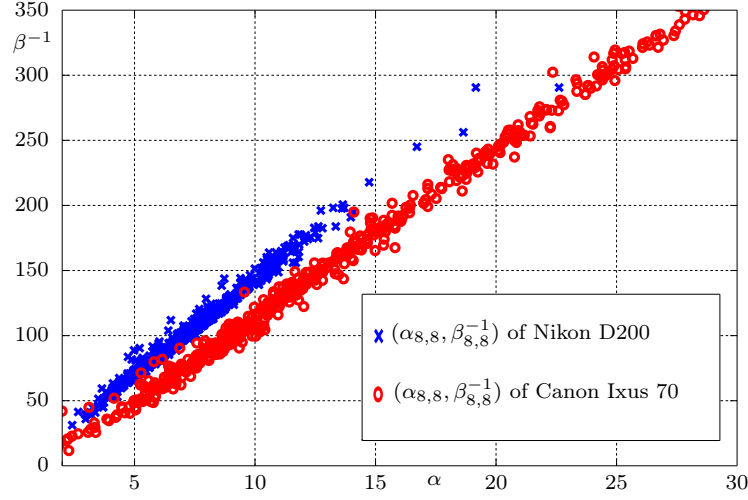


Figure 7.2: Estimated parameters (α, β) at frequency $(8, 8)$ of natural JPEG images issued from *Canon Ixus 70* and *Nikon D200* camera models.

with

$$\mathbf{H}_k = \begin{pmatrix} \hat{\alpha}_{k,1} & 1 \\ \vdots & \vdots \\ \hat{\alpha}_{k,M} & 1 \end{pmatrix} \quad \text{and} \quad \mathbf{v}_k = \begin{pmatrix} \hat{\beta}_{k,1}^{-1} \\ \vdots \\ \hat{\beta}_{k,M}^{-1} \end{pmatrix},$$

The LS estimates $(\hat{c}_k, \hat{d}_k)$ are unbiased and asymptotically equivalent to ML estimates in large samples [161].

As showed above, the LS approach requires several couples (α_k, β_k) for estimation of parameters $(\tilde{c}_k, \tilde{d}_k)$. One can collect M images and estimate a couple $(\alpha_{k,m}, \beta_{k,m})$ from all homogeneous blocks of each image following the ML approach (see Section 4.4.2). However, from the practical point of view, it is necessary to estimate the parameters $(\tilde{c}_k, \tilde{d}_k)$ from a single image. This is accomplished by extracting randomly a subset of homogeneous blocks from N_b blocks, then performing the ML estimation of parameters $(\alpha_{k,m}, \beta_{k,m})$ on this subset.

7.2.3 Property of Camera Fingerprint

When the image content is removed perfectly, the parameters $(\tilde{c}_k, \tilde{d}_k)$ remain identical for images with different image contents. However, in practice, due to the fact that the perfect denoising filter $\mathcal{D}$ is difficult to obtain, the DCT coefficients at low frequencies are still affected by image content. Meanwhile, the coefficients at high frequencies contain mostly noises because of the energy compaction property of DCT operation. Thus they are more relevant to exploit for camera model identification. Figure 7.2 shows the the linear relation (7.1) at frequency $(8, 8)$ of natural JPEG images taken by *Canon Ixus 70* and *Nikon D200* camera models. It should be noted that each point (α, β) in Figure 7.2 corresponds to one image. Figure 7.2 involves the JPEG images with different imaged scenes, different camera settings,

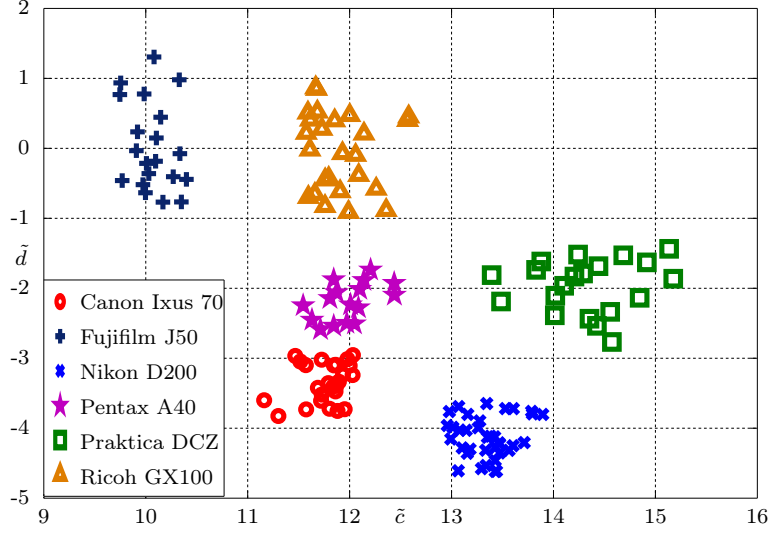


Figure 7.3: Estimated parameters $(\tilde{c}, \tilde{d})$ at frequency $(8, 8)$ of natural JPEG images issued from different camera models in Dresden database.

different devices per model, and different environmental conditions. This indicates that the parameters $(\tilde{c}_k, \tilde{d}_k)$ remain similar under those conditions.

Moreover, for camera model identification problem, it is necessary to verify the discriminability of parameters $(\tilde{c}_k, \tilde{d}_k)$ for different camera models. The Figure 7.3 shows the parameters $(\tilde{c}_k, \tilde{d}_k)$ estimated from JPEG images at frequency $(8, 8)$ for different camera models. This figure clearly shows their discriminability between different camera models. Therefore, the parameters $(\tilde{c}_k, \tilde{d}_k)$ are exploited as camera fingerprint to identify camera models in this chapter.

7.3 Optimal Detector for Camera Model Identification Problem

7.3.1 Hypothesis Testing Formulation

The camera model $\mathcal{S}_j$, $j \in \{0, 1\}$, is now characterized by the parameters $(\tilde{c}_{k,j}, \tilde{d}_{k,j})$, $k \in \{1, \dots, K\}$, where K is the number of usable frequencies for camera model identification. It is assumed that $(\tilde{c}_{k,0}, \tilde{d}_{k,0}) \neq (\tilde{c}_{k,1}, \tilde{d}_{k,1})$. The goal of the test is to decide between two hypotheses defined by

$$\begin{cases} \mathcal{H}_0 = \left\{ I_{k,i} \sim P_{\theta_{k,0}}, \beta_{k,0}^{-1} = \tilde{c}_{k,0}\alpha_k + \tilde{d}_{k,0}, \forall k \in \{1, \dots, K\}, \forall i \in \{1, \dots, N_b\} \right\} \\ \mathcal{H}_1 = \left\{ I_{k,i} \sim P_{\theta_{k,1}}, \beta_{k,1}^{-1} = \tilde{c}_{k,1}\alpha_k + \tilde{d}_{k,1}, \forall k \in \{1, \dots, K\}, \forall i \in \{1, \dots, N_b\} \right\}, \end{cases} \quad (7.7)$$

where $P_{\theta_{k,j}}, \theta_{k,j} = (\alpha_k, \tilde{c}_{k,j}, \tilde{d}_{k,j})$, denotes the probability distribution of DCT coefficients $I_{k,i}$ under hypothesis $\mathcal{H}_j$. Let

$$\mathcal{K}_{\alpha_0} = \left\{ \delta : \sup_{\theta_0} \mathbb{P}_{\mathcal{H}_0} [\delta(\tilde{\mathbf{Z}}) = \mathcal{H}_1] \leq \alpha_0 \right\}$$

be the class of tests whose the false alarm probability is upper-bounded by the prescribed rate α_0 where $\theta_0 = (\theta_{1,0}, \dots, \theta_{K,0})$ is the vector containing all parameters. Among all the tests in the class $\mathcal{K}_{\alpha_0}$, it is aimed at finding a test δ which maximizes the power function $\beta(\delta)$. The camera parameters $(\tilde{c}_{k,0}, \tilde{d}_{k,0})$ are assumed to be known in advance. The main goal of this chapter is to study the LRT and to design the GLRTs to address the difficulty of unknown parameters $(\alpha_k, \tilde{c}_{k,1}, \tilde{d}_{k,1})$.

7.3.2 LRT for Two Simple Hypotheses

When all model parameters are known, in virtue of the Neyman-Pearson lemma, the MP test δ_{dct}^* solving the problem (7.7) is the LRT given by the following decision rule

$$\delta_{\text{dct}}^*(\tilde{\mathbf{Z}}) = \begin{cases} \mathcal{H}_0 & \text{if } \Lambda_{\text{dct}}(\tilde{\mathbf{Z}}) = \sum_{k=1}^K \sum_{i=1}^{N_b} \Lambda_{\text{dct}}(I_{k,i}) < \tau_{\text{dct}}^* \\ \mathcal{H}_1 & \text{if } \Lambda_{\text{dct}}(\tilde{\mathbf{Z}}) = \sum_{k=1}^K \sum_{i=1}^{N_b} \Lambda_{\text{dct}}(I_{k,i}) \geq \tau_{\text{dct}}^* \end{cases} \quad (7.8)$$

where the LR $\Lambda_{\text{dct}}(I_{k,i})$ is defined as

$$\Lambda_{\text{dct}}(I_{k,i}) = \log \frac{P_{\theta_{k,1}}[I_{k,i}]}{P_{\theta_{k,0}}[I_{k,i}]}, \quad (7.9)$$

assuming that the DCT coefficients are statistically independent. From (4.70), it can be noted that the expression of the LR $\Lambda_{\text{dct}}(I_{k,i})$ is difficult to exploit for subsequent stages, e.g. the design of the GLRT and analytic establishment of its statistical performance. Therefore it is proposed to simplify the LR $\Lambda_{\text{dct}}(I_{k,i})$ to facilitate the study in the manner that it does not cause any loss of optimality.

Using the Laplace's approximation [180, 181] (see more details in Appendix 7.7.2), the pdf $f_I(x)$ (4.70) can be approximated as

$$f_I(x) \approx \frac{|x|^{\alpha-1}}{(2\beta)^{\frac{\alpha}{2}} \Gamma(\alpha)} \exp\left(-|x| \sqrt{\frac{2}{\beta}}\right). \quad (7.10)$$

Consequently, the LR $\Lambda_{\text{dct}}(I_{k,i})$ can be given as

$$\begin{aligned} \Lambda_{\text{dct}}(I_{k,i}) &= \log \frac{\frac{|I_{k,i}|^{\alpha_k-1}}{(2\beta_{k,1})^{\frac{\alpha_k}{2}} \Gamma(\alpha_k)} \exp\left(-|I_{k,i}| \sqrt{\frac{2}{\beta_{k,1}}}\right)}{\frac{|I_{k,i}|^{\alpha_k-1}}{(2\beta_{k,0})^{\frac{\alpha_k}{2}} \Gamma(\alpha_k)} \exp\left(-|I_{k,i}| \sqrt{\frac{2}{\beta_{k,0}}}\right)} \\ &= \frac{\alpha_k}{2} \log \frac{\beta_{k,1}^{-1}}{\beta_{k,0}^{-1}} - \sqrt{2} |I_{k,i}| \left(\sqrt{\beta_{k,1}^{-1}} - \sqrt{\beta_{k,0}^{-1}} \right). \end{aligned} \quad (7.11)$$

It should be noted that other polynomial expansions for the modified Bessel function $K_\nu(x)$ are provided in [170], so a polynomial approximation of $f_I(x)$ can be derived. However, those approximations are not considered in this chapter. The main advantage of the Laplace's approximation (7.10) is to provide an approximation of the form of exponential family function, which allows us to simplify the expression of the LR $\Lambda_{\text{dct}}(I_{k,i})$. The approximating function (7.10) is used only for simplification of the LR. The estimation of parameters (α_k, β_k) is always based on the exact function (4.70).

Proposition 7.1. *Under hypothesis $\mathcal{H}_j$, the first two moments of the LR $\Lambda_{\text{dct}}(I_{k,i})$ are given by*

$$\mathbb{E}_{\mathcal{H}_j}[\Lambda_{\text{dct}}(I_{k,i})] = \frac{\alpha_k}{2} \log \frac{\beta_{k,1}^{-1}}{\beta_{k,0}^{-1}} - \frac{2}{\sqrt{\pi}} \beta_{k,j}^{\frac{1}{2}} \frac{\Gamma(\alpha_k + \frac{1}{2})}{\Gamma(\alpha_k)} \left(\sqrt{\beta_{k,1}^{-1}} - \sqrt{\beta_{k,0}^{-1}} \right) \quad (7.12)$$

$$\text{Var}_{\mathcal{H}_j}[\Lambda_{\text{dct}}(I_{k,i})] = 2 \left(\sqrt{\beta_{k,1}^{-1}} - \sqrt{\beta_{k,0}^{-1}} \right)^2 \left(\alpha_k \beta_{k,j} - \frac{2\beta_{k,j}}{\pi} \frac{\Gamma^2(\alpha_k + \frac{1}{2})}{\Gamma^2(\alpha_k)} \right). \quad (7.13)$$

Proof. of Proposition 7.1 is given in Appendix 7.7.3. $\square$

In virtue of Lindeberg CLT, the statistical distribution of the LR $\Lambda_{\text{dct}}(\tilde{\mathbf{Z}})$ under hypothesis $\mathcal{H}_j$ is derived as

$$\Lambda(\tilde{\mathbf{Z}}) \xrightarrow{d} \mathcal{N}(m_{\text{dct}}^{(j)}, v_{\text{dct}}^{(j)}), \quad (7.14)$$

where the expectation $m_{\text{dct}}^{(j)}$ and variance $v_{\text{dct}}^{(j)}$ are given by

$$m_{\text{dct}}^{(j)} = \sum_{k=1}^K N_b \mathbb{E}_{\mathcal{H}_j}[\Lambda(I_{k,i})] \quad (7.15)$$

$$v_{\text{dct}}^{(j)} = \sum_{k=1}^K N_b \text{Var}_{\mathcal{H}_j}[\Lambda(I_{k,i})]. \quad (7.16)$$

The normalized LR $\Lambda_{\text{dct}}^*(\tilde{\mathbf{Z}})$ is defined as

$$\Lambda_{\text{dct}}^*(\tilde{\mathbf{Z}}) = \frac{\Lambda_{\text{dct}}(\tilde{\mathbf{Z}}) - m_{\text{dct}}^{(0)}}{\sqrt{v_{\text{dct}}^{(0)}}}. \quad (7.17)$$

Finally, the corresponding LRT δ_{dct}^* is rewritten as follows

$$\delta_{\text{dct}}^*(\tilde{\mathbf{Z}}) = \begin{cases} \mathcal{H}_0 & \text{if } \Lambda_{\text{dct}}^*(\tilde{\mathbf{Z}}) < \tau_{\text{dct}}^* \\ \mathcal{H}_1 & \text{if } \Lambda_{\text{dct}}^*(\tilde{\mathbf{Z}}) \geq \tau_{\text{dct}}^* \end{cases} \quad (7.18)$$

where the decision threshold τ_{dct}^* is the solution of the equation $\mathbb{P}_{\mathcal{H}_0}[\Lambda_{\text{dct}}^*(\tilde{\mathbf{Z}}) \geq \tau_{\text{dct}}^*] = \alpha_0$. The decision threshold τ_{dct}^* and the power $\beta(\delta_{\text{dct}}^*)$ are given in following theorem.

Theorem 7.1. *In an ideal context where all the model parameters $(\alpha_k, \tilde{c}_{k,j}, \tilde{d}_{k,j})$ are exactly known, the decision threshold and the power function of the LRT δ_{dct}^* are given by*

$$\tau_{\text{dct}}^* = \Phi^{-1}(1 - \alpha_0) \quad (7.19)$$

$$\beta(\delta_{\text{dct}}^*) = 1 - \Phi \left(\frac{m_{\text{dct}}^{(0)} - m_{\text{dct}}^{(1)} + \tau_{\text{dct}}^* \sqrt{v_{\text{dct}}^{(0)}}}{\sqrt{v_{\text{dct}}^{(1)}}} \right). \quad (7.20)$$

Proof. From (7.14), based on the definitions of decision threshold and power function, the proof follows immediately. $\square$

7.4 Practical Context: GLRT

The scenario studied in the LRT may not be realistic because the parameters $(\alpha_k, \tilde{c}_{k,1}, \tilde{d}_{k,1})$ are unknown in practice. This section designs two GLRTs to deal with unknown parameters. It is proposed to replace unknown parameters by their ML estimates in the LR $\Lambda_{\text{dct}}(I_{k,i})$ (7.11).

7.4.1 GLRT with Unknown Parameters α_k

In this subsection the camera parameters $(\tilde{c}_{k,j}, \tilde{d}_{k,j})$ are assumed to be known. We only deal with unknown nuisance parameters α_k . By replacing α_k by $\hat{\alpha}_k$ given by the ML estimation in Section 4.4.2, the GLR $\bar{\Lambda}_{\text{dct}}(I_{k,i})$ can be given by

$$\bar{\Lambda}_{\text{dct}}(I_{k,i}) = \frac{\hat{\alpha}_k}{2} \log \frac{\tilde{c}_{k,1}\hat{\alpha}_k + \tilde{d}_{k,1}}{\tilde{c}_{k,0}\hat{\alpha}_k + \tilde{d}_{k,0}} - \sqrt{2}|I_{k,i}| \left(\sqrt{\tilde{c}_{k,1}\hat{\alpha}_k + \tilde{d}_{k,1}} - \sqrt{\tilde{c}_{k,0}\hat{\alpha}_k + \tilde{d}_{k,0}} \right). \quad (7.21)$$

The ML estimate $\hat{\alpha}_k$ is asymptotically consistent [20], i.e. it asymptotically converges in probability to its true value: $\hat{\alpha}_k \xrightarrow{p} \alpha_k$. Therefore, from the Slutsky's theorem, the GLR $\bar{\Lambda}_{\text{dct}}(\tilde{\mathbf{Z}}) = \sum_{k=1}^K \sum_{i=1}^{N_b} \bar{\Lambda}_{\text{dct}}(I_{k,i})$ asymptotically follows the Gaussian distribution under each hypothesis $\mathcal{H}_j$ as

$$\bar{\Lambda}_{\text{dct}}(\tilde{\mathbf{Z}}) \xrightarrow{d} \mathcal{N}(m_{\text{dct}}^{(j)}, v_{\text{dct}}^{(j)}), \quad (7.22)$$

where the expectation $m_{\text{dct}}^{(j)}$ and variance $v_{\text{dct}}^{(j)}$ are given in (7.15) and (7.16), respectively. Finally, the GLRT δ_{dct}^* based on the normalized GLR $\bar{\Lambda}_{\text{dct}}^*(\tilde{\mathbf{Z}}) = \frac{\bar{\Lambda}_{\text{dct}}(\tilde{\mathbf{Z}}) - \hat{m}_{\text{dct}}^{(0)}}{\sqrt{\hat{v}_{\text{dct}}^{(0)}}}$ is given by

$$\delta_{\text{dct}}^*(\tilde{\mathbf{Z}}) = \begin{cases} \mathcal{H}_0 & \text{if } \bar{\Lambda}_{\text{dct}}^*(\tilde{\mathbf{Z}}) < \bar{\tau}_{\text{dct}}^* \\ \mathcal{H}_1 & \text{if } \bar{\Lambda}_{\text{dct}}^*(\tilde{\mathbf{Z}}) \geq \bar{\tau}_{\text{dct}}^* \end{cases} \quad (7.23)$$

where the decision threshold $\bar{\tau}_{\text{dct}}^*$ is the solution of the equation $\mathbb{P}_{\mathcal{H}_0}[\bar{\Lambda}_{\text{dct}}^*(\tilde{\mathbf{Z}}) \geq \bar{\tau}_{\text{dct}}^*] = \alpha_0$, and $\hat{m}_{\text{dct}}^{(0)}$ and $\hat{v}_{\text{dct}}^{(0)}$ are estimates of $m_{\text{dct}}^{(0)}$ and $v_{\text{dct}}^{(0)}$ by replacing α_k by $\hat{\alpha}_k$

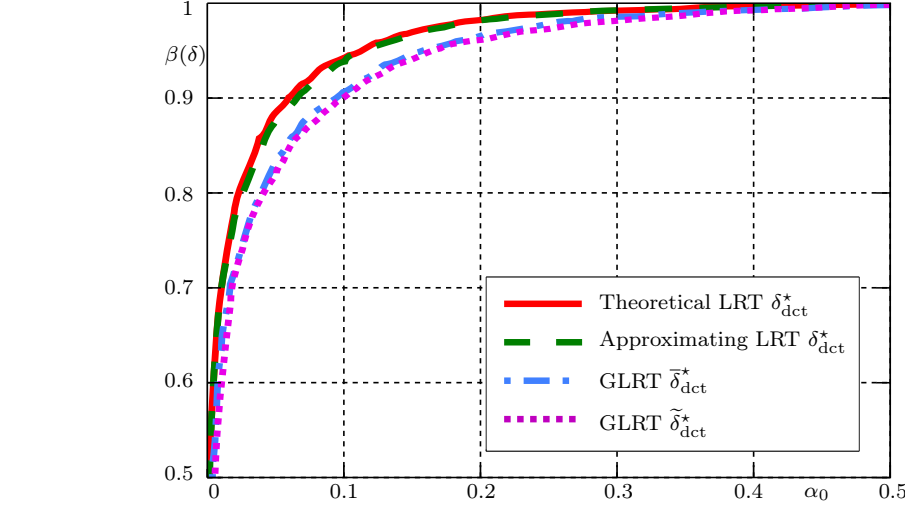


Figure 7.4: Detection performance of proposed tests on simulated vectors with 1024 coefficients.

in (7.15) and (7.16), respectively. From the Slutsky's theorem, the decision threshold $\bar{\tau}_{\text{dct}}^*$ and the power $\beta(\bar{\delta}_{\text{dct}}^*)$ of the GLRT $\bar{\delta}_{\text{dct}}^*$ can be accordingly defined as in the Theorem 7.1.

7.4.2 GLRT with Unknown Parameters $(\alpha_k, \tilde{c}_{k,1}, \tilde{d}_{k,1})$

Before designing the GLRT, the LS estimation of camera parameters $(\tilde{c}_{k,1}, \tilde{d}_{k,1})$ is performed on the inspected image $\tilde{\mathbf{Z}}$; see Section 7.2.2. The LS estimates $(\hat{\tilde{c}}_{k,1}, \hat{\tilde{d}}_{k,1})$ are asymptotically equivalent to ML estimates in large samples [161]. Moreover, they are unbiased and follow the asymptotic bivariate Gaussian distribution

$$\begin{pmatrix} \hat{\tilde{c}}_{k,1} \\ \hat{\tilde{d}}_{k,1} \end{pmatrix} \sim \mathcal{N} \left(\begin{pmatrix} \tilde{c}_{k,1} \\ \tilde{d}_{k,1} \end{pmatrix}, \begin{pmatrix} \sigma_{\tilde{c}_{k,1}}^2 & \sigma_{\tilde{c}_{k,1}\tilde{d}_{k,1}} \\ \sigma_{\tilde{c}_{k,1}\tilde{d}_{k,1}} & \sigma_{\tilde{d}_{k,1}}^2 \end{pmatrix} \right), \quad (7.24)$$

where $\sigma_{\tilde{c}_{k,1}}^2$, $\sigma_{\tilde{d}_{k,1}}^2$, $\sigma_{\tilde{c}_{k,1}\tilde{d}_{k,1}}$ denote the variance of $\hat{\tilde{c}}_{k,1}$, variance of $\hat{\tilde{d}}_{k,1}$, and covariance between $\hat{\tilde{c}}_{k,1}$ and $\hat{\tilde{d}}_{k,1}$, respectively. This covariance matrix is empirically calculated as discussed in Section 6.5.1. The parameters $(\tilde{c}_{k,1}, \tilde{d}_{k,1})$ would characterize an unknown camera model. It is required to take into account the variability of LS estimates $(\hat{\tilde{c}}_{k,1}, \hat{\tilde{d}}_{k,1})$ in the establishment of the statistical performance of the GLRT.

By replacing unknown parameters $(\alpha_k, \tilde{c}_{k,1}, \tilde{d}_{k,1})$ by $(\hat{\alpha}_k, \hat{\tilde{c}}_{k,1}, \hat{\tilde{d}}_{k,1})$ in the LR $\Lambda(I_{k,i})$ (7.11), the GLR $\tilde{\Lambda}_{\text{dct}}(I_{k,i})$ is given as

$$\tilde{\Lambda}_{\text{dct}}(I_{k,i}) = \frac{\hat{\alpha}_k}{2} \log \frac{\hat{\tilde{c}}_{k,1}\hat{\alpha}_k + \hat{\tilde{d}}_{k,1}}{\tilde{c}_{k,0}\hat{\alpha}_k + \tilde{d}_{k,0}} - \sqrt{2}|I_{k,i}| \left(\sqrt{\hat{\tilde{c}}_{k,1}\hat{\alpha}_k + \hat{\tilde{d}}_{k,1}} - \sqrt{\tilde{c}_{k,0}\hat{\alpha}_k + \tilde{d}_{k,0}} \right). \quad (7.25)$$

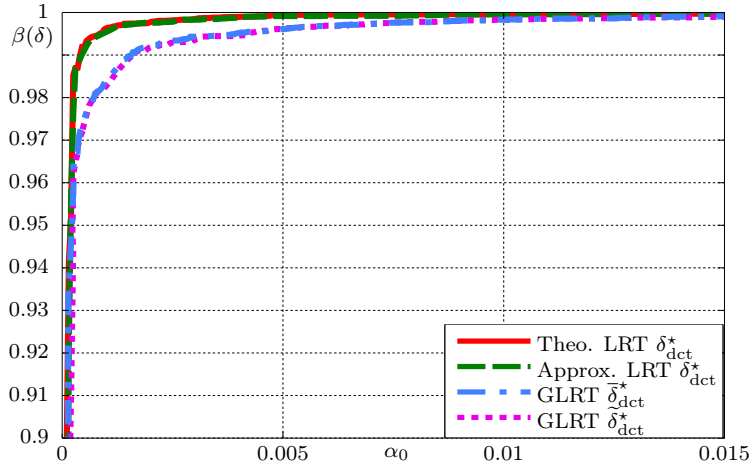


Figure 7.5: Detection performance of proposed tests on simulated vectors with 4096 coefficients.

Proposition 7.2. *Under hypothesis $\mathcal{H}_j$, the first two moments of the GLR $\tilde{\Lambda}_{\text{dct}}(I_{k,i})$ can be approximately defined by*

$$\mathbb{E}_{\mathcal{H}_j} [\tilde{\Lambda}_{\text{dct}}(I_{k,i})] = \mathbb{E}_{\mathcal{H}_j} [\Lambda_{\text{dct}}(I_{k,i})] \quad (7.26)$$

$$\begin{aligned} \text{Var}_{\mathcal{H}_j} [\tilde{\Lambda}_{\text{dct}}(I_{k,i})] &= \text{Var}_{\mathcal{H}_j} [\Lambda_{\text{dct}}(I_{k,i})] \\ &\quad + \frac{\beta_{k,1}^2 \alpha_k (\alpha_k + 2)}{4} \left(\alpha_k^2 \sigma_{\hat{c}_{k,1}}^2 + \sigma_{\hat{d}_{k,1}}^2 + 2\alpha_k \sigma_{\hat{c}_{k,1}} \hat{d}_{k,1} \right). \end{aligned} \quad (7.27)$$

Proof. of Proposition 7.2 is given in Appendix 7.7.4. $\square$

It can be noted that the second term in (7.27) aims to take into account the variability of LS estimates $(\hat{c}_{k,1}, \hat{d}_{k,1})$. In virtue of Lindeberg CLT, the GLR $\tilde{\Lambda}_{\text{dct}}(\tilde{\mathbf{Z}}) = \sum_{k=1}^K \sum_{i=1}^{N_b} \tilde{\Lambda}_{\text{dct}}(I_{k,i})$ follows the Gaussian distribution under each hypothesis $\mathcal{H}_j$

$$\tilde{\Lambda}_{\text{dct}}(\tilde{\mathbf{Z}}) \xrightarrow{d} \mathcal{N}(m_{\text{dct}}^{(j)}, v_{\text{dct}}^{(j)}), \quad (7.28)$$

where the expectation $m_{\text{dct}}^{(j)}$ is given in (7.15) and the variance $v_{\text{dct}}^{(j)}$ is defined as

$$v_{\text{dct}}^{(j)} = \sum_{k=1}^K N_b \text{Var}_{\mathcal{H}_j} [\tilde{\Lambda}_{\text{dct}}(I_{k,i})]. \quad (7.29)$$

Finally, the GLRT $\tilde{\delta}_{\text{dct}}^*$ based on the normalized GLR $\tilde{\Lambda}_{\text{dct}}^*(\tilde{\mathbf{Z}}) = \frac{\tilde{\Lambda}_{\text{dct}}(\tilde{\mathbf{Z}}) - \hat{m}_{\text{dct}}^{(0)}}{\sqrt{\hat{v}_{\text{dct}}^{(0)}}}$ is written as

$$\tilde{\delta}_{\text{dct}}^*(\tilde{\mathbf{Z}}) = \begin{cases} \mathcal{H}_0 & \text{if } \tilde{\Lambda}_{\text{dct}}^*(\tilde{\mathbf{Z}}) < \tilde{\tau}_{\text{dct}}^* \\ \mathcal{H}_1 & \text{if } \tilde{\Lambda}_{\text{dct}}^*(\tilde{\mathbf{Z}}) \geq \tilde{\tau}_{\text{dct}}^* \end{cases} \quad (7.30)$$

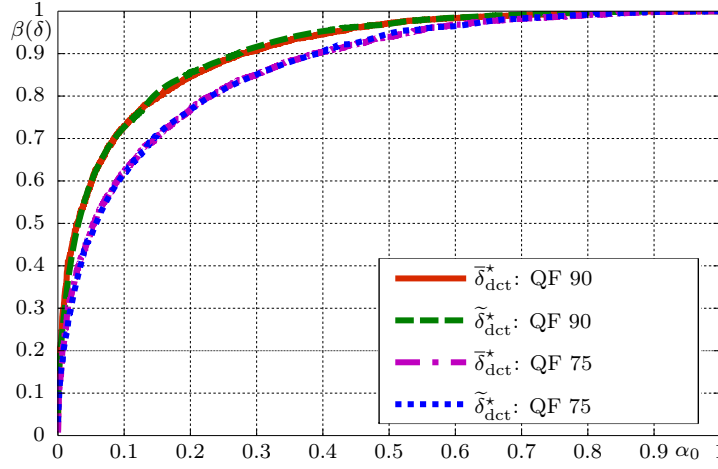


Figure 7.6: Detection performance of proposed GLRTs for 1024 coefficients at frequency (8, 8) extracted randomly from simulated images with different quality factors.

where the decision threshold $\tilde{\tau}_{\text{dct}}^*$ is the solution of the equation $\mathbb{P}_{\mathcal{H}_0}[\tilde{\Lambda}_{\text{dct}}^*(\tilde{\mathbf{Z}}) \geq \tilde{\tau}_{\text{dct}}^*] = \alpha_0$, and $\hat{m}_{\text{dct}}^{(0)}$ and $\hat{v}_{\text{dct}}^{(0)}$ are estimates of $m_{\text{dct}}^{(0)}$ and $v_{\text{dct}}^{(0)}$ by replacing $(\alpha_k, \tilde{c}_{k,1}, \tilde{d}_{k,1})$ by $(\hat{\alpha}_k, \hat{c}_{k,1}, \hat{d}_{k,1})$ in (7.15) and (7.29), respectively. From the Slutsky's theorem, the decision threshold and the power of the GLRT $\tilde{\delta}_{\text{dct}}^*$ are given in the following theorem.

Theorem 7.2. *When the image $\tilde{\mathbf{Z}}$ is tested against the known camera model $\mathcal{S}_0$ characterized by the parameters $(\tilde{c}_{k,0}, \tilde{d}_{k,0})$, the decision threshold and the power of the GLRT $\tilde{\delta}_{\text{dct}}^*$ are given by*

$$\tilde{\tau}_{\text{dct}}^* = \Phi^{-1}(1 - \alpha_0) \quad (7.31)$$

$$\beta(\tilde{\delta}_{\text{dct}}^*) = 1 - \Phi\left(\frac{m_{\text{dct}}^{(0)} - m_{\text{dct}}^{(1)} + \tilde{\tau}_{\text{dct}}^* \sqrt{\hat{v}_{\text{dct}}^{(0)}}}{\sqrt{\hat{v}_{\text{dct}}^{(1)}}}\right). \quad (7.32)$$

Proof. From (7.28), based on the definitions of decision threshold and power function, the proof follows immediately. $\square$

7.5 Numerical Experiments

In this chapter, to suppress image content, the wavelet-based denoising filter proposed in [29, 165] is employed because of its relative accuracy and computational efficiency. Besides, the selection of homogeneous blocks requires an appropriate threshold ϑ . This threshold should be fixed independently of image content. The threshold ϑ is set at $\vartheta = 0.5$.

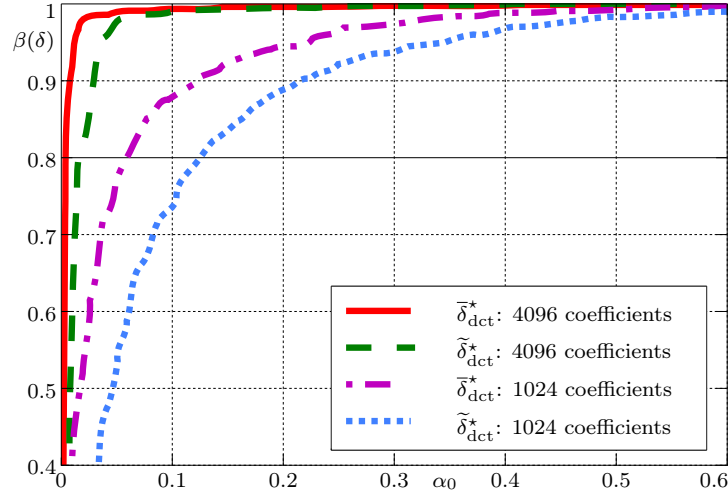


Figure 7.7: Detection performance of proposed tests for different number of coefficients at frequency (8,8) of natural JPEG images taken by *Canon Ixus 70* and *Nikon D200* camera models.

7.5.1 Detection Performance on Simulated Database

The detection performance of proposed tests is first theoretically studied on simulated database. The camera models $\mathcal{S}_0$ and $\mathcal{S}_1$ are characterized by the parameters $(c_0, d_0) = (11.8, -3.5)$ and $(c_1, d_1) = (13.5, -4.5)$, respectively. These parameters correspond to frequency (8,8) of JPEG images taken by *Canon Ixus 70* and *Nikon D200* camera models in the Dresden image database [168], respectively (see Figure 7.2). They are used to generate randomly 5000 vectors of 1024 and 4096 coefficients under $\mathcal{H}_0$ and $\mathcal{H}_1$. Because the LR $\Lambda_{\text{dct}}(I_{k,i})$ is simplified to facilitate the study, it is desirable to compare the detection performance of the LRT based on the approximating LR with the one based on the exact LR. The expectation and variance of the exact LR are calculated numerically. Moreover, it is necessary to compare the detection performance of the proposed GLRTs with the LRT since the GLRTs utilize ML estimates of unknown parameters, which may cause a loss of power. Figure 7.4 and Figure 7.5 show the detection performance of all proposed tests for 1024 and 4096 coefficients, respectively. For clarity, only regions of interest are illustrated in the figures. It is worth noting that the loss of power between the theoretical LRT and approximating LRT is negligible. Besides, a small loss of power is revealed between the GLRTs and LRT due to the estimation of unknown parameters. Nevertheless this loss of power decreases when the number of coefficients increases. It can be also noted that the loss of power between two GLRTs $\bar{\delta}_{\text{dct}}^*$ and $\tilde{\delta}_{\text{dct}}^*$ is negligible, i.e. the variability of estimates $(\hat{c}_{k,1}, \hat{d}_{k,1})$ are well taken into account in the GLRT $\tilde{\delta}_{\text{dct}}^*$. The power function of all proposed tests is perfect (e.g. $\beta(\delta) = 1$) from 2^{14} coefficients for any false alarm rate α_0 .

Moreover, it is desirable to study the detection performance of the proposed tests on simulated images that follow the image processing pipeline as described in

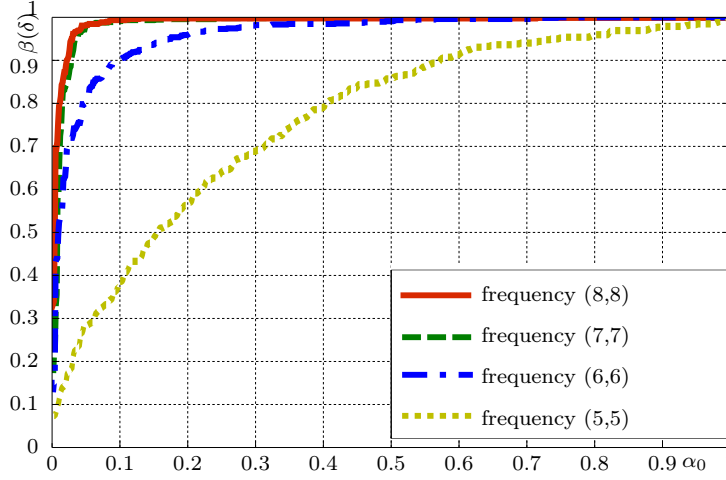


Figure 7.8: Detection performance of the GLRT $\tilde{\delta}_{\text{dct}}^*$ for 4096 coefficients at different frequencies of natural JPEG images taken by *Canon Ixus 70* and *Nikon D200* camera models.

Chapter 4. To this end, suppose the camera models $\mathcal{S}_0$ and $\mathcal{S}_1$ are characterized by the parameters $(\tilde{a}_0, \tilde{b}_0, \gamma_0) = (0.1, 2, 2.2)$ and $(\tilde{a}_1, \tilde{b}_1, \gamma_1) = (0.2, 2, 2.2)$. These parameters are used together with the reference image *lena* to generate randomly 5000 images under $\mathcal{H}_0$ and $\mathcal{H}_1$. The simulated images are then compressed with quality factor of 90 and 75. The detection performance of the proposed GLRTs for 1024 coefficients at frequency (8,8) extracted randomly from those simulated images is shown in Figure 7.6. As expected, a small loss of power is revealed with the decline of quality factor.

7.5.2 Detection Performance on Two *Canon Ixus 70* and *Nikon D200* Camera Models

In practice, the reference camera parameters $(\tilde{c}_{k,0}, \tilde{d}_{k,0})$ are given by estimating the parameters $(\tilde{c}_k, \tilde{d}_k)$ on each image of 50 images taken by the camera model $\mathcal{S}_0$ and averaging 50 previous estimates.

To highlight the relevance of the proposed GLRTs, two *Canon Ixus 70* and *Nikon D200* camera models of the Dresden image database [168] are chosen to conduct experiments. The *Canon Ixus 70* and *Nikon D200* cameras are respectively set at $\mathcal{H}_0$ and $\mathcal{H}_1$. The Figure 7.7 shows the detection performance of the GLRTs $\bar{\delta}_{\text{dct}}^*$ and $\tilde{\delta}_{\text{dct}}^*$ for 1024 and 4096 coefficients extracted randomly at frequency (8,8) of natural JPEG images taken by *Canon Ixus 70* and *Nikon D200* camera models. We can note a similar behavior to the detection performance on simulated database. Besides, there is a small loss of power between the two GLRTs because different estimates $(\hat{c}_{k,1}, \hat{d}_{k,1})$ used in the design of the GLRT $\tilde{\delta}_{\text{dct}}^*$ are still influenced by image content. Nevertheless, this loss of power also decreases when the number of coefficients increases. Besides, Figure 7.8 illustrates detection performance of the GLRT $\tilde{\delta}_{\text{dct}}^*$ for 4096 coefficients randomly extracted at different frequencies.

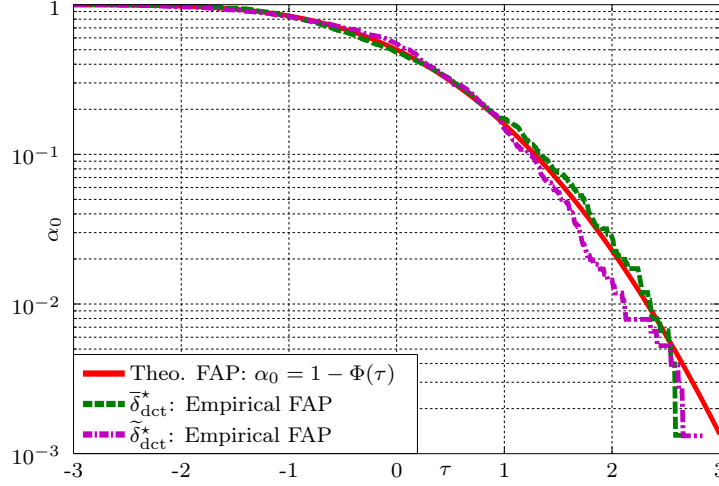


Figure 7.9: Comparison between the theoretical false alarm probability (FAP) and the empirical FAP, plotted as a function of decision threshold τ , of the proposed tests at the frequency (8,8) of natural images.

Camera Model		No. devices	Resolution	No. images
Canon Ixus 70	Cn2	3	3072×2304	500
Fujifilm FinePix J50	F	3	3264×2448	500
Kodak M1063	K	5	3664×2748	550
Nikon Coolpix S710	N1	5	4352×3264	550
Nikon D200	N2	2	3872×2592	500
Nikon D70	N3	2	3008×2000	300
Pentax Optio A40	Pe	4	4000×3000	600
Praktica DCZ 5.9	Pr	5	2560×1920	700
Ricoh Capilo GX100	Ri	5	3648×2736	700
Rollei RCP-7325XS	Ro	3	3072×2304	350
Sony DSC-H50	S	2	3456×2592	250
Σ	11	39		5500

Table 7.1: Camera Model Used in Experiments

It can be noted that the detection performance decreases with the reverse zig-zag order. Meanwhile, the Figure 7.9 shows the comparison between the theoretical and empirical false alarm probability, which are plotted as a function of decision threshold τ , of the proposed tests at the frequency (8,8) of natural images. The two proposed GLRTs $\hat{\delta}_{\text{dct}}^*$ and $\tilde{\delta}_{\text{dct}}^*$ show an ability of guaranteeing a prescribed false alarm rate.

7.5.3 Detection Performance on a Large Image Database

Experiments are then conducted on the Dresden database to verify the efficiency of the proposed approach. Technical specifications of the cameras are shown in Table 7.1. Firstly, the GLRT $\tilde{\delta}_{\text{dct}}^*$ is used to verify whether a given image is acquired by the camera model of interest. The decision threshold $\tilde{\tau}_{\text{dct}}^*$ is given by the Theorem

		Tested images											
		Cn2	N1	Pe	F	Ro	K	N3	N2	Pr	Ri	S	Avg.
$\mathcal{H}_0$	Cn2	100	26.54	20.42	*	*	*	*	*	*	*	*	95.31
	N1	*	84.61	10.56	*	*	*	*	*	*	*	*	
	Pe	*	7.79	82.04	*	*	*	*	*	*	*	*	
	F	*	*	*	91.67	6.63	*	*	*	*	*	*	
	Ro	*	*	*	22.92	100	*	*	*	*	*	*	
	K	*	*	*	*	*	99.81	14.15	*	*	*	*	
	N3	*	*	*	*	*	*	97.62	*	*	*	*	
	N2	*	*	*	*	*	*	*	100	*	*	*	
	Pr	*	*	*	*	*	*	*	*	95.78	*	*	
	Ri	*	*	*	*	*	*	*	*	*	100	*	
	S	*	*	*	*	*	*	*	*	*	*	96.81	
													95.31

Table 7.2: Detection performance of proposed detector $\tilde{\delta}_{\text{dct}}^*$ (the symbol * represents values smaller than 2%)

		Tested images											
		Cn2	N1	Pe	F	Ro	K	N3	N2	Pr	Ri	S	Avg.
$\mathcal{H}_0$	Cn2	98.26	*	*	*	*	*	*	2.61	*	*	*	
	N1	*	100	3.97	*	*	*	*	*	*	*	*	
	Pe	*	*	90.86	*	*	*	*	*	*	*	*	
	F	*	*	*	99.94	*	*	*	*	*	*	*	
	Ro	*	*	*	*	99.79	*	*	*	*	*	*	
	K	*	*	*	*	*	99.19	*	*	*	*	*	
	N3	*	*	*	*	*	*	84.57	*	*	*	*	
	N2	*	*	*	*	*	*	*	93.73	*	*	*	
	Pr	*	*	*	*	*	*	*	*	98.42	*	*	
	Ri	*	*	*	*	*	*	*	*	*	100	*	
S	*	*	*	*	*	*	*	*	*	*	98.23		
													96.63

Table 7.3: Detection performance of SVM-based detector

7.2 corresponding to the false alarm rate $\alpha_0 = 10^{-5}$. It is proposed to use the last 21 high frequencies for the test since other frequencies may be still affected by the image content. The detection performance of the test $\tilde{\delta}_{\text{dct}}^*$ is shown in Table 7.2. It can be noted that the incorrect detection in some groups of camera models, such as (Cn2,N1,Pe), (F,Ro), and (K,N3), is important. This may be justified due to a similarity in JPEG compression scheme used in the camera and the very false alarm probability that is set in this experiment. To deal with this scenario, a second testing round involves performing the GLRT $\bar{\delta}_{\text{dct}}^*$ on the camera models of conflict. The detection performance of the test $\bar{\delta}_{\text{dct}}^*$ is shown in Table 7.5. The images are almost correctly classified. It should be noted that the GLRT $\bar{\delta}_{\text{dct}}^*$ aims to give a decision rule between two different camera models, thus the experiment on the diagonal of Table 7.5 is not performed. The SVM-based detector [168] is presented for comparison. Its detection performance is shown in Table 7.3. The proposed detector $\tilde{\delta}_{\text{dct}}^*$ is slightly equivalent to the SVM-based detector in terms of average correct detection performance but the misclassification of the former is more severe.

		Tested images											
		Cn2	N1	Pe	F	Ro	K	N3	N2	Pr	Ri	S	Avg.
$\mathcal{H}_0$	Cn2	100	*	*	*	*	*	*	*	*	*	*	
	N1	*	98.92	*	*	*	*	*	*	*	*	*	
	Pe	*	*	100	*	*	*	*	*	*	*	*	
	F	*	*	*	73.81	*	*	*	*	*	*	*	
	Ro	*	*	*	*	99.49	*	*	*	*	*	*	
	K	*	*	*	*	*	99.19	*	*	*	*	*	
	N3	*	*	*	*	*	*	100	*	*	*	*	
	N2	*	*	*	*	*	*	*	98.92	*	*	*	
	Pr	*	*	*	*	*	*	*	*	96.65	*	*	
	Ri	*	*	*	*	*	*	*	*	*	100	*	
	S	*	*	*	*	*	*	*	*	*	*	100	
													96.99

Table 7.4: Detection performance of PRNU-based detector

		$\mathcal{H}_1$						
		Cn2	N1	Pe	F	Ro	K	N3
$\mathcal{H}_0$	Cn2	-	0	3.52	-	-	-	-
	N1	0	-	3.13	-	-	-	-
	Pe	0	0	-	-	-	-	-
	F	-	-	-	-	0	-	-
	Ro	-	-	-	2.12	-	-	-
	K	-	-	-	-	-	-	0
	N3	-	-	-	-	-	0	-

Table 7.5: Detection performance of proposed detector $\bar{\delta}_{\text{dct}}^*$

		Tested images			
		Canon 400D	Pentax K20D	Canon Rebel XSi	Leica M9
$\mathcal{H}_0$	Canon 400D	98	78	*	*
	Pentax K20D	72	94	*	*
	Canon Rebel XSi	*	*	96	53
	Leica M9	*	*	49	93

Table 7.6: Detection performance of proposed detector $\tilde{\delta}_{\text{dct}}^*$ on 4 camera models of BOSS database

The PRNU-based detector [32] is also performed in this experiment. This PRNU-based detector is only conducted on one device per model. Its detection performance is shown in Table 7.4. The detection performance of the proposed detector $\tilde{\delta}_{\text{dct}}^*$ is slightly worse than other ones.

The proposed approach is based on the difference in statistical properties in the DCT domain that jointly results from different sensor noises and in-camera processing algorithms. Thus it is desirable to observe the detection performance of the proposed tests on the images that go through the same post-acquisition processes. Therefore, it is proposed to perform the same operations on 4000 RAW images from 4 camera models in the BOSS database [171] and compress them at the quality factor

95. Consequently, the difference in statistical properties of DCT coefficients mainly results from different sensor noises. The detection performance of the detector $\tilde{\delta}_{\text{dct}}^*$ on these images is shown in Table 7.6. It can be noted that the probability of miss detection is considerably poor. The detector $\tilde{\delta}_{\text{dct}}^*$ can not provide better results as its detection performance is similar to the random guess. This experiment shows a difficulty in camera model identification as the image processing techniques remain identical or similar between camera models.

7.6 Conclusion

This chapter presents a new approach based on the state-of-the-art model of DCT coefficients to capture their statistical difference, which jointly results from different sensor noises and in-camera processing algorithms. The parameters $(\tilde{c}, \tilde{d})$ characterizing the simplistic linear relation between α and β^{-1} , which are two parameters of the DCT coefficient model, are proposed to be exploited as camera fingerprint for camera model identification. Based on the parametric model of DCT coefficients, this chapter studies the most powerful LRT and proposes two GLRTs that can be straightforwardly applied in practice.

7.7 Appendix

7.7.1 Relation between the Parameters $(\tilde{a}, \tilde{b}, \gamma)$ and $(\alpha_{u,v}, \beta_{u,v})$

For the sake of simplification, it is assumed that the pixels are independent and identically distributed within each 8×8 block. By taking variance on the both sides of the equation (2.11), we obtain

$$\mathbb{V}\text{ar}[I_{u,v}] = \mathbb{V}\text{ar}[\tilde{z}_{m,n}] \quad (7.33)$$

$$\mathbb{V}\text{ar}[I_{u,v}^2] = A_{u,v} \mathbb{V}\text{ar}[\tilde{z}_{m,n}^2] + (1 - A_{u,v}) \mathbb{V}\text{ar}^2[\tilde{z}_{m,n}], \quad (7.34)$$

with

$$A_{u,v} = \frac{1}{4^4} T_u^4 T_v^4 \sum_{m=0}^7 \sum_{n=0}^7 \cos^4\left(\frac{(2m+1)u\pi}{16}\right) \cos^4\left(\frac{(2n+1)v\pi}{16}\right). \quad (7.35)$$

On the one hand, it follows from (4.71) and (4.72) that

$$\mathbb{V}\text{ar}[I_{u,v}] = \alpha_{u,v} \beta_{u,v} \quad (7.36)$$

$$\mathbb{V}\text{ar}[I_{u,v}^2] = \mathbb{E}[I_{u,v}^4] - \mathbb{E}^2[I_{u,v}^2] = 2\alpha_{u,v}^2 \beta_{u,v}^2 + 3\alpha_{u,v} \beta_{u,v}^2. \quad (7.37)$$

On the other hand, from the noise model (4.42), we obtain

$$\mathbb{V}\text{ar}[\tilde{z}_{m,n}] = \frac{1}{\gamma^2} \tilde{\mu}^{2-2\gamma} (\tilde{a} \tilde{\mu}^\gamma + \tilde{b}) \quad (7.38)$$

$$\mathbb{V}\text{ar}[\tilde{z}_{m,n}^2] = \frac{4}{\gamma^2} \tilde{\mu}^{4-2\gamma} (\tilde{a} \tilde{\mu}^\gamma + \tilde{b}). \quad (7.39)$$

Consequently, it follows that

$$\alpha_{u,v}\beta_{u,v} = \frac{1}{\gamma^2}\tilde{\mu}^{2-2\gamma}(\tilde{a}\tilde{\mu}^\gamma + \tilde{b}) \quad (7.40)$$

$$(A_{u,v} + 1)\alpha_{u,v}\beta_{u,v} + 3\beta_{u,v} = 4A_{u,v}\tilde{\mu}^2. \quad (7.41)$$

By resolving this system of equations, the relation between $\alpha_{u,v}$ and $\beta_{u,v}^{-1}$ is given as

$$\beta_{u,v}^{-1} = \frac{(A_{u,v} + 1)\alpha_{u,v} + 3}{4\tilde{b}^{\frac{2}{\gamma}}A_{u,v}} \left(\sqrt{\gamma^2\alpha_{u,v}\tilde{b}\frac{(A_{u,v} + 1)\alpha_{u,v} + 3}{4A_{u,v}} + \frac{\tilde{a}^2}{4} - \frac{\tilde{a}}{2}} \right)^{\frac{2}{\gamma}}. \quad (7.42)$$

This relation is too complicated to exploit. Therefore, it is proposed to employ the polynomial expansion and only keep the first two terms

$$\beta_{u,v}^{-1} = \tilde{c}_{u,v}\alpha_{u,v} + \tilde{d}_{u,v}, \quad (7.43)$$

where the parameters $(\tilde{c}_{u,v}, \tilde{d}_{u,v})$ depend on the parameters $(\tilde{a}, \tilde{b}, \gamma)$. Numerical experiments show that this simplified equation sufficiently characterizes the relation between the parameters $(\tilde{a}, \tilde{b}, \gamma)$ and $(\alpha_{p,q}, \beta_{p,q})$ (see Figure 7.1 and Figure 7.2).

7.7.2 Laplace's Approximation of DCT Coefficient Model

Let us briefly describe the idea behind the Laplace's approximation [180]. The Laplace's method aims to provide an approximation for integrals of the form

$$\mathfrak{J} = \int \exp(-g(t))dt, \quad (7.44)$$

when the function $g(t)$ reaches the global minimum at t^* . By using the Taylor expansion of the function $g(t)$ at t^* , we have

$$g(t) = g(t^*) + \frac{g''(t^*)}{2}(t - t^*)^2 + o((t - t^*)^2), \quad (7.45)$$

where $g''(t)$ denotes the second derivative of the function $g(t)$. Therefore, the integral $\mathfrak{J}$ can be approximated as

$$\mathfrak{J} \approx \exp(-g(t^*)) \int \exp\left[-\frac{g''(t^*)}{2}(t - t^*)^2\right] dt. \quad (7.46)$$

This integral takes the form of Gaussian integral. We derive

$$\mathfrak{J} \approx \sqrt{\frac{2\pi}{|g''(t^*)|}} \exp(-g(t^*)). \quad (7.47)$$

A generalization has been made in [181] with an arbitrary function $h(t)$

$$\mathfrak{J} = \int h(t) \exp(-g(t))dt \approx \sqrt{\frac{2\pi}{|g''(t^*)|}} h(t^*) \exp(-g(t^*)). \quad (7.48)$$

From (4.69), the DCT coefficient model $f_I(x)$ can be rewritten as

$$f_I(x) = \frac{1}{\sqrt{2\pi}\beta^\alpha\Gamma(\alpha)} \int_0^\infty h(t) \exp(-g(t)) dt \quad (7.49)$$

where

$$g(t) = \frac{t}{\beta} + \frac{x^2}{2t} \quad \text{and} \quad h(t) = t^{\alpha-\frac{3}{2}}. \quad (7.50)$$

The function $g(t)$ reaches the minimum at $t^* = |x|\sqrt{\frac{\beta}{2}}$ and its second derivative is defined by $g''(t) = \frac{x^2}{t^3}$. Consequently, the function $f_I(x)$ can be approximated as

$$f_I(x) \approx \frac{|x|^{\alpha-1}}{(2\beta)^{\frac{\alpha}{2}}\Gamma(\alpha)} \exp\left(-|x|\sqrt{\frac{2}{\beta}}\right). \quad (7.51)$$

It can be noted that this approximating model is a special case of the GF model [146] when $\gamma = 1$ (the variable γ is given in [146, Eq. (6)]).

7.7.3 Expectation and Variance of the LR $\Lambda_{\text{dct}}(I_{k,i})$ under Hypothesis $\mathcal{H}_j$

It can be noted from (7.11) that it is necessary to calculate expectation and variance of the random variable $|I|$. Given a known variance σ_{blk}^2 , the random variable I is normally distributed with zero-mean and variance σ_{blk}^2 . Thus, the random variable $|I|$ follows the half-Normal distribution [160]. Therefore, we obtain

$$\mathbb{E}_{I|\sigma_{\text{blk}}^2} [|I| | \sigma_{\text{blk}}^2] = \sqrt{\frac{2}{\pi}} \sigma_{\text{blk}}. \quad (7.52)$$

Based on the law of total expectation, the mathematical expectation of $|I|$ is given by

$$\mathbb{E}_I [|I|] = \mathbb{E}_{\sigma_{\text{blk}}^2} [\mathbb{E}_{I|\sigma_{\text{blk}}^2} [|I| | \sigma_{\text{blk}}^2]] = \sqrt{\frac{2}{\pi}} \mathbb{E}_{\sigma_{\text{blk}}^2} [\sigma_{\text{blk}}] = \sqrt{\frac{2}{\pi}} \beta^{\frac{1}{2}} \frac{\Gamma(\alpha + \frac{1}{2})}{\Gamma(\alpha)}. \quad (7.53)$$

Besides, the variance of $|I|$ is given by

$$\text{Var}_I [|I|] = \mathbb{E}_I [|I|^2] - \mathbb{E}_I^2 [|I|] = \alpha\beta - \frac{2\beta}{\pi} \frac{\Gamma^2(\alpha + \frac{1}{2})}{\Gamma^2(\alpha)}. \quad (7.54)$$

Consequently, the expectation and variance of the LR $\Lambda_{\text{dct}}(I_{k,i})$ under hypothesis $\mathcal{H}_j$ can be defined by

$$\mathbb{E}_{\mathcal{H}_j} [\Lambda_{\text{dct}}(I_{k,i})] = \frac{\alpha_k}{2} \log \frac{\beta_{k,1}^{-1}}{\beta_{k,0}^{-1}} - \frac{2}{\sqrt{\pi}} \beta_{k,j}^{\frac{1}{2}} \frac{\Gamma(\alpha_k + \frac{1}{2})}{\Gamma(\alpha_k)} (\sqrt{\beta_{k,1}^{-1}} - \sqrt{\beta_{k,0}^{-1}}) \quad (7.55)$$

$$\text{Var}_{\mathcal{H}_j} [\Lambda_{\text{dct}}(I_{k,i})] = 2 \left(\sqrt{\beta_{k,1}^{-1}} - \sqrt{\beta_{k,0}^{-1}} \right)^2 \left(\alpha_k \beta_{k,j} - \frac{2\beta_{k,j}}{\pi} \frac{\Gamma^2(\alpha_k + \frac{1}{2})}{\Gamma^2(\alpha_k)} \right). \quad (7.56)$$

7.7.4 Asymptotic Expectation and Variance of the GLR $\tilde{\Lambda}_{\text{dct}}(I_{k,i})$ under Hypothesis $\mathcal{H}_j$

The variance of the estimate $\hat{\alpha}_k$ is assumed to be negligible when the number of coefficients N_b is very large. Thus the estimate $\hat{\alpha}_k$ can be treated as a constant α_k . Besides, since the estimates $(\hat{c}_{k,1}, \hat{d}_{k,1})$ are consistent, the asymptotic mathematical expectation of the GLR $\tilde{\Lambda}_{\text{dct}}(I_{k,i})$ under hypothesis $\mathcal{H}_j$ does not change, i.e. $\mathbb{E}_{\mathcal{H}_j}[\tilde{\Lambda}_{\text{dct}}(I_{k,i})] = \mathbb{E}_{\mathcal{H}_j}[\Lambda_{\text{dct}}(I_{k,i})]$.

Meanwhile, the variance of the GLR $\tilde{\Lambda}_{\text{dct}}(I_{k,i})$ needs to take into account the variability of the estimates $(\hat{c}_{k,1}, \hat{d}_{k,1})$. Based on the definitions of mathematical expectation and variance, we have

$$\mathbb{E}_{\mathcal{H}_j}[\hat{c}_{k,1}\hat{\alpha}_k + \hat{d}_{k,1}] = \beta_{k,1}^{-1} \quad (7.57)$$

$$\mathbb{V}\text{ar}_{\mathcal{H}_j}[\hat{c}_{k,1}\hat{\alpha}_k + \hat{d}_{k,1}] = \alpha_k^2 \sigma_{\hat{c}_{k,1}}^2 + \sigma_{\hat{d}_{k,1}}^2 + 2\alpha_k \sigma_{\hat{c}_{k,1}\hat{d}_{k,1}}. \quad (7.58)$$

Subsequently, from the Delta method in Lemma A.2, we derive that

$$\begin{aligned} \mathbb{V}\text{ar}_{\mathcal{H}_j} \left[\log \frac{\hat{c}_{k,1}\hat{\alpha}_k + \hat{d}_{k,1}}{\tilde{c}_{k,0}\hat{\alpha}_k + \tilde{d}_{k,0}} \right] &= \mathbb{V}\text{ar}_{\mathcal{H}_j} \left[\log (\hat{c}_{k,1}\hat{\alpha}_k + \hat{d}_{k,1}) \right] = \frac{\mathbb{V}\text{ar}_{\mathcal{H}_j}[\hat{c}_{k,1}\hat{\alpha}_k + \hat{d}_{k,1}]}{\mathbb{E}_{\mathcal{H}_j}^2[\hat{c}_{k,1}\hat{\alpha}_k + \hat{d}_{k,1}]} \\ &= \beta_{k,1}^2 \left(\alpha_k^2 \sigma_{\hat{c}_{k,1}}^2 + \sigma_{\hat{d}_{k,1}}^2 + 2\alpha_k \sigma_{\hat{c}_{k,1}\hat{d}_{k,1}} \right), \end{aligned} \quad (7.59)$$

and

$$\begin{aligned} \mathbb{V}\text{ar}_{\mathcal{H}_j} \left[\sqrt{\hat{c}_{k,1}\hat{\alpha}_k + \hat{d}_{k,1}} - \sqrt{\tilde{c}_{k,0}\hat{\alpha}_k + \tilde{d}_{k,0}} \right] &= \mathbb{V}\text{ar}_{\mathcal{H}_j} \left[\sqrt{\hat{c}_{k,1}\hat{\alpha}_k + \hat{d}_{k,1}} \right] \\ &= \frac{\mathbb{V}\text{ar}_{\mathcal{H}_j}[\hat{c}_{k,1}\hat{\alpha}_k + \hat{d}_{k,1}]}{4\mathbb{E}_{\mathcal{H}_j}[\hat{c}_{k,1}\hat{\alpha}_k + \hat{d}_{k,1}]} \\ &= \frac{\beta_{k,1}}{4} \left(\alpha_k^2 \sigma_{\hat{c}_{k,1}}^2 + \sigma_{\hat{d}_{k,1}}^2 + 2\alpha_k \sigma_{\hat{c}_{k,1}\hat{d}_{k,1}} \right). \end{aligned} \quad (7.60)$$

Finally, the asymptotic variance of the GLR $\tilde{\Lambda}_{\text{dct}}(I_{k,i})$ can be given as

$$\begin{aligned} \mathbb{V}\text{ar}_{\mathcal{H}_j}[\tilde{\Lambda}_{\text{dct}}(I_{k,i})] &= \frac{\alpha_k^2}{4} \mathbb{V}\text{ar}_{\mathcal{H}_j} \left[\log \frac{\hat{c}_{k,1}\hat{\alpha}_k + \hat{d}_{k,1}}{\tilde{c}_{k,0}\hat{\alpha}_k + \tilde{d}_{k,0}} \right] \\ &\quad + 2\mathbb{V}\text{ar}_{\mathcal{H}_j} \left[\sqrt{\hat{c}_{k,1}\hat{\alpha}_k + \hat{d}_{k,1}} - \sqrt{\tilde{c}_{k,0}\hat{\alpha}_k + \tilde{d}_{k,0}} \right] \mathbb{E}_{\mathcal{H}_j}[|I|^2] \\ &\quad + 2\mathbb{E}_{\mathcal{H}_j}^2 \left[\sqrt{\hat{c}_{k,1}\hat{\alpha}_k + \hat{d}_{k,1}} - \sqrt{\tilde{c}_{k,0}\hat{\alpha}_k + \tilde{d}_{k,0}} \right] \mathbb{V}\text{ar}_{\mathcal{H}_j}[|I|^2] \\ &= \frac{\beta_{k,1}^2 \alpha_k (\alpha_k + 2)}{4} \left(\alpha_k^2 \sigma_{\hat{c}_{k,1}}^2 + \sigma_{\hat{d}_{k,1}}^2 + 2\alpha_k \sigma_{\hat{c}_{k,1}\hat{d}_{k,1}} \right) \\ &\quad + 2 \left(\sqrt{\beta_{k,1}^{-1}} - \sqrt{\beta_{k,0}^{-1}} \right)^2 \left(\alpha_k \beta_{k,j} - \frac{2\beta_{k,j}}{\pi} \frac{\Gamma^2(\alpha_k + \frac{1}{2})}{\Gamma^2(\alpha_k)} \right). \end{aligned} \quad (7.61)$$

Part IV

Statistical Detection of Hidden Data in Natural Images

Statistical Detection of Data Embedded in Least Significant Bits of Clipped Images

Contents

8.1	Introduction	140
8.2	Cover-Image Model	141
8.2.1	Non-Clipped Image Model	141
8.2.2	Clipped Image Model	142
8.3	GLRT for Non-Clipped Images	142
8.3.1	Impact of LSB Replacement: Stego-Image Model	142
8.3.2	Hypothesis Testing Formulation	143
8.3.3	ML Estimation of Image Parameters	144
8.3.4	Design of GLRT	145
8.4	GLRT for Clipped Images	148
8.4.1	ML Estimation of Image Parameters	148
8.4.2	Design of GLRT	148
8.5	Numerical Experiments	151
8.5.1	Detection Performance on Simulated Database	151
8.5.2	Detection Performance on Real Image Database	153
8.6	Conclusion	155
8.7	Appendix	155
8.7.1	Denoising Filter for Non-Clipped RAW Images Corrupted by Signal-Dependent Noise	155
8.7.2	Statistical distribution of the GLR $\hat{\Lambda}_{\text{ncl}}(\mathbf{Z})$ under hypothesis $\mathcal{H}_0$	157
8.7.3	Statistical distribution of the GLR $\hat{\Lambda}_{\text{ncl}}(\mathbf{Z})$ under hypothesis $\mathcal{H}_1$	158
8.7.4	ML Estimation of Parameters in Truncated Gaussian Data	159
8.7.5	Statistical distribution of the GLR $\hat{\Lambda}_{\text{cl}}(\mathbf{Z})$	160

8.1 Introduction

The goal of this chapter is to design a statistical test for the steganalysis of LSB replacement in a natural RAW image. In spite of lots of existing methods proposed for the steganalysis of LSB replacement in the literature (see Section 2.5.2), an improvement of existing steganalysis methods is still desirable. Like Chapter 5, 6, 7, hypothesis testing theory is the main framework of our researches for the hidden data detection problem. This framework has been already applied in the steganalysis of LSB replacement, see [107, 109–114]. However, those detectors are only based on a simplistic image model that can not totally characterize a natural image, resulting in inaccurate estimation of pixels' expectation and variance. This leads to a loss of power even on a simulated database.

Another problem that has not been studied yet in the literature is the phenomenon of clipping. This phenomenon is due to limited dynamic range of the imaging system. None of existing detectors have considered the clipping phenomenon, thus its impact on their detection performance is unknown.

The main contributions are the following:

- The proposed approach is based on the heteroscedastic noise model that is more relevant to characterize a natural RAW image than the conventional AWGN.
- In order to estimate expectation of the pixels, a state-of-the-art denoising method [155] is exploited. Using together with the heteroscedastic noise model, this method can significantly improve the estimation of pixels' expectation and variance, resulting in a higher detection performance.
- The clipping phenomenon, which makes the so-called *clipped* pixels being overexposed or underexposed, is taken into account in a rigorous statistical model. Based on this non-linear model of imaging device response, a generalized accurate detector is provided for any natural RAW image (non-clipped or clipped). It is also shown that this phenomenon dramatically reduces the performance of prior-art detectors.
- Using the accurate statistical model of natural images, the GLRTs are designed such that they allow the warranting of a prescribed false alarm probability and the setting of decision threshold independently of the image content while other statistical detectors [109] fail in practice.

The work in this chapter has been published in [182]. The chapter is organized as follows. Section 8.2 presents a cover model for a non-clipped and clipped RAW image. Section 8.3 casts the hidden data detection problem into the framework of hypothesis testing theory and proposes a GLRT to deal with the difficulty of unknown image parameters in a non-clipped image. Section 8.4 addresses a GLRT taking into account the impact of clipping phenomenon. This detector serves as an upper bound of the previously proposed detector or any practical detector which does not

regard this kind of degradation. Section 8.5 presents numerical performances of the proposed tests on synthetic and real natural images and some comparisons with other state-of-the-art detectors. Finally, Section 8.6 concludes the chapter.

8.2 Cover-Image Model

As discussed in Section 3.2.1, a RAW image can be accurately characterized by the heteroscedastic noise model. Besides, this model has been successfully applied to the problem of camera model identification, see Chapter 5. Therefore, this chapter proposes to start from the heteroscedastic noise model in order to design a statistical test for hidden data detection.

8.2.1 Non-Clipped Image Model

It can be noted that the statistical analysis of heteroscedastic noise model in Section 3.2.1 has not taken into account yet the impact of quantization. For the sake of clarity, let $\mathbf{X} = (x_1, \dots, x_N)^T$ and $\mathbf{Z} = (z_1, \dots, z_N)^T$ denote the non-clipped RAW image before quantization and after quantization, respectively. According to the heteroscedastic noise model (4.1), the pixel x_i follows the Gaussian distribution with expectation μ_i and variance $\sigma_i^2 = a\mu_i + b$,

$$x_i \sim \mathcal{N}(\mu_i, \sigma_i^2). \quad (8.1)$$

For the sake of simplification, suppose the quantization is performed with unitary step. Hence the quantized non-clipped RAW pixel z_i can be defined as $z_i = \lfloor x_i \rfloor$. Let P_{θ_i} , given by

$$P_{\theta_i} = \{p_{\theta_i}[0], p_{\theta_i}[1], \dots, p_{\theta_i}[B]\}, \quad (8.2)$$

be the pmf of the non-clipped RAW pixel z_i where $\theta_i = (\mu_i, a, b)$, $B = 2^\nu - 1$ and $p_{\theta_i}[l]$, $\forall l \in \{0, \dots, B\}$, is given by

$$\begin{aligned} p_{\theta_i}[l] &= \mathbb{P}[z_i = l] = \mathbb{P}\left[l - \frac{1}{2} \leq x_i \leq l + \frac{1}{2}\right] \\ &= \frac{1}{\sigma_i} \int_{l-\frac{1}{2}}^{l+\frac{1}{2}} \phi\left(\frac{t - \mu_i}{\sigma_i}\right) dt \\ &= \Phi\left(\frac{l + \frac{1}{2} - \mu_i}{\sigma_i}\right) - \Phi\left(\frac{l - \frac{1}{2} - \mu_i}{\sigma_i}\right). \end{aligned} \quad (8.3)$$

From the mean value theorem, the probability $p_{\theta_i}[l]$ can be simplified as

$$p_{\theta_i}[k] = \frac{1}{\sigma_i} \phi\left(\frac{l - \mu_i}{\sigma_i}\right) + o(\sigma_i^{-2}), \quad (8.4)$$

where the term $o(\sigma_i^{-2})$ accounts for the error of approximation. It is assumed that the error $o(\sigma_i^{-2})$ is negligible because of the fact that the quantization step is very small compared with noise in a natural RAW image [110]. This assumption is quite

realist because a RAW image is often coded with $\nu \in \{12, 14, 16\}$ bits. At this stage it is assumed that the clipping phenomenon is absent, i.e. the probability that one observation x_i exceeds over the boundary 0 or B is negligible.

8.2.2 Clipped Image Model

In practice, the dynamic range of acquisition system is always limited. Indeed, recorded pixels are bounded by the interval $[0, B]$. The clipping phenomenon may happen when noise is very important, which may cause the pixel x to exceed these bounds. Consequently, the recorded pixels exceeding these bounds are replaced by the bounds themselves. Let $\underline{\mathbf{X}} = (x_1, \dots, x_N)^T$ and $\underline{\mathbf{Z}} = (z_1, \dots, z_N)^T$ denote the clipped RAW image before quantization and after quantization, respectively. Therefore, the *clipped* pixel $\underline{x}_i$ is defined by

$$\underline{x}_i = \max(0, \min(x_i, B)), \quad (8.5)$$

where x_i follows the heteroscedastic noise model. As explained in [155], the generalized pdf of clipped pixel $\underline{x}_i$ is defined by

$$f_{\theta_i}(t) = \Phi\left(\frac{-\mu_i}{\sigma_i}\right)\delta_0(t) + \frac{\frac{1}{\sigma_i}\phi\left(\frac{t-\mu_i}{\sigma_i}\right)}{\Phi\left(\frac{B-\mu_i}{\sigma_i}\right) - \Phi\left(\frac{-\mu_i}{\sigma_i}\right)} + \Phi\left(\frac{\mu_i - B}{\sigma_i}\right)\delta_0(B - t), \quad (8.6)$$

where δ_0 is the Dirac delta impulse at 0. The first and the last term in (8.6) correspond to the probabilities of clipping from below and from above while the second term represents the pdf of a truncated Gaussian random variables (see details about the truncated Gaussian distribution in [160]). The clipping phenomenon presents two main difficulties. On the one hand, it modifies the statistical distribution of RAW pixels, which now follow the double-censored Gaussian distribution (8.6). On the other hand, the pixel's variance can no longer be treated as the linear function of its expectation. Similar to Section 8.2.1, the impact of quantization is assumed to be negligible, i.e. the pmf $P_{\theta_i} = \{p_{\theta_i}[0], p_{\theta_i}[1], \dots, p_{\theta_i}[B]\}$ of the quantized clipped RAW pixel $\underline{z}_i$ can be defined as (8.6).

8.3 GLRT for Non-Clipped Images

8.3.1 Impact of LSB Replacement: Stego-Image Model

Suppose the cover-image $\mathbf{C} = (c_1, \dots, c_N)^T$ and its probability distribution is given in Equation (8.2). The LSB replacement technique is performed by replacing a ratio R of the cover pixels' LSB by the secret message bits to create a stego-image $\mathbf{S} = (s_1, \dots, s_N)^T$, see Section 2.5.1. Due to message embedding, the pmf of cover pixels c_i is changed. Let Q_{R,θ_i} , given by

$$Q_{R,\theta_i} = \{q_{R,\theta_i}[0], q_{R,\theta_i}[1], \dots, q_{R,\theta_i}[B]\}, \quad (8.7)$$

be the pmf of the stego-pixel s_i at embedding rate R . The statistical model of LSB replacement mechanism was described in Proposition 2.1. From the total probability theorem, the probability $q_{R,\theta_i}[l]$ is given by

$$\begin{aligned} q_{R,\theta_i}[l] &= \mathbb{P}[s_i = \bar{c}_i] \mathbb{P}[c_i = \bar{l}] + \mathbb{P}[s_i = c_i] \mathbb{P}[c_i = l] \\ &= \frac{R}{2} p_{\theta_i}[\bar{l}] + \left(1 - \frac{R}{2}\right) p_{\theta_i}[l], \end{aligned} \quad (8.8)$$

where $\bar{l}$ indicates the integer l with LSB flipped $\bar{l} = l + (-1)^l$.

8.3.2 Hypothesis Testing Formulation

The hidden bit detection problem is cast in the framework of hypothesis testing theory. Given a non-clipped RAW image $\mathbf{Z} = (z_1, \dots, z_N)^T$, the steganalysis problem described in Definition 2.2 can be rewritten as

$$\begin{cases} \mathcal{H}_0 &= \{z_i \sim P_{\theta_i}, \forall i \in \{1, \dots, N\}\} \\ \mathcal{H}_1 &= \{z_i \sim Q_{R,\theta_i}, \forall i \in \{1, \dots, N\}\}. \end{cases} \quad (8.9)$$

Let

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

be the class of tests whose false alarm probability is upper-bounded by the rate α_0 where $\boldsymbol{\theta} = (\theta_1, \dots, \theta_N)$ is the parameter vector. Among all the tests in $\mathcal{K}_{\alpha_0}$, it is aimed at finding a test δ which maximizes the power function $\beta(\delta)$.

The problem (8.9) involves two main difficulties. First, the hypotheses $\mathcal{H}_0$ and $\mathcal{H}_1$ are composite because the embedding rate R is unknown. Second, the parameters $\boldsymbol{\theta}$ are unknown in practice. Moreover, they do not contain any information about the existence of the secret message. Hence they can be considered as nuisance parameters.

For the first difficulty, an approach proposed in [113] attempted to reduce the problem with two composite hypotheses to the one with two simple hypotheses based on a local asymptotic approach [183]. The idea is to design a test that is AUMP in the neighborhood of the border between two hypotheses. This test is referred to as Locally Asymptotically Uniformly Most Powerful (LAUMP) test. However it showed a considerable loss of optimality.

This chapter addresses the difficulty of unknown nuisance parameters θ_i while the embedding rate R is assumed to be known in advance. Since the parameters (a, b) can be accurately estimated from a single image (see WLS estimation in Section 4.2.2 and ML estimation in [133]), the main goal of this section is to design a GLRT to deal with unknown parameters μ_i for the steganalysis of LSB replacement in non-clipped RAW image.

8.3.3 ML Estimation of Image Parameters

For estimation of image parameters μ_i , the proposed method consists of two key steps: image denoising and image segmentation. The denoising step aims to estimate approximately the image structure. Then the segmentation step is based on the approximate image structure to divide the image in question into non-overlapping homogeneous segments in which the pixels are supposedly i.i.d. These steps allow us to provide analytically ML estimates of μ_i in each segment.

A majority of existing denoising algorithms in the literature is designed based on AWGN model. However, as discussed in Chapter 3, noise in a natural image is inherently signal-dependent. More particularly, noise variance in a RAW image linearly depends on the image content according to the heteroscedastic noise model. Since this chapter deals with LSB replacement in a RAW image, the heteroscedasticity of noise needs to be taken into account in the denoising step.

There exist two approaches proposed in the literature for the denoising of an image corrupted by signal-dependent noise. The first approach consists in directly designing a denoising filter for such heteroscedastic noise [184, 185]. The second exploits a variance-stabilizing homomorphic transformation [186] so that noise in the image after transformation becomes signal-independent, which allows us to apply a homoscedastic filter for AWGN on the transformed image [155, 187, 188]. Especially, the denoising filter designed in [155] has taken into account the clipping phenomenon. The denoising filter proposed in this chapter relies on the second approach, which involves similar steps as proposed in [155]. For the sake of clarity, the proposed filter is detailed in Appendix 8.7.1.

Formally, let $\mathcal{D}$ be the ideal denoising filter. It means that the image content of a noisy non-clipped image can be very accurately recovered:

$$\mathcal{D}(z_i) = \mathbb{E}[z_i] = \mu_i. \quad (8.10)$$

The segmentation step consists in diving the image $\mathbf{Z}$ into K non-overlapping segments S_k of size n_k , $k \in \{1, \dots, K\}$. Each segment S_k , which is characterized by its central value u_k and allowed deviation $\kappa_k > 0$, is defined as :

$$S_k = \left\{ z_i : \mathcal{D}(z_i) \in \left[u_k - \frac{\kappa_k}{2}, u_k + \frac{\kappa_k}{2} \right], i \in \{1, \dots, N\} \right\}. \quad (8.11)$$

The number of segments K is set to the number of quantization levels, e.g. $K = 2^\nu$ and $\kappa_k = 1$. Without loss of generality, suppose that each segment S_k is represented by a vector $\mathbf{z}_k = \{z_{k,i}\}_{i=1}^{n_k}$ where pixels share the same expectation μ_k . Thus the parameters $\boldsymbol{\theta}_k = (\mu_k, a, b)$ characterizes the statistical distribution of pixels in the segment S_k . Consequently, the ML estimate of the image parameter μ_k in the segment S_k is given by

$$\hat{\mu}_k = \frac{1}{n_k} \sum_{i=1}^{n_k} z_{k,i}. \quad (8.12)$$

This ML estimate is unbiased and follows the Normal distribution

$$\hat{\mu}_k \sim \mathcal{N}\left(\mu_k, \frac{\sigma_k^2}{n_k}\right), \quad \text{with} \quad \sigma_k^2 = a\mu_k + b. \quad (8.13)$$

Remark 8.1. A method for ML estimation of image parameters μ_k has been already presented in Section 4.2.2 when dealing with a natural RAW image. That method involves transforming the RAW image into wavelet domain and performing the ML estimation of image parameters in this domain. However, it seems impossible to model statistically the mechanism of LSB replacement in the wavelet domain. Since the secret message bits are embedded in the spatial domain, it is required to perform ML estimation of image parameters directly in this domain. The method proposed in this section serves this purpose.

Remark 8.2. It can be noted that the latter method is also based on similar steps as the former. The main difference is that the former method works in a blind manner, providing estimates of the parameters (μ_i, a, b) simultaneously, while the latter method requires prior knowledge of the camera parameters (a, b) and only provides estimates of the image parameters μ_i .

8.3.4 Design of GLRT

Using above segmentation, the steganalysis problem (8.9) can be equivalently rewritten as

$$\begin{cases} \mathcal{H}_0 &= \{z_{k,i} \sim P_{\theta_k}, \forall k \in \{1, \dots, K\}, \forall i \in \{1, \dots, n_k\}\} \\ \mathcal{H}_1 &= \{z_{k,i} \sim Q_{R, \theta_k}, \forall k \in \{1, \dots, K\}, \forall i \in \{1, \dots, n_k\}\}. \end{cases} \quad (8.14)$$

When all the parameters θ_k are known, the MP test solving the problem (8.14) is the LRT given by the following decision rule

$$\delta_{\text{ncl}}^*(\mathbf{Z}) = \begin{cases} \mathcal{H}_0 & \text{if } \Lambda_{\text{ncl}}(\mathbf{Z}) = \sum_{k=1}^K \sum_{i=1}^{n_k} \Lambda_{\text{ncl}}(z_{k,i}) < \tau_{\text{ncl}}^* \\ \mathcal{H}_1 & \text{if } \Lambda_{\text{ncl}}(\mathbf{Z}) = \sum_{k=1}^K \sum_{i=1}^{n_k} \Lambda_{\text{ncl}}(z_{k,i}) \geq \tau_{\text{ncl}}^*, \end{cases} \quad (8.15)$$

where the LR $\Lambda_{\text{ncl}}(z_{k,i})$ is defined by

$$\begin{aligned} \Lambda_{\text{ncl}}(z_{k,i}) &= \log \frac{q_{R, \theta_k}[z_{k,i}]}{p_{\theta_k}[z_{k,i}]} = \log \left(\frac{R}{2} \frac{p_{\theta_k}[\bar{z}_{k,i}]}{p_{\theta_k}[z_{k,i}]} + 1 - \frac{R}{2} \right) \\ &= \log \left[1 + \frac{R}{2} \left(\exp \left(\frac{\gamma_{k,i}(z_{k,i} - \mu_k)}{\sigma_k^2} \right) - 1 \right) \right], \end{aligned} \quad (8.16)$$

where $\gamma_{k,i} = z_{k,i} - \bar{z}_{k,i}$ represents the embedding impact and takes value of 1 and -1 depending on the parity of $z_{k,i}$. By using the first-order series Taylor expansion of $\log(1+x)$ and $\exp(x)$, a simplified expression of the LR $\Lambda_{\text{ncl}}(z_{k,i})$ is derived as

$$\Lambda_{\text{ncl}}(z_{k,i}) = \frac{R}{2\sigma_k^2} \gamma_{k,i}(z_{k,i} - \mu_k). \quad (8.17)$$

In practice the image parameters μ_k are unknown. In such situation, a usual solution consists in replacing the unknown parameters by their ML estimates in the LR $\Lambda_{\text{ncl}}(z_{k,i})$ (8.17), which leads to the GLRT defined by

$$\hat{\delta}_{\text{ncl}}(\mathbf{Z}) = \begin{cases} \mathcal{H}_0 & \text{if } \hat{\Lambda}_{\text{ncl}}(\mathbf{Z}) = \sum_{k=1}^K \sum_{i=1}^{n_k} \hat{\Lambda}_{\text{ncl}}(z_{k,i}) < \hat{\tau}_{\text{ncl}} \\ \mathcal{H}_1 & \text{if } \hat{\Lambda}_{\text{ncl}}(\mathbf{Z}) = \sum_{k=1}^K \sum_{i=1}^{n_k} \hat{\Lambda}_{\text{ncl}}(z_{k,i}) \geq \hat{\tau}_{\text{ncl}}, \end{cases} \quad (8.18)$$

where the GLR $\hat{\Lambda}_{\text{ncl}}(z_{k,i})$ is given by

$$\hat{\Lambda}_{\text{ncl}}(z_{k,i}) = \frac{1}{\hat{\sigma}_k^2} \gamma_{k,i}(z_{k,i} - \hat{\mu}_k), \quad \text{with } \hat{\sigma}_k^2 = a\hat{\mu}_k + b. \quad (8.19)$$

Theorem 8.1. *Under hypothesis $\mathcal{H}_0$, the GLR $\hat{\Lambda}_{\text{ncl}}(\mathbf{Z})$ follows the Gaussian distribution with zero-mean and variance $v_{\text{ncl}}^{(0)}$ as*

$$\hat{\Lambda}_{\text{ncl}}(\mathbf{Z}) \xrightarrow{d} \mathcal{N}(0, v_{\text{ncl}}^{(0)}), \quad (8.20)$$

where

$$v_{\text{ncl}}^{(0)} = \sum_{k=1}^K \frac{n_k + 1}{\sigma_k^2 + \frac{a^2}{n_k}}. \quad (8.21)$$

Proof. of Theorem 8.1 is given in Appendix 8.7.2. $\square$

Theorem 8.2. *Under hypothesis $\mathcal{H}_1$, the GLR $\hat{\Lambda}_{\text{ncl}}(\mathbf{Z})$ follows the Gaussian distribution with expectation $m_{\text{ncl}}^{(1)}$ and variance $v_{\text{ncl}}^{(1)}$ as*

$$\hat{\Lambda}_{\text{ncl}}(\mathbf{Z}) \xrightarrow{d} \mathcal{N}(m_{\text{ncl}}^{(1)}, v_{\text{ncl}}^{(1)}) \quad (8.22)$$

where

$$m_{\text{ncl}}^{(1)} = \frac{R}{2} \sum_{k=1}^K \frac{n_k}{\sigma_k^2} \quad (8.23)$$

$$v_{\text{ncl}}^{(1)} = \sum_{k=1}^K \left(\frac{n_k + 1 + \frac{n_k R}{2\sigma_k^2}}{\sigma_k^2 + \frac{a^2}{n_k}} - \frac{R^2 n_k}{4\sigma_k^4} \right). \quad (8.24)$$

Proof. of Theorem 8.2 is given in Appendix 8.7.3. $\square$

It can be noted that $(m_{\text{ncl}}^{(1)}, v_{\text{ncl}}^{(1)})$ depend on the camera parameters (a, b) and embedding rate R . The smaller the embedding rate R is or the more important the noise variance characterized by parameters (a, b) is, the smaller the detectability is, because a small change in the LSB due to the message insertion is not significant compared to the noise level.

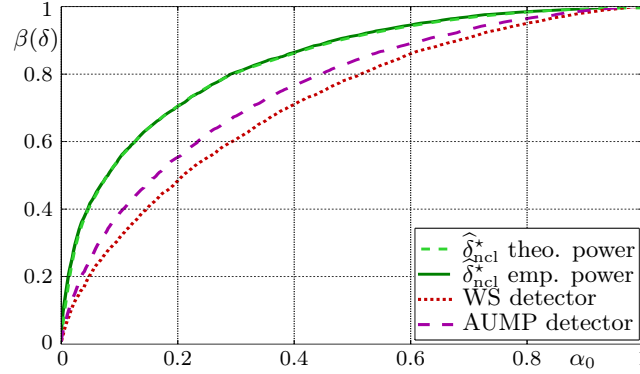


Figure 8.1: Detection performance on non-clipped simulated images for embedding rate $R = 0.05$.

The normalized GLR $\hat{\Lambda}_{\text{ncl}}^*(\mathbf{Z})$ is defined by

$$\hat{\Lambda}_{\text{ncl}}^*(\mathbf{Z}) = \frac{\hat{\Lambda}_{\text{ncl}}(\mathbf{Z})}{\sqrt{\hat{v}_{\text{ncl}}^{(0)}}} = \sum_{k=1}^K \sum_{i=1}^{n_k} \frac{1}{\hat{\sigma}_k^2 \sqrt{\hat{v}_{\text{ncl}}^{(0)}}} (z_{k,i} - \bar{z}_{k,i})(z_{k,i} - \hat{\mu}_k). \quad (8.25)$$

where $\hat{v}_{\text{ncl}}^{(0)}$ is the estimate of the variance $v_{\text{ncl}}^{(0)}$ by replacing μ_k by $\hat{\mu}_k$ in (8.21). It is worth noting that the normalized GLR $\hat{\Lambda}_{\text{ncl}}^*(\mathbf{Z})$ has a similar form of WS detector. Accordingly, the corresponding GLRT $\hat{\delta}_{\text{ncl}}^*$ based on the normalized GLR $\hat{\Lambda}_{\text{ncl}}^*(\mathbf{Z})$ is given as follows

$$\hat{\delta}_{\text{ncl}}^*(\mathbf{Z}) = \begin{cases} \mathcal{H}_0 & \text{if } \hat{\Lambda}_{\text{ncl}}^*(\mathbf{Z}) < \hat{\tau}_{\text{ncl}}^* \\ \mathcal{H}_1 & \text{if } \hat{\Lambda}_{\text{ncl}}^*(\mathbf{Z}) \geq \hat{\tau}_{\text{ncl}}^* \end{cases} \quad (8.26)$$

where, to ensure the GLRT $\hat{\delta}_{\text{ncl}}^*$ to be in the class $\mathcal{K}_{\alpha_0}$, the decision threshold $\hat{\tau}_{\text{ncl}}^*$ is the solution of the equation $\mathbb{P}_{\mathcal{H}_0}[\hat{\Lambda}_{\text{ncl}}^*(\mathbf{Z}) \geq \hat{\tau}_{\text{ncl}}^*] = \alpha_0$.

Theorem 8.3. For any non-clipped RAW image whose parameters μ_i are unknown, the decision threshold and the power of the GLRT $\hat{\delta}_{\text{ncl}}^*$ based on the decision function $\hat{\Lambda}_{\text{ncl}}^*(\mathbf{Z})$ are given by :

$$\hat{\tau}_{\text{ncl}}^* = \Phi^{-1}(1 - \alpha_0) \quad (8.27)$$

$$\beta(\hat{\delta}_{\text{ncl}}^*) = 1 - \Phi \left(\frac{\hat{\tau}_{\text{ncl}}^* \sqrt{v_{\text{ncl}}^{(0)}} - m_{\text{ncl}}^{(1)}}{\sqrt{v_{\text{ncl}}^{(1)}}} \right). \quad (8.28)$$

Proof. Since the estimate $\hat{v}_{\text{ncl}}^{(0)}$ is consistent, it follows from the Slutsky's theorem that

$$\begin{cases} \hat{\Lambda}_{\text{ncl}}^*(\mathbf{Z}) \xrightarrow{d} \mathcal{N}(0, 1) & \text{under } \mathcal{H}_0, \\ \hat{\Lambda}_{\text{ncl}}^*(\mathbf{Z}) \xrightarrow{d} \mathcal{N} \left(\frac{m_{\text{ncl}}^{(1)}}{\sqrt{v_{\text{ncl}}^{(0)}}}, \frac{v_{\text{ncl}}^{(1)}}{v_{\text{ncl}}^{(0)}} \right) & \text{under } \mathcal{H}_1. \end{cases} \quad (8.29)$$

Based on the definition of decision threshold and power function, the proof follows immediately. $\square$

8.4 GLRT for Clipped Images

The purpose of this section is to design a GLRT taking into account the clipping phenomenon. This GLRT is designed following similar steps as in Section 8.3 but the calculation (e.g. the statistical distribution of the GLR) is different.

8.4.1 ML Estimation of Image Parameters

To perform ML estimation of image parameters μ_i in a clipped RAW image, it is proposed to rely on similar steps such as image denoising and segmentation. Here it is crucial to understand the relation between the clipped data and non-clipped data [155] when dealing with a clipped image. The output image after performing the denoising filter on a clipped image can only be treated as an estimate of the expectation $\mathbb{E}[z_i]$, not the parameter μ_i . Therefore it is necessary to combine the denoising filter with a declipped operator [155] to transform a clipped value to the corresponding non-clipped value. Let $\underline{\mathcal{D}}$ be the ideal denoising and also declipped operator

$$\underline{\mathcal{D}}(z_i) = \mu_i. \quad (8.30)$$

Then the clipped RAW image $\underline{\mathbf{Z}}$ is segmented into K non-overlapping segments S_k of size n_k , $k \in \{1, \dots, K\}$, defined by

$$S_k = \left\{ z_i : \underline{\mathcal{D}}(z_i) \in \left[u_k - \frac{\kappa_k}{2}, u_k + \frac{\kappa_k}{2} \right], i \in \{1, \dots, N\} \right\}. \quad (8.31)$$

The ML estimate $\hat{\mu}_k$ is given in (8.81). It follows the asymptotic Gaussian distribution

$$\hat{\mu}_k \sim \mathcal{N}\left(\mu_k, \frac{1}{n_k \mathbf{F}(\mu_k)}\right) \quad (8.32)$$

where $\mathbf{F}$ is the Fisher information given in (8.84). More details are given in Appendix 8.7.4.

8.4.2 Design of GLRT

Now the cover image and stego image are respectively denoted by $\underline{\mathbf{C}} = (\underline{c}_1, \dots, \underline{c}_N)^T$ and $\underline{\mathbf{S}} = (\underline{s}_1, \dots, \underline{s}_N)^T$. Here again, without taking into account the impact of quantization, the pmf $\underline{P}_{\theta_i} = \{p_{\theta_i}[0], p_{\theta_i}[1], \dots, p_{\theta_i}[B]\}$ of the cover clipped pixel $\underline{c}_i$ can be defined by (8.6). Due to the insertion of hidden bits, the pmf $\underline{Q}_{R, \theta_i} = \{q_{R, \theta_i}[0], q_{R, \theta_i}[1], \dots, q_{R, \theta_i}[B]\}$ of the stego clipped pixel $\underline{s}_i$ is given by

$$\underline{q}_{R, \theta_i}[l] = \frac{R}{2} p_{\theta_i}[\bar{l}] + \left(1 - \frac{R}{2}\right) p_{\theta_i}[l]. \quad (8.33)$$

Let $\underline{\mathbf{Z}} = \{z_1, \dots, z_N\}^T$ be the inspected clipped image. The hidden data detection problem (8.14) in the case of clipped images is rewritten as follows

$$\begin{cases} \mathcal{H}_0 &= \{z_{k,i} \sim \underline{P}_{\theta_k}, \forall k \in \{1, \dots, K\}, \forall i \in \{1, \dots, n_k\}\} \\ \mathcal{H}_1 &= \{z_{k,i} \sim \underline{Q}_{R, \theta_k}, \forall k \in \{1, \dots, K\}, \forall i \in \{1, \dots, n_k\}\}. \end{cases} \quad (8.34)$$

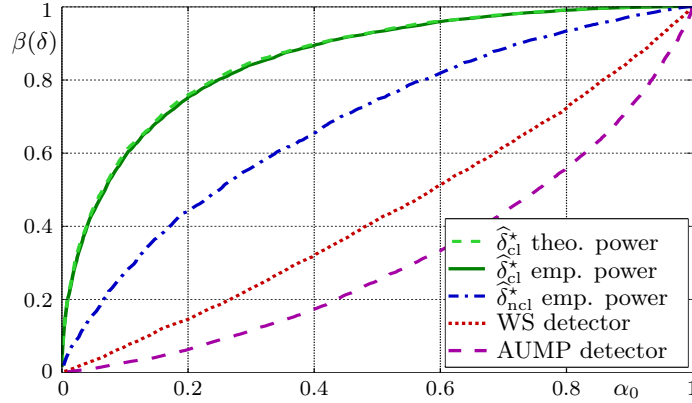


Figure 8.2: Detection performance on clipped simulated images for embedding rate $R = 0.05$.

The LR of one clipped observation $\underline{z}_{k,i}$ can be defined as

$$\begin{aligned}\Lambda_{\text{cl}}(\underline{z}_{k,i}) &= \log \left(\frac{R \underline{p}_{\theta_k}[\underline{z}_{k,i}]}{2 \underline{p}_{\theta_k}[\underline{z}_{k,i}]} + 1 - \frac{R}{2} \right) \\ &= \log \left[1 + \frac{R}{2} \left(\exp \left(\frac{\gamma_{k,i}(\underline{z}_{k,i} - \mu_k)}{\sigma_k^2} \right) - 1 \right) \right],\end{aligned}\quad (8.35)$$

with $\gamma_{k,i} = \underline{z}_{k,i} - \bar{z}_{k,i}$. Here it is proposed to neglect the first and last terms in (8.6) for the sake of simplification. Using again Taylor series expansion, the LR $\Lambda_{\text{cl}}(\underline{z}_{k,i})$ can be simplified as

$$\Lambda_{\text{cl}}(\underline{z}_{k,i}) = \frac{R}{2\sigma_k^2} \gamma_{k,i} (\underline{z}_{k,i} - \mu_k). \quad (8.36)$$

It can be noted that the LR $\Lambda_{\text{cl}}(\underline{z}_{k,i})$ of clipped data has a similar form as the LR $\Lambda_{\text{ncl}}(\underline{z}_{k,i})$ but the image parameter μ_k is now no longer the expected value of the clipped pixel $\underline{z}_{k,i}$. This bias can lead to the loss of power of the GLRT $\hat{\delta}_{\text{ncl}}^*$ when performing on a clipped image. The image parameter μ_k is always unknown in practice.

The GLRT solving the problem (8.34) is given by

$$\hat{\delta}_{\text{cl}}(\mathbf{Z}) = \begin{cases} \mathcal{H}_0 & \text{if } \hat{\Lambda}_{\text{cl}}(\mathbf{Z}) = \sum_{k=1}^K \sum_{i=1}^{n_k} \hat{\Lambda}_{\text{cl}}(\underline{z}_{k,i}) < \hat{\tau}_{\text{cl}} \\ \mathcal{H}_1 & \text{if } \hat{\Lambda}_{\text{cl}}(\mathbf{Z}) = \sum_{k=1}^K \sum_{i=1}^{n_k} \hat{\Lambda}_{\text{cl}}(\underline{z}_{k,i}) \geq \hat{\tau}_{\text{cl}}, \end{cases} \quad (8.37)$$

where the GLR $\hat{\Lambda}_{\text{cl}}(\underline{z}_{k,i})$ is defined by replacing the unknown parameter μ_k by the ML estimate $\hat{\mu}_k$ in the LR $\Lambda_{\text{cl}}(\underline{z}_{k,i})$ (8.36)

$$\hat{\Lambda}_{\text{cl}}(\underline{z}_{k,i}) = \frac{1}{\hat{\sigma}_k^2} \gamma_{k,i} (\underline{z}_{k,i} - \hat{\mu}_k). \quad (8.38)$$

Theorem 8.4. Under hypothesis $\mathcal{H}_j$, the GLR $\hat{\Lambda}_{\text{cl}}(\underline{\mathbf{Z}})$ follows the Gaussian distribution with expectation $m_{\text{cl}}^{(j)}$ and variance $v_{\text{cl}}^{(j)}$, defined as

$$\hat{\Lambda}_{\text{cl}}(\underline{\mathbf{Z}}) \xrightarrow{d} \mathcal{N}(m_{\text{cl}}^{(j)}, v_{\text{cl}}^{(j)}), \quad (8.39)$$

with

$$m_{\text{cl}}^{(0)} = 0 \quad (8.40)$$

$$v_{\text{cl}}^{(0)} = \sum_{k=1}^K \frac{n_k \sigma_k^2 \left(1 + t_{0,k} T_{0,k} - t_{B,k} T_{B,k}\right) + \frac{1}{\mathbf{F}(\mu_k)}}{\sigma_k^4 + \frac{a^2}{n_k \mathbf{F}(\mu_k)}} \quad (8.41)$$

$$m_{\text{cl}}^{(1)} = \frac{R}{2} \sum_{k=1}^K \frac{n_k}{\sigma_k^2} \quad (8.42)$$

$$v_{\text{cl}}^{(1)} = \sum_{k=1}^K \left[\frac{n_k \sigma_k^2 \left(1 + t_{0,k} T_{0,k} - t_{B,k} T_{B,k}\right) + \frac{1}{\mathbf{F}(\mu_k)} + \frac{n_k R}{2}}{\sigma_k^4 + \frac{a^2}{n_k \mathbf{F}(\mu_k)}} - \frac{n_k R^2}{4\sigma_k^4} \right], \quad (8.43)$$

and the terms $t_{0,k}$, $t_{B,k}$, $T_{0,k}$, $T_{B,k}$ take into account the impact of truncation in each segment S_k

$$t_{0,k} = -\frac{\mu_k}{\sigma_k} \quad \text{and} \quad t_{B,k} = \frac{B - \mu_k}{\sigma_k} \quad (8.44)$$

$$T_{0,k} = \frac{\phi(t_{0,k})}{\Phi(t_{B,k}) - \Phi(t_{0,k})} \quad \text{and} \quad T_{B,k} = \frac{\phi(t_{B,k})}{\Phi(t_{B,k}) - \Phi(t_{0,k})}. \quad (8.45)$$

Proof. of Theorem 8.4 is given in Appendix 8.7.5. $\square$

Therefore, the normalized GLR $\hat{\Lambda}_{\text{cl}}^*(\underline{\mathbf{Z}})$ is defined by

$$\hat{\Lambda}_{\text{cl}}^*(\underline{\mathbf{Z}}) = \frac{\hat{\Lambda}_{\text{cl}}(\underline{\mathbf{Z}})}{\sqrt{\hat{v}_{\text{cl}}^{(0)}}} = \sum_{k=1}^K \sum_{i=1}^{n_k} \frac{1}{\hat{\sigma}_k^2 \sqrt{\hat{v}_{\text{cl}}^{(0)}}} (z_{k,i} - \bar{z}_{k,i})(z_{k,i} - \hat{\mu}_k) \quad (8.46)$$

where $\hat{v}_{\text{cl}}^{(0)}$ is the estimate of $v_{\text{cl}}^{(0)}$ defined by replacing μ_k by $\hat{\mu}_k$ in (8.41). The GLRT $\hat{\delta}_{\text{cl}}^*$ based on the normalized GLR $\hat{\Lambda}_{\text{cl}}^*(\underline{\mathbf{Z}})$ is defined as follows

$$\hat{\delta}_{\text{cl}}^*(\underline{\mathbf{Z}}) = \begin{cases} \mathcal{H}_0 & \text{if } \hat{\Lambda}_{\text{cl}}^*(\underline{\mathbf{Z}}) < \hat{\tau}_{\text{cl}}^* \\ \mathcal{H}_1 & \text{if } \hat{\Lambda}_{\text{cl}}^*(\underline{\mathbf{Z}}) \geq \hat{\tau}_{\text{cl}}^* \end{cases} \quad (8.47)$$

Theorem 8.5. For any clipped RAW image whose parameters μ_i are unknown, the decision threshold and the power of the GLRT $\hat{\delta}_{\text{cl}}^*$ based on the decision function $\hat{\Lambda}_{\text{cl}}^*(\underline{\mathbf{Z}})$ are given by

$$\hat{\tau}_{\text{cl}}^* = \Phi^{-1}(1 - \alpha_0) \quad (8.48)$$

$$\beta(\hat{\delta}_{\text{cl}}^*) = 1 - \Phi \left(\frac{\hat{\tau}_{\text{cl}}^* \sqrt{v_{\text{cl}}^{(0)}} - m_{\text{cl}}^{(1)}}{\sqrt{v_{\text{cl}}^{(1)}}} \right). \quad (8.49)$$

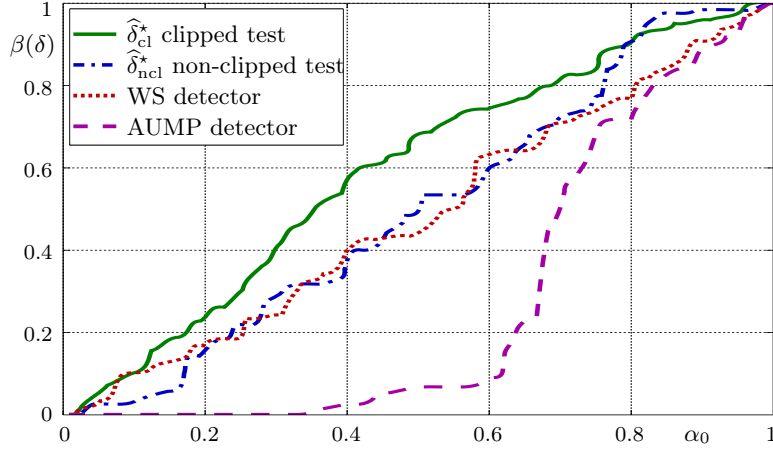


Figure 8.3: Detection performance on real clipped images for embedding rate $R = 0.2$.

Proof. According to Theorem 8.4, based on the definition of decision threshold and power function, the proof follows immediately. $\square$

The clipping phenomenon whose impact on any steganalysis detector has never been studied yet has been taken into account in the design of the proposed GLRT. It can be noted that when the clipping phenomenon does not happen in a natural RAW image, the power function $\beta(\hat{\delta}_{cl}^*)$ tends to $\beta(\hat{\delta}_{ncl}^*)$. In other words, $\beta(\hat{\delta}_{ncl}^*)$ can be regarded as a lower bound of $\beta(\hat{\delta}_{cl}^*)$ when not taking into account the clipping.

8.5 Numerical Experiments

8.5.1 Detection Performance on Simulated Database

The detection performance of the proposed tests $\hat{\delta}_{ncl}^*$ and $\hat{\delta}_{cl}^*$ is first theoretically studied on simulated database. The parameters (a, b) characterizing the heteroscedastic noise model are set at $a = 0.0115$ and $b = 0.0002$. These values correspond to the *Nikon D70* camera with ISO 200 estimated from the Dresden image database [168]. These parameters are used with 8-bit synthetic image of size 512×512 to generate randomly 5000 cover images. The secret message bits are drawn from a binomial distribution $\mathcal{B}(1, 1/2)$ and then embedded in pixel's LSB. The number of segments K is set to the number of grayscale levels, $K = 2^8$ and the embedding rate R is set at 0.05. Potentially, a large number of detectors could be compared with the proposed tests. In this chapter, it is proposed to include in the comparison the revised version of WS [103, Eq.(2)], because it is well-known for a good detection performance. In addition, it is also proposed to compare the proposed tests with the AUMP detector [109] which is also designed within hypothesis testing framework based on a simplistic piece-wise polynomial model. The Figure 8.1 and Figure 8.2 illustrate respectively the detection performance of all detectors in case of non-clipped images and clipped images. In both scenario, the empirical power of the

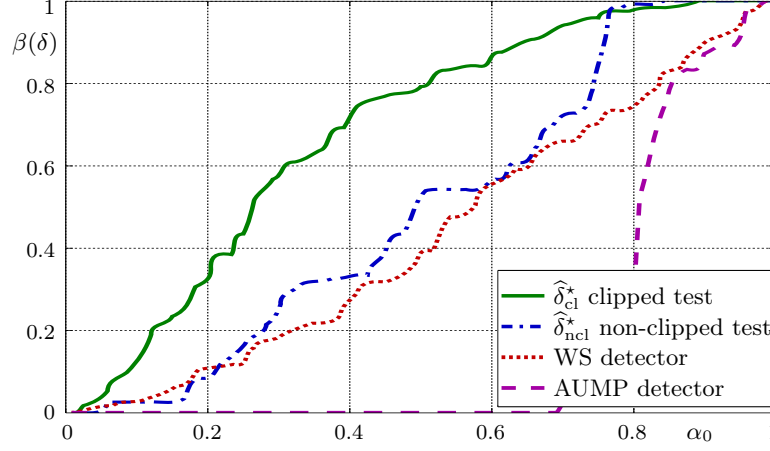


Figure 8.4: Detection performance on real clipped images for embedding rate $R = 0.4$.

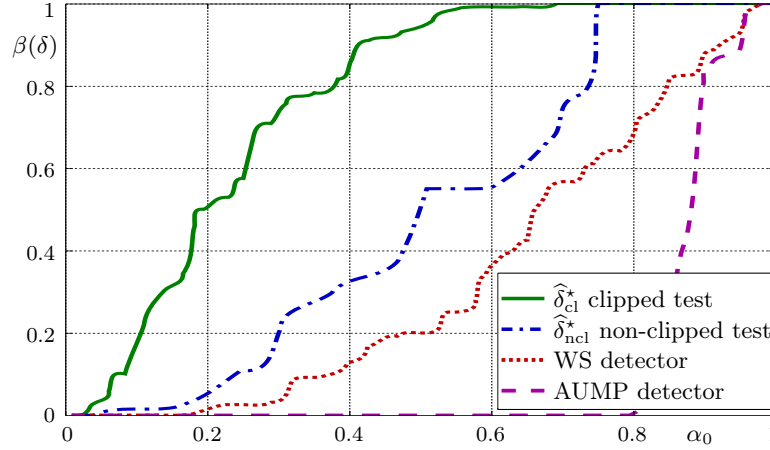


Figure 8.5: Detection performance on real clipped images for embedding rate $R = 0.6$.

proposed GLRTs $\hat{\delta}_{\text{ncl}}^*$ and $\hat{\delta}_{\text{cl}}^*$ is identical to the theoretical one, which shows no loss of optimality. It can be noted from Figure 8.1 that since the AUMP test only utilizes the local polynomial model that can not provide accurate estimation of pixels' expectation and variance, this bias leads to a loss of optimality of the AUMP test even in the simulated database [109]. The Figure 8.2 also shows that the AUMP test can not tolerate the impact of the clipping phenomenon. These experiments emphasize the importance of taking into account both the heteroscedastic noise model and the clipping phenomenon. The heteroscedastic noise model, which is more relevant to characterize a RAW image, allows an accurate estimation of pixels' expectation and variance, thus improving the detection performance of the GLRTs. To deal with the clipping phenomenon, it is crucial to rely on the heteroscedastic noise model.

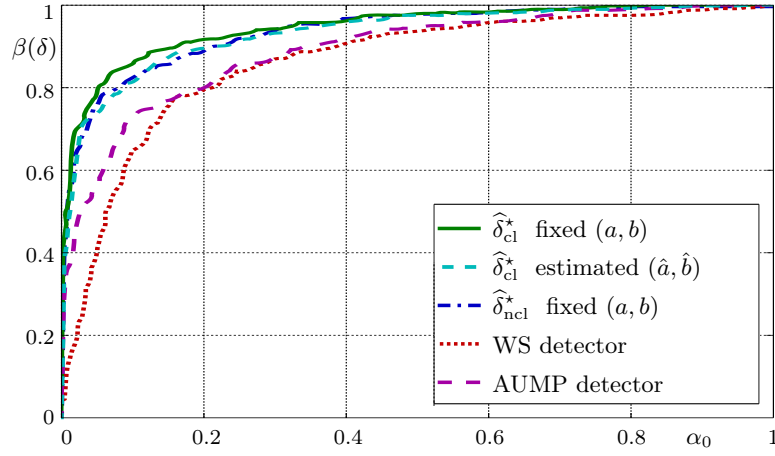


Figure 8.6: Detection performance on 12-bit images taken by Canon 400D with ISO 100 from BOSS database for embedding rate $R = 0.05$.

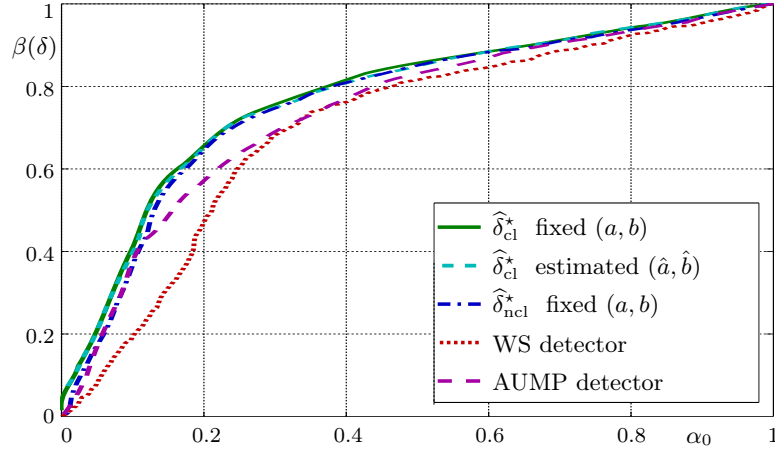


Figure 8.7: Detection performance on 5000 images from BOSS database for embedding rate $R = 0.05$.

8.5.2 Detection Performance on Real Image Database

Experiments are then conducted on real image database to highlight the relevance of the proposed GLRTs. Real clipped images captured by the *Nikon D300* camera are provided at <http://www.cs.tut.fi/~foi/sensornoise>. Prior to our experiments, every RAW image was converted to an uncompressed image format using Ddraw (with parameters -D - T -4 -j -v). Only the red color channel is used in these experiments. The camera parameters are estimated on each clipped image based on the ML approach [133]. The reference parameters (a, b) are then provided by averaging the previously estimated values. The denoising filter proposed in [155] is employed for estimation of image parameters μ_i in a clipped image. The images are standardized to the size of 512×512 by cropping. The experiments are respectively performed for the embedding rate $R \in \{0.2, 0.4, 0.6\}$. The detection performance for each embedding rate is illustrated in Figure 8.3, 8.4 and 8.5. Similar to the case of

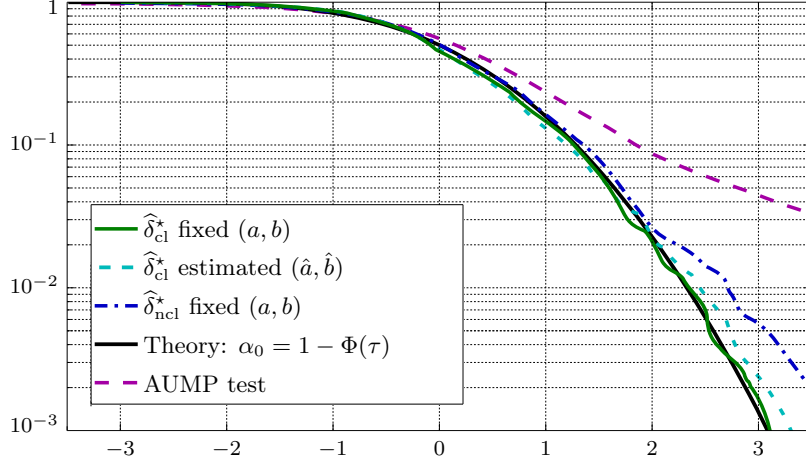


Figure 8.8: Empirical false-alarm probability from real images of BOSS database plotted as a function of decision threshold, compared with theoretical FAP.

simulation, the proposed GLRT $\hat{\delta}_{cl}^*$ remains ensuring a high detection performance while the WS and AUMP detector fail.

To verify the efficiency of the proposed approach, experiments are conducted a large image data base with 5000 RAW images from the BOSS base [171]. On the one hand, the parameters $(\hat{a}, \hat{b})$ are estimated on each image. On the other hand, the reference parameters (a, b) are provided by averaging the previously estimates obtained from each image with the same ISO sensitivity per camera model. The power function of both GLRTs $\hat{\delta}_{ncl}^*$ and $\hat{\delta}_{cl}^*$ is drawn using the fixed camera parameters (a, b) and the estimated $(\hat{a}, \hat{b})$. The Figure 8.6 shows the detection performance of all detectors on 12-bit RAW images of the *Canon 400D* camera with ISO 100. The Figure 8.7 shows the detection performance on 5000 RAW images from the BOSS base [171]. The clipping phenomenon seldom happens in the images of the *Canon 400D* camera with ISO 100. Even in non-clipped images, the proposed tests also outperform other detectors (see Figure 8.6). Moreover, it is worth noting that there is no great difference between the test $\hat{\delta}_{cl}^*$ with the reference parameters (a, b) estimated from all image with the same ISO sensitivity per camera model and with $(\hat{a}, \hat{b})$ estimated from each image. This indicates that the fact of estimating the camera parameters on each image does not importantly affect the power function of the test. In other words, the prior knowledge of camera parameters may not be necessary when designing the GLRTs.

Finally, it is proposed to show the possibility of the proposed test to guarantee in practice a prescribed false alarm probability. To this end, Figure 8.8 shows a comparison between theoretical and empirical false-alarm probability as a function of decision threshold. Note that, because the WS detector aims at estimating the embedding rate and does not enable to warrant a prescribed false alarm probability, this detector is omitted in this experiment. The Figure 8.8 obviously shows that the AUMP detector [109] fails in practice to meet a prescribed false-alarm probability. This is mainly due to a rather inaccurate image model as well as assumptions which

seldom hold true such as constant variance on each block of pixels. On the opposite, the proposed GLRTs allows us in practice to guarantee a false alarm probability. It should be noted that, again, there is a no important difference of empirical false-alarm probability of the test $\hat{\delta}_{\text{cl}}^*$ with fixed parameters (a, b) and with parameters $(\hat{a}, \hat{b})$ estimated on each image. Hence the use of estimated parameters is also capable in practice to guarantee a false alarm probability. Similarly, the test $\hat{\delta}_{\text{ncl}}^*$ which does not take into account the clipping phenomenon slightly fails to meet a false-alarm probability, particularly for a decision threshold typically corresponding to $\alpha_0 \lesssim 10^{-2}$. This is mainly due to the small number of images whose the number of clipped pixels is not negligible.

8.6 Conclusion

This chapter proposes a statistical test to detect secret data hidden in the LSB plane of a natural RAW image. The hidden data detection problem is cast in the framework of hypothesis testing theory. The approach is based on the heteroscedastic noise model that characterizes the RAW image more accurately than the AWGN or a piece-wise polynomial model. The GLRTs are designed to deal with the difficulty of unknown image parameters, which can be straightforwardly applied to any natural RAW image. It is also shown that the WS detector is a variant of the proposed GLRTs, similar to other statistical detectors [109–113]. But the proposed approach has gone further by relying on a more accurate image model and exploiting the state-of-the-art denoising filter. This allows a more accurate estimation of pixels' expectation and variance, resulting a considerable improvement of detection performance. The specificity of the proposed approach is that the clipping phenomenon is taken into account in the design of GLRT while other detectors can not tolerate its impact and fail in practice. Moreover, the proposed GLRTs can warrant a prescribed false alarm rate while ensuring a high detection performance. Numerical experiments on simulated database and real image database highlight the relevance of the proposed approach.

8.7 Appendix

8.7.1 Denoising Filter for Non-Clipped RAW Images Corrupted by Signal-Dependent Noise

We propose a denoising filter to remove the heteroscedastic noise corrupting non-clipped RAW images. Firstly, a homomorphic transformation is employed to transform a signal-dependent noise into a signal-independent noise and stabilize pixels' variance to unity. Let denote this homomorphic transformation by the function $f : [0, B] \rightarrow \mathbb{R}^+$. As explained in [186], this function is defined as the solution of the problem

$$f(t) = \int_{-\frac{b}{a}}^t \frac{1}{\sigma(x)} dx = \int_{-\frac{b}{a}}^t \frac{1}{\sqrt{ax+b}} dx. \quad (8.50)$$

Hence, the transformation f can be given by

$$f(t) = \begin{cases} \frac{2}{a}\sqrt{at+b} & t > -\frac{b}{a} \\ 0 & t \leq -\frac{b}{a}. \end{cases} \quad (8.51)$$

The transformation f is monotonically increasing, which implies the existence of the inverse transformation f^{-1} , given by

$$f^{-1}(t) = \frac{a}{4}t^2 - \frac{b}{a}. \quad (8.52)$$

By using the transformation f , one obtains an image $\mathbf{Y}$ that is corrupted by AWGN with unitary variance

$$y \sim \mathcal{N}(\vartheta, 1). \quad (8.53)$$

The index of pixel y is omitted for simplicity. At this stage, a homoscedastic filter $\mathcal{D}_{\text{ho}}$ can be applied to remove the AWGN. Assuming that the free-noise image is successfully recovered, the denoised value $\mathcal{D}_{\text{ho}}(y)$ can be treated as the expected value of y

$$\mathcal{D}_{\text{ho}}(y) = \mathbb{E}[y] = \vartheta. \quad (8.54)$$

However, due to the non-linearity of f , the inverse transformation of the denoised value $\mathcal{D}_{\text{ho}}(y)$ can not be treated as the expected value of the pixel z

$$f^{-1}(\mathcal{D}_{\text{ho}}(y)) = f^{-1}(\vartheta) \neq \mathbb{E}[z] = \mu. \quad (8.55)$$

This leads to a systematic estimation bias, which needs to be corrected.

Let the function h defined by :

$$h : f(\mu) \mapsto \vartheta = h(f(\mu)).$$

By using the Taylor series expansion of $f(z)$ around $z = \mu$, one obtains :

$$f(z) = f(\mu) + \sum_{n=1}^{\infty} \frac{(z-\mu)^n}{n!} f^{(n)}(\mu) \quad (8.56)$$

where $f^{(n)}$ is the n -th derivative of f . The recursive formula of $f^{(n)}$ is given by

$$f^{(n)}(t) = \left(-\frac{a}{2}\right)^{n-1} (2n-3)!! (at+b)^{-n+\frac{1}{2}}, \quad \forall n \geq 2 \quad (8.57)$$

where $n!!$ denotes the double factorial of n . Taking expectations on both sides of (8.56), the calculation shows that

$$\vartheta = \mathbb{E}[f(z)] = f(\mu) - \sum_{k=1}^{\infty} \left(\frac{a}{2}\right)^{2k-1} \frac{(4k-3)!!}{(2k)!!} \frac{1}{\sigma^{2k-1}}. \quad (8.58)$$

Replacing $t = f(\mu)$ in (8.58), it follows that

$$\vartheta = t - \sum_{k=1}^{\infty} \frac{(4k-3)!!}{(2k)!!} \frac{1}{t^{2k-1}}. \quad (8.59)$$

Therefore, the function h can be defined by

$$h(t) = t - \sum_{k=1}^{\infty} \frac{(4k-3)!!}{(2k)!!} \frac{1}{t^{2k-1}}. \quad (8.60)$$

The first derivative of h is positive, hence the function h is monotonically increasing on $\mathbb{R}^+$. This ensures the invertibility of h . Consequently, the noisy RAW pixel is ideally mapped to its denoised value by the relation

$$\mu = f^{-1}(h^{-1}(\mathcal{D}_{\text{ho}}(f(z)))). \quad (8.61)$$

Due to (8.61), the special denoising filter for heteroscedastic noise in non-clipped images can be defined by

$$\mathcal{D} = f \circ \mathcal{D}_{\text{ho}} \circ h^{-1} \circ f^{-1}. \quad (8.62)$$

8.7.2 Statistical distribution of the GLR $\hat{\Lambda}_{\text{ncl}}(\mathbf{Z})$ under hypothesis $\mathcal{H}_0$

For brevity, let denote $\rho_{k,i} = \frac{1}{\hat{\sigma}_k^2}(z_{k,i} - \hat{\mu}_k)$ such that the GLR $\hat{\Lambda}_{\text{ncl}}(z_{k,i})$ can be rewritten as $\hat{\Lambda}_{\text{ncl}}(z_{k,i}) = \gamma_{k,i}\rho_{k,i}$. Based on the total probability theorem, it is easily shown that

$$\begin{aligned} \mathbb{E}_{\mathcal{H}_0}[\hat{\Lambda}_{\text{ncl}}(z_{k,i})] &= \mathbb{P}_{\mathcal{H}_0}[\gamma_{k,i} = 1]\mathbb{E}_{\mathcal{H}_0}[\rho_{k,i} \mid \gamma_{k,i} = 1] - \mathbb{P}_{\mathcal{H}_0}[\gamma_{k,i} = -1]\mathbb{E}_{\mathcal{H}_0}[\rho_{k,i} \mid \gamma_{k,i} = -1] \\ &= \frac{1}{2}\mathbb{E}_{\mathcal{H}_0}[\rho_{k,i}] - \frac{1}{2}\mathbb{E}_{\mathcal{H}_0}[\rho_{k,i}] \\ &= 0, \end{aligned} \quad (8.63)$$

Meanwhile, the variance of the GLR $\hat{\Lambda}_{\text{ncl}}(z_{k,i})$ can be expressed as

$$\mathbb{V}\text{ar}_{\mathcal{H}_0}[\hat{\Lambda}_{\text{ncl}}(z_{k,i})] = \mathbb{E}_{\mathcal{H}_0}[\gamma_{k,i}^2 \rho_{k,i}^2] = \mathbb{E}_{\mathcal{H}_0}[\rho_{k,i}^2] = \mathbb{E}_{\mathcal{H}_0}\left[\frac{(z_{k,i} - \hat{\mu}_k)^2}{\hat{\sigma}_k^4}\right]. \quad (8.64)$$

It follows from (8.13) that

$$\hat{\sigma}_k^2 \sim \mathcal{N}\left(\sigma_k^2, \frac{a^2}{n_k} \sigma_k^2\right). \quad (8.65)$$

Using the second Delta method in Lemma A.3, it follows that

$$\mathbb{V}\text{ar}_{\mathcal{H}_0}[\hat{\Lambda}_{\text{ncl}}(z_{k,i})] = \frac{\mathbb{E}_{\mathcal{H}_0}[(z_{k,i} - \hat{\mu}_k)^2]}{\mathbb{E}_{\mathcal{H}_0}[\hat{\sigma}_k^4]} = \frac{1 + \frac{1}{n_k}}{\sigma_k^2 + \frac{a^2}{n_k}}. \quad (8.66)$$

Finally, from the Lindeberg CLT, we derive the statistical distribution of the GLR $\hat{\Lambda}(\mathbf{Z})$ under hypothesis $\mathcal{H}_0$

$$\hat{\Lambda}_{\text{ncl}}(\mathbf{Z}) \xrightarrow{d} \mathcal{N}(0, v_{\text{ncl}}^{(0)}), \quad (8.67)$$

with

$$v_{\text{ncl}}^{(0)} = \sum_{k=1}^K \sum_{i=1}^{n_k} \mathbb{V}\text{ar}_{\mathcal{H}_0} [\hat{\Lambda}(z_{k,i})] = \sum_{k=1}^K \frac{n_k + 1}{\sigma_k^2 + \frac{a^2}{n_k}}. \quad (8.68)$$

8.7.3 Statistical distribution of the GLR $\hat{\Lambda}_{\text{ncl}}(\mathbf{Z})$ under hypothesis $\mathcal{H}_1$

Under hypothesis $\mathcal{H}_1$, the estimates $\hat{\mu}_k$ are obviously impacted by the insertion of the secret message. However, under assumption that the impact of quantization is negligible and the denoising operator $\mathcal{D}$ is ideal, one can obtain an almost identical denoised image

$$\mathcal{D}(\mathbf{C}) = \mathcal{D}(\mathbf{S}). \quad (8.69)$$

Hence, the probability that $\mathcal{D}(z_i)$ falls into the interval $\left[u_k - \frac{\kappa_k}{2}, u_k + \frac{\kappa_k}{2}\right]$ does not change under hypothesis $\mathcal{H}_1$. It is straightforward to derive that the impact of insertion on the estimates $\hat{\mu}_k$ is negligible.

The image $\mathbf{Z}$ contains both cover and stego-pixels under hypothesis $\mathcal{H}_1$. Hence, the law of total expectation and of total variance allow to calculate the first two moments of $\hat{\Lambda}_{\text{ncl}}(z_{k,i})$. The mathematical expectation of $\hat{\Lambda}_{\text{ncl}}(z_{k,i})$ is defined as

$$\begin{aligned} \mathbb{E}_{\mathcal{H}_1} [\hat{\Lambda}_{\text{ncl}}(z_{k,i})] &= \frac{R}{2} \mathbb{E}_{\mathcal{H}_0} [\hat{\Lambda}_{\text{ncl}}(\bar{z}_{k,i})] + \left(1 - \frac{R}{2}\right) \mathbb{E}_{\mathcal{H}_0} [\hat{\Lambda}_{\text{ncl}}(z_{k,i})] \\ &= \frac{R}{2} \mathbb{E}_{\mathcal{H}_0} [\hat{\Lambda}_{\text{ncl}}(\bar{z}_{k,i})] \\ &= \frac{R}{2} \left(\mathbb{P}_{\mathcal{H}_0}[\bar{\gamma}_{k,i} = 1] \mathbb{E}_{\mathcal{H}_0} \left[\frac{\bar{z}_{k,i} - \hat{\mu}_k}{\hat{\sigma}_k^2} | \bar{\gamma}_{k,i} = 1 \right] \right. \\ &\quad \left. - \mathbb{P}_{\mathcal{H}_0}[\bar{\gamma}_{k,i} = -1] \mathbb{E}_{\mathcal{H}_0} \left[\frac{\bar{z}_{k,i} - \hat{\mu}_k}{\hat{\sigma}_k^2} | \bar{\gamma}_{k,i} = -1 \right] \right) \\ &= \frac{R}{2} \left(\frac{1}{2} \mathbb{E}_{\mathcal{H}_0} \left[\frac{z_{k,i} + 1 - \hat{\mu}_k}{\hat{\sigma}_k^2} \right] - \frac{1}{2} \mathbb{E}_{\mathcal{H}_0} \left[\frac{z_{k,i} - 1 - \hat{\mu}_k}{\hat{\sigma}_k^2} \right] \right) \\ &= \frac{R}{2} \mathbb{E}_{\mathcal{H}_0} \left[\frac{1}{\hat{\sigma}_k^2} \right] \\ &= \frac{R}{2\sigma_k^2}, \end{aligned} \quad (8.70)$$

with $\bar{\gamma}_{k,i} = \bar{z}_{k,i} - \left(\bar{z}_{k,i} + (-1)^{\bar{z}_{k,i}}\right) = (-1)^{\bar{z}_{k,i}} = \bar{z}_{k,i} - z_{k,i}$.

Similarly, a direct calculation yields to

$$\begin{aligned}
\mathbb{E}_{\mathcal{H}_1} \left[\widehat{\Lambda}_{\text{ncl}}^2(z_{k,i}) \right] &= \mathbb{E}_{\mathcal{H}_1} \left[\frac{(z_{k,i} - \hat{\mu}_k)^2}{\hat{\sigma}_k^4} \right] \\
&= \frac{R}{2} \mathbb{E}_{\mathcal{H}_0} \left[\frac{(\bar{z}_{k,i} - \hat{\mu}_k)^2}{\hat{\sigma}_k^4} \right] + \left(1 - \frac{R}{2} \right) \mathbb{E}_{\mathcal{H}_0} \left[\frac{(z_{k,i} - \hat{\mu}_k)^2}{\hat{\sigma}_k^4} \right] \\
&= \mathbb{E}_{\mathcal{H}_0} \left[\frac{(z_{k,i} - \hat{\mu}_k)^2}{\hat{\sigma}_k^4} \right] + \frac{R}{2} \mathbb{E}_{\mathcal{H}_0} \left[\frac{1}{\hat{\sigma}_k^4} \right] \\
&= \frac{1 + \frac{1}{n_k} + \frac{R}{2\sigma_k^2}}{\sigma_k^2 + \frac{a^2}{n_k}}.
\end{aligned} \tag{8.71}$$

It follows that

$$\begin{aligned}
\text{Var}_{\mathcal{H}_1} \left[\widehat{\Lambda}_{\text{ncl}}(z_{k,i}) \right] &= \mathbb{E}_{\mathcal{H}_1} \left[\widehat{\Lambda}_{\text{ncl}}^2(z_{k,i}) \right] - \mathbb{E}_{\mathcal{H}_1}^2 \left[\widehat{\Lambda}_{\text{ncl}}(z_{k,i}) \right] \\
&= \frac{1 + \frac{1}{n_k} + \frac{R}{2\sigma_k^2}}{\sigma_k^2 + \frac{a^2}{n_k}} - \frac{R^2}{4\sigma_k^4}.
\end{aligned} \tag{8.72}$$

Hence, using again the Lindeberg CLT, one obtains the statistical distribution of the GLR $\widehat{\Lambda}_{\text{ncl}}(\mathbf{Z})$ under hypothesis $\mathcal{H}_1$

$$\widehat{\Lambda}_{\text{ncl}}(\mathbf{Z}) \xrightarrow{d} \mathcal{N} \left(m_{\text{ncl}}^{(1)}, v_{\text{ncl}}^{(1)} \right) \tag{8.73}$$

where m_1 and v_1 are defined by

$$m_{\text{ncl}}^{(1)} = \frac{R}{2} \sum_{k=1}^K \frac{n_k}{\sigma_k^2} \tag{8.74}$$

$$v_{\text{ncl}}^{(1)} = \sum_{k=1}^K \left[\frac{n_k + 1 + \frac{n_k R}{2\sigma_k^2}}{\sigma_k^2 + \frac{a^2}{n_k}} - \frac{R^2 n_k}{4\sigma_k^4} \right]. \tag{8.75}$$

8.7.4 ML Estimation of Parameters in Truncated Gaussian Data

This section proposes a ML estimation of parameters in truncated Gaussian data. For the sake of simplification, we neglect the first and the third terms in (8.6). Suppose a random variable X follows the truncated Gaussian distribution taking values on the interval $(0, B)$. The data that takes value at the boundary 0 and B is excluded. Such truncated Gaussian distribution is denoted by $\mathcal{N}_{(0,B)}(\mu, \sigma^2)$ with $\sigma^2 = a\mu + b$. This distribution also takes advantage of the heteroscedastic relation between the mathematical expectation and variance. The pdf of X is simplified from (8.6) as

$$f_X(u) = \frac{1}{\Phi(t_B) - \Phi(t_0)} \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left(-\frac{(u - \mu)^2}{2\sigma^2} \right) \tag{8.76}$$

where

$$t_0 = \frac{-\mu}{\sigma} \quad \text{and} \quad t_B = \frac{B - \mu}{\sigma} \quad (8.77)$$

characterize the truncation. It must be noted that μ now is not the expected value of X . According to [160], the mathematical expectation and variance of the random variable X are given by

$$\mathbb{E}[X] = \mu + (T_0 - T_B)\sigma \quad (8.78)$$

$$\text{Var}[X] = \sigma^2 \left[1 + t_0 T_0 - t_B T_B - (T_0 - T_B)^2 \right] \quad (8.79)$$

where

$$T_0 = \frac{\phi(t_0)}{\Phi(t_B) - \Phi(t_0)} \quad \text{and} \quad T_B = \frac{\phi(t_B)}{\Phi(t_B) - \Phi(t_0)}. \quad (8.80)$$

Suppose a vector of n independent realizations $\mathbf{x} = (x_i)_{i=1}^n$ drawing from the truncated Gaussian distribution $\mathcal{N}_{(0,B)}(\mu, \sigma^2)$. By definition, the ML estimate $\hat{\mu}$ is defined as the solution which maximizes the log likelihood function of the samples $\mathbf{x}$

$$\hat{\mu} = \arg \max_{\mu} \mathcal{L}(\mathbf{x}; \mu) = \arg \max_{\mu} \sum_{i=1}^n \mathcal{L}(x_i; \mu) \quad (8.81)$$

where

$$\mathcal{L}(x_i; \mu) = -\log \left[\Phi(t_B) - \Phi(t_0) \right] - \log(\sqrt{2\pi\sigma^2}) - \frac{(x_i - \mu)^2}{2\sigma^2}. \quad (8.82)$$

Unfortunately, the ML estimate $\hat{\mu}$ can not be analytically given. In this section, the problem (8.81) is numerically solved using the Nelder-Mead optimization method [157] and taking the empirical mean $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$ as initial solution. According to [189], the ML estimate $\hat{\mu}$ follows the asymptotic Gaussian distribution

$$\hat{\mu} \sim \mathcal{N}\left(\mu, \frac{1}{n\mathbf{F}(\mu)}\right), \quad (8.83)$$

where $\mathbf{F}(\mu)$ represents the Fisher information

$$\begin{aligned} \mathbf{F}(\mu) &= -\mathbb{E} \left[\frac{\partial^2 \mathcal{L}(x_i; \mu)}{\partial \mu^2} \right] \\ &= \frac{\partial^2 t_B}{\partial \mu^2} T_B + \frac{\partial t_B}{\partial \mu} \frac{\partial T_B}{\partial \mu} - \frac{\partial^2 t_0}{\partial \mu^2} T_0 - \frac{\partial t_0}{\partial \mu} \frac{\partial T_0}{\partial \mu} - \frac{a^2 \sigma^2 - 2b^2}{2\sigma^6} \\ &\quad + \frac{a(a^2 \mathbb{E}[x_i^2] + 2b \mathbb{E}[x_i])}{\sigma^6}. \end{aligned} \quad (8.84)$$

8.7.5 Statistical distribution of the GLR $\hat{\Lambda}_{\text{cl}}(\underline{\mathbf{Z}})$

The proof follows similar calculations as in Section 8.7.2 and 8.7.3. For simplicity, we will not recall intermediate calculations.

Under hypothesis $\mathcal{H}_0$, from the total probability theorem, it can be shown that $\mathbb{E}_{\mathcal{H}_0}[\hat{\Lambda}_{\text{cl}}(\underline{z}_{k,i})] = 0$. Besides, it follows from (8.32) that

$$\hat{\sigma}_k^2 \sim \mathcal{N}\left(\sigma_k^2, \frac{a^2}{n_k \mathbf{F}(\mu_k)}\right). \quad (8.85)$$

The variance of the GLR $\hat{\Lambda}_{\text{cl}}(\underline{z}_{k,i})$ under hypothesis $\mathcal{H}_0$ can be calculated as

$$\begin{aligned} \mathbb{V}\text{ar}_{\mathcal{H}_0}[\hat{\Lambda}_{\text{cl}}(\underline{z}_{k,i})] &= \frac{\mathbb{E}_{\mathcal{H}_0}[(\underline{z}_{k,i} - \hat{\mu}_k)^2]}{\mathbb{E}_{\mathcal{H}_0}[\hat{\sigma}_k^4]} \\ &= \frac{\mathbb{E}_{\mathcal{H}_0}[(\underline{z}_{k,i} - \mu_k)^2] + \mathbb{V}\text{ar}_{\mathcal{H}_0}[\hat{\mu}_k]}{\mathbb{E}_{\mathcal{H}_0}^2[\hat{\sigma}_k^2] + \mathbb{V}\text{ar}_{\mathcal{H}_0}[\hat{\sigma}_k^2]} \\ &= \frac{\sigma_k^2 \left(1 + t_{0,k}T_{0,k} - t_{B,k}T_{B,k}\right) + \frac{1}{n_k \mathbf{F}(\mu_k)}}{\sigma_k^4 + \frac{a^2}{n_k \mathbf{F}(\mu_k)}}, \end{aligned} \quad (8.86)$$

where the terms $t_{0,k}$, $t_{B,k}$, $T_{0,k}$, $T_{B,k}$ take into account the impact of truncation in each segment S_k

$$t_{0,k} = -\frac{\mu_k}{\sigma_k} \quad \text{and} \quad t_{B,k} = \frac{B - \mu_k}{\sigma_k} \quad (8.87)$$

$$T_{0,k} = \frac{\phi(t_{0,k})}{\Phi(t_{B,k}) - \Phi(t_{0,k})} \quad \text{and} \quad T_{B,k} = \frac{\phi(t_{B,k})}{\Phi(t_{B,k}) - \Phi(t_{0,k})}. \quad (8.88)$$

Under hypothesis $\mathcal{H}_1$, the mathematical expectation of the GLR $\hat{\Lambda}_{\text{cl}}(\underline{z}_{k,i})$ is given as

$$\mathbb{E}_{\mathcal{H}_1}[\hat{\Lambda}_{\text{cl}}(\underline{z}_{k,i})] = \frac{R}{2} \mathbb{E}_{\mathcal{H}_0}\left[\frac{1}{\hat{\sigma}_k^2}\right] = \frac{R}{2\sigma_k^2}. \quad (8.89)$$

Besides, the variance of the GLR $\hat{\Lambda}_{\text{cl}}(\underline{z}_{k,i})$ under hypothesis $\mathcal{H}_1$ is given as

$$\begin{aligned} \mathbb{V}\text{ar}_{\mathcal{H}_1}[\hat{\Lambda}_{\text{cl}}(\underline{z}_{k,i})] &= \mathbb{E}_{\mathcal{H}_1}[\hat{\Lambda}_{\text{cl}}^2(\underline{z}_{k,i})] - \mathbb{E}_{\mathcal{H}_1}^2[\hat{\Lambda}_{\text{cl}}(\underline{z}_{k,i})] \\ &= \mathbb{E}_{\mathcal{H}_0}\left[\frac{(\underline{z}_{k,i} - \hat{\mu}_k)^2}{\hat{\sigma}_k^4}\right] + \frac{R}{2} \mathbb{E}_{\mathcal{H}_0}\left[\frac{1}{\hat{\sigma}_k^4}\right] - \frac{R^2}{4\sigma_k^4} \\ &= \frac{\sigma_k^2 \left(1 + t_{0,k}T_{0,k} - t_{B,k}T_{B,k}\right) + \frac{1}{n_k \mathbf{F}(\mu_k)} + \frac{R}{2}}{\sigma_k^4 + \frac{a^2}{n_k \mathbf{F}(\mu_k)}} - \frac{R^2}{4\sigma_k^4}. \end{aligned} \quad (8.90)$$

Therefore, in virtue of Lindeberg CLT, the statistical distribution of the GLR $\hat{\Lambda}_{\text{cl}}(\underline{\mathbf{Z}})$ under hypothesis $\mathcal{H}_j$ is given by

$$\hat{\Lambda}_{\text{cl}}(\underline{\mathbf{Z}}) \xrightarrow{d} \mathcal{N}(m_{\text{cl}}^{(j)}, v_{\text{cl}}^{(j)}), \quad (8.91)$$

with

$$m_{\text{cl}}^{(0)} = 0 \quad (8.92)$$

$$v_{\text{cl}}^{(0)} = \sum_{k=1}^K \frac{n_k \sigma_k^2 \left(1 + t_{0,k} T_{0,k} - t_{B,k} T_{B,k} \right) + \frac{1}{\mathbf{F}(\mu_k)}}{\sigma_k^4 + \frac{a^2}{n_k \mathbf{F}(\mu_k)}} \quad (8.93)$$

$$m_{\text{cl}}^{(1)} = \frac{R}{2} \sum_{k=1}^K \frac{n_k}{\sigma_k^2} \quad (8.94)$$

$$v_{\text{cl}}^{(1)} = \sum_{k=1}^K \left[\frac{n_k \sigma_k^2 \left(1 + t_{0,k} T_{0,k} - t_{B,k} T_{B,k} \right) + \frac{1}{\mathbf{F}(\mu_k)} + \frac{n_k R}{2}}{\sigma_k^4 + \frac{a^2}{n_k \mathbf{F}(\mu_k)}} - \frac{n_k R^2}{4 \sigma_k^4} \right]. \quad (8.95)$$

Steganalysis of Jsteg Algorithm Based on a Novel Statistical Model of Quantized DCT Coefficients

Contents

9.1	Introduction	163
9.2	Optimal Detector for Steganalysis of Jsteg Algorithm	164
9.2.1	Hypothesis Testing Formulation	164
9.2.2	LRT for Two Simple Hypotheses	165
9.3	Quantitative Steganalysis of Jsteg Algorithm	168
9.3.1	ML Estimation of Embedding Rate	168
9.3.2	Revisiting WS estimator	169
9.4	Numerical Experiments	169
9.4.1	Detection Performance of the proposed LRT	169
9.4.2	Accuracy of the Proposed Estimator	172
9.5	Conclusion	173

9.1 Introduction

The goal of this chapter is to design a statistical test for steganalysis of Jsteg algorithm within hypothesis testing framework. As discussed in Section 2.5.1, Jsteg algorithm is a variant of LSB replacement mechanism. Instead of embedding hidden bits into pixel's LSB, Jsteg algorithm embeds them into the LSB of quantized DCT coefficients. Like in steganalysis of LSB replacement, several detectors have been proposed for the steganalysis of Jsteg algorithm, including structural detectors [122–126], WS detectors [127, 128], statistical detectors [129] and blind detectors. Unlike statistical detectors proposed for steganalysis of LSB replacement, hypothesis testing theory is of limited exploitation in the steganalysis of Jsteg algorithm. A first step proposed in [129] is based on the Laplacian model to characterize DCT coefficients. However, due to the fact that the Laplacian model is not relevant (see

Section 4.4), a considerable loss of optimality is revealed. Other detectors perform efficiently but their performance can not analytically established and they can not warrant a prescribed false alarm probability.

The main contributions are the following:

- The approach is based on the highly accurate model of DCT coefficients proposed in Section 4.4 that can detect any small change in statistical properties of cover image due to message embedding.
- By formulating the hidden data detection as a hypothesis testing problem, this chapter studies an optimal detector given by the LRT assuming that all model parameters are known in practice. The statistical performance of the LRT is analytically established. The test allows us to warrant a prescribed false alarm probability and maximize the correct detection probability.
- Based on the proposed model of quantized DCT coefficients, a ML estimator of embedding rate is also derived for quantitative steganalysis. Besides, the WS detector [128] is revisited and a more efficient version is proposed relying on a modified filter.

The chapter is organized as follows. Section 9.2 states the hidden data detection problem in the framework of hypothesis testing theory and studies the LRT assuming that all model parameters are known. Section 9.3 derives the ML estimator of embedding rate based on the proposed model of quantized DCT coefficients and revisits the WS detector. Section 9.4 presents numerical results on simulated database and real image database and the comparison with prior-art detectors. Numerical results on simulated database and a large image database not only show the accuracy of the proposed model of quantized DCT coefficients but also emphasize the relevance of the proposed detector and ML estimator. Finally, Section 9.5 concludes the chapter. The work in this chapter has been published in [154]. Furthermore, the same methodology has been proposed for steganalysis of Outguess algorithm [190].

9.2 Optimal Detector for Steganalysis of Jsteg Algorithm

9.2.1 Hypothesis Testing Formulation

Let $\mathbf{C}$ be a matrix representing the cover image that is composed of 64 vectors of quantized DCT coefficients $\mathbf{C}_k = (C_{k,1}, \dots, C_{k,N_{\text{blk}}})$, $k \in \{1, \dots, 64\}$, where N_{blk} is the number of blocks. The pmf of the quantized DCT coefficient $\mathbf{C}_k$ is denoted by $P_{\boldsymbol{\theta}_k, \Delta_k}$ characterized by the parameter vector $\boldsymbol{\theta}_k$ and the corresponding quantization step Δ_k . For instance, in our proposed model, $\boldsymbol{\theta}_k = (\alpha_k, \beta_k)$. In the Jsteg algorithm, each hidden bit, that is either 0 or 1, is statistically independent of the cover coefficients. Moreover, the probability of insertion is equal for every coefficient. The Jsteg does not embed in the coefficients that are equal to 0 and 1 since artifacts caused

by such insertion can be easily detected. For the same reason, the DC coefficient is not used for insertion as well. The number of usable coefficients in each vector $\mathbf{C}_k$, $k \in \{2, \dots, 64\}$, is represented by a random variable $n_k \leq N_{\text{blk}}$. The number of usable coefficients n_k depends on the image content and the quantization matrix (hence the quality factor). Without loss of generality, the n_k first components of the vector are usable and the remaining $n - n_k$ components are excluded.

The secret message is embedded at rate R in the cover image $\mathbf{C}$ to create a stego-image $\mathbf{S} = (\mathbf{S}_1, \dots, \mathbf{S}_{64})$. As described in Section 8.3.1, the pmf of $\mathbf{S}_k$, say $Q_{R, \boldsymbol{\theta}_k, \Delta_k}$, can be given as

$$Q_{R, \boldsymbol{\theta}_k, \Delta_k}(l) = \begin{cases} \frac{R}{2} P_{\boldsymbol{\theta}_k, \Delta_k}(\bar{l}) + \left(1 - \frac{R}{2}\right) P_{\boldsymbol{\theta}_k, \Delta_k}(l) & \forall l \in \mathbb{Z} \setminus \{0, 1\} \\ Q_{R, \boldsymbol{\theta}_k, \Delta_k}(l) = P_{\boldsymbol{\theta}_k, \Delta_k}(l) & l \in \{0, 1\}, \end{cases} \quad (9.1)$$

where $\bar{l}$ indicates the integer l with LSB flipped $\bar{l} = l + (-1)^l$. Since coefficients with value 0 and 1 are not used for insertion, the pmf $P_{\boldsymbol{\theta}_k, \Delta_k}$ does not change for $l \in \{0, 1\}$ after insertion.

When inspecting the image of DCT coefficients $\mathbf{D}$ that is either a cover image $\{\mathbf{D} = \mathbf{C}\}$ or a stego-one $\{\mathbf{D} = \mathbf{S}\}$, the goal of the test is to decide between two hypotheses defined as

$$\begin{cases} \mathcal{H}_0 = \left\{ D_{k,i} \sim P_{\boldsymbol{\theta}_k, \Delta_k}, \forall k \in \{2, \dots, 64\}, \forall i \in \{1, \dots, N_{\text{blk}}\} \right\} \\ \mathcal{H}_1 = \left\{ D_{k,i} \sim Q_{R, \boldsymbol{\theta}_k, \Delta_k}, \forall k \in \{2, \dots, 64\}, \forall i \in \{1, \dots, N_{\text{blk}}\} \right\}. \end{cases} \quad (9.2)$$

The DC coefficients $\mathbf{D}_1$ are excluded in the problem (9.2) because they are not used for secret message embedding. Let

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

be the class of tests whose false alarm probability is upper-bounded by the rate α_0 where $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_{64})$ is the parameter vector. Among all the tests in the class $\mathcal{K}_{\alpha_0}$, it is aimed at finding a test δ which maximizes the power function $\beta(\delta)$.

9.2.2 LRT for Two Simple Hypotheses

Assuming that the model parameters are known, in virtue of the Neyman-Pearson lemma, the MP test over the class $\mathcal{K}_{\alpha_0}$ is the LRT given by the following decision rule

$$\delta_{\text{jst}}^*(\mathbf{D}) = \begin{cases} \mathcal{H}_0 & \text{if } \Lambda_{\text{jst}}(\mathbf{D}) = \sum_{k=2}^{64} \Lambda_{\text{jst}}(\mathbf{D}_k) < \tau_{\text{jst}}^* \\ \mathcal{H}_1 & \text{if } \Lambda_{\text{jst}}(\mathbf{D}) = \sum_{k=2}^{64} \Lambda_{\text{jst}}(\mathbf{D}_k) \geq \tau_{\text{jst}}^* \end{cases} \quad (9.3)$$

where the LR $\Lambda_{\text{jst}}(\mathbf{D}_k) = \sum_{i=1}^{n_k} \Lambda_{\text{jst}}(D_{k,i})$ and $\Lambda_{\text{jst}}(D_{k,i})$ is defined by

$$\Lambda_{\text{jst}}(D_{k,i}) = \log \frac{Q_{R,\theta_k,\Delta_k}(D_{k,i})}{P_{\theta_k,\Delta_k}(D_{k,i})} = \log \left[1 - \frac{R}{2} + \frac{R}{2} \frac{P_{\theta_k,\Delta_k}(D_{k,i} + (-1)^{D_{k,i}})}{P_{\theta_k,\Delta_k}(D_{k,i})} \right]. \quad (9.4)$$

Here, it is assumed that the DCT coefficients are statistically independent. Accordingly, the LR $\Lambda_{\text{jst}}(D_{k,i})$ can be interpreted as a function of the coefficient $D_{k,i}$. Suppose the function $h(l; \theta_k, \Delta_k)$ defined as

$$h(l; \theta_k, \Delta_k) \triangleq \log \left[1 - \frac{R}{2} + \frac{R}{2} \frac{P_{\theta_k,\Delta_k}(l + (-1)^l)}{P_{\theta_k,\Delta_k}(l)} \right], \quad (9.5)$$

such that $\Lambda_{\text{jst}}(D_{k,i}) = h(D_{k,i}; \theta_k, \Delta_k)$.

Because the proposed model of quantized DCT coefficients is complicated, it seems impossible to simplify the expression of the LR $\Lambda_{\text{jst}}(D_{k,i})$ and establish analytically its expectation and variance. Nevertheless, it is proposed to take advantage of the fact that the quantized DCT coefficient $D_{k,i}$ only takes integer values in a finite set, say $\mathbb{D}$. The expectation $m_{\text{jst}}^{(k,j)}$ and variance $v_{\text{jst}}^{(k,j)}$ of the LR $\Lambda_{\text{jst}}(D_{k,i})$ under hypothesis $\mathcal{H}_j$, $j \in \{0, 1\}$, can be numerically given by

$$m_{\text{jst}}^{(k,0)} = \mathbb{E}_{\mathcal{H}_0} [\Lambda_{\text{jst}}(D_{k,i})] = \sum_{l \in \mathbb{D}} h(l; \theta_k, \Delta_k) P_{\theta_k,\Delta_k}(l) \quad (9.6)$$

$$v_{\text{jst}}^{(k,0)} = \text{Var}_{\mathcal{H}_0} [\Lambda_{\text{jst}}(D_{k,i})] = \sum_{l \in \mathbb{D}} (h(l; \theta_k, \Delta_k) - m_{\text{jst}}^{(k,0)})^2 P_{\theta_k,\Delta_k}(l) \quad (9.7)$$

$$m_{\text{jst}}^{(k,1)} = \mathbb{E}_{\mathcal{H}_1} [\Lambda_{\text{jst}}(D_{k,i})] = \sum_{l \in \mathbb{D}} h(l; \theta_k, \Delta_k) Q_{R,\theta_k,\Delta_k}(l) \quad (9.8)$$

$$v_{\text{jst}}^{(k,1)} = \text{Var}_{\mathcal{H}_1} [\Lambda_{\text{jst}}(D_{k,i})] = \sum_{l \in \mathbb{D}} (h(l; \theta_k, \Delta_k) - m_{\text{jst}}^{(k,1)})^2 Q_{R,\theta_k,\Delta_k}(l). \quad (9.9)$$

Theorem 9.1. *Under hypothesis $\mathcal{H}_j$, the LR $\Lambda_{\text{jst}}(\mathbf{D})$ follows the Gaussian distribution with expectation $m_{\text{jst}}^{(j)}$ and variance $v_{\text{jst}}^{(j)}$ as*

$$\Lambda_{\text{jst}}(\mathbf{D}) \xrightarrow{d} \mathcal{N}(m_{\text{jst}}^{(j)}, v_{\text{jst}}^{(j)}), \quad (9.10)$$

with

$$m_{\text{jst}}^{(j)} = \sum_{k=2}^{64} N_{\text{blk}} p_k^* m_{\text{jst}}^{(k,j)} \quad (9.11)$$

$$v_{\text{jst}}^{(j)} = \sum_{k=2}^{64} \left[N_{\text{blk}} p_k^* v_{\text{jst}}^{(k,j)} + N_{\text{blk}} p_k^* (1 - p_k^*) (m_{\text{jst}}^{(k,j)})^2 \right], \quad (9.12)$$

where $p_k^* = 1 - P_{\theta_k,\Delta_k}(0) - P_{\theta_k,\Delta_k}(1)$ is the probability that a coefficient differs from 0 and 1.

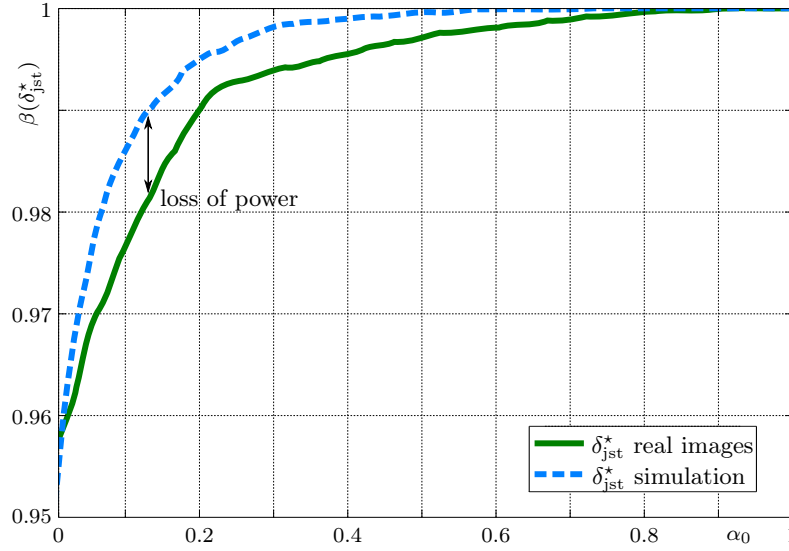


Figure 9.1: Detection performance of the test δ_{jst}^* based on the proposed model with embedding rate $R = 0.05$ on the simulated images and real images.

Proof. It can be noted that the LR $\Lambda_{\text{jst}}(\mathbf{D}_k)$ is a random sum of n_k elements since n_k , which corresponds to the number of coefficients that differ from 0 and 1, is itself a random variable. The number n_k follows the binomial distribution $\mathcal{B}(N_{\text{blk}}, p_k^*)$. The probability p_k^* remains identical under every hypothesis since the insertion is not performed in coefficients that are equal to 0 and 1. It follows from the Wald's identity [191] that the mathematical expectation and variance of the LR $\Lambda_{\text{jst}}(\mathbf{D}_k)$ can be given by

$$\begin{aligned}\mathbb{E}_{\mathcal{H}_j}[\Lambda_{\text{jst}}(\mathbf{D}_k)] &= \mathbb{E}_{\mathcal{H}_j}[n_k] \mathbb{E}_{\mathcal{H}_j}[\Lambda_{\text{jst}}(D_{k,i})] = N_{\text{blk}} p_k^* m_{\text{jst}}^{(k,j)} \\ \text{Var}_{\mathcal{H}_j}[\Lambda_{\text{jst}}(\mathbf{D}_k)] &= \mathbb{E}_{\mathcal{H}_j}[n_k] \text{Var}_{\mathcal{H}_j}[\Lambda_{\text{jst}}(D_{k,i})] + \mathbb{E}_{\mathcal{H}_j}^2[\Lambda_{\text{jst}}(D_{k,i})] \text{Var}_{\mathcal{H}_j}[n_k] \\ &= N_{\text{blk}} p_k^* v_{\text{jst}}^{(k,j)} + N_{\text{blk}} p_k^* (1 - p_k^*) (m_{\text{jst}}^{(k,j)})^2.\end{aligned}$$

In virtue of the Lindeberg CLT, the LR $\Lambda_{\text{jst}}(\mathbf{D}_k)$ follows the Gaussian distribution as

$$\Lambda_{\text{jst}}(\mathbf{D}_k) \xrightarrow{d} \mathcal{N}\left(N_{\text{blk}} p_k^* m_{\text{jst}}^{(k,j)}, N_{\text{blk}} p_k^* v_{\text{jst}}^{(k,j)} + N_{\text{blk}} p_k^* (1 - p_k^*) (m_{\text{jst}}^{(k,j)})^2\right).$$

Finally, from the linearity property of the Gaussian distribution, the proof follows immediately. $\square$

Finally, the LRT δ_{jst}^* based on the normalized LR $\Lambda_{\text{jst}}^*(\mathbf{D}) = \frac{\Lambda_{\text{jst}}(\mathbf{D}) - m_{\text{jst}}^{(0)}}{\sqrt{v_{\text{jst}}^{(0)}}}$ is given as follows

$$\delta_{\text{jst}}^*(\mathbf{D}) = \begin{cases} \mathcal{H}_0 & \text{if } \Lambda_{\text{jst}}^*(\mathbf{D}) < \tau_{\text{jst}}^* \\ \mathcal{H}_1 & \text{if } \Lambda_{\text{jst}}^*(\mathbf{D}) \geq \tau_{\text{jst}}^* \end{cases} \quad (9.13)$$

where the decision threshold τ_{jst}^* is the solution of the equation $\mathbb{P}_0[\Lambda_{\text{jst}}^*(\mathbf{D}) \geq \tau_{\text{jst}}^*] = \alpha_0$. The decision threshold τ_{jst}^* and the power function $\beta(\delta_{\text{jst}}^*)$ of the LRT δ_{jst}^* are given in the following theorem.

Theorem 9.2. *In an ideal context where the embedding rate R and the parameters θ_k are known, the decision threshold and the power function of the LRT δ_{jst}^* are given by*

$$\tau_{\text{jst}}^* = \Phi^{-1}(1 - \alpha_0) \quad (9.14)$$

$$\beta(\delta_{\text{jst}}^*) = 1 - \Phi\left(\frac{m_{\text{jst}}^{(0)} - m_{\text{jst}}^{(1)} + \tau_{\text{jst}}^* \sqrt{v_{\text{jst}}^{(0)}}}{\sqrt{v_{\text{jst}}^{(1)}}}\right). \quad (9.15)$$

Proof. From Theorem 9.1, based on the definitions of the decision threshold and the power of the test, the proof follows immediately. $\square$

The statistical performance of the LRT δ_{jst}^* is analytically established. The decision threshold is set independently of the image content. It can be noted that the scenario studied by the test δ_{jst}^* may not be realistic because the parameters θ_k can not be known in advance in a real image. An usual approach in practice is to replace the parameters θ_k by ML estimates $\hat{\theta}_k^{\text{ML}}$. Thus, the detection performance of the test δ_{jst}^* depends on the accuracy of the proposed model of DCT coefficients and ML estimates $\hat{\theta}_k^{\text{ML}}$. The fact of replacing the unknown parameters θ_k by ML estimates $\hat{\theta}_k^{\text{ML}}$ shows that the test δ_{jst}^* seems to coincide with the GLRT. However, the test δ_{jst}^* does not consider the variability of ML estimates $\hat{\theta}_k^{\text{ML}}$. Because the ML estimates $\hat{\theta}_k^{\text{ML}}$ are numerically derived by the optimization method (see Section 4.4.3), their statistical properties can not be easily studied. This leads to a difficulty of establishing analytically the statistical performance of the GLRT.

9.3 Quantitative Steganalysis of Jsteg Algorithm

9.3.1 ML Estimation of Embedding Rate

Here the estimation of embedding rate is formulated into the ML framework using the proposed model of DCT coefficients. As discussed in [99], ML estimators are more statistically rigorous, but their performance is weak due to lack of accurate models for cover images. An extension for ML framework is derived in [126] that is based on the concept of a precover introduced in [99] and the Generalized Cauchy distribution for unquantized DCT coefficients [87]. On the contrary, the proposed model of quantized DCT coefficients is exploited to estimate the embedding rate R . Given an inspected image $\mathbf{D}$, the ML estimator $\hat{R}$ is given by

$$\hat{R} = \arg \max_{0 \leq R \leq 1} \sum_{k=2}^{64} \sum_{i=1}^{n_k} \log Q_{R, \theta_k, \Delta_k}(D_{k,i}) \quad (9.16)$$

where $Q_{R,\theta_k,\Delta_k}(D_{k,i})$ is given in (9.1). Again, the DC coefficients and the coefficients that are equal to 0 and 1 are excluded. The maximization problem (9.16) is resolved numerically by the Nelder-Mead method [157].

9.3.2 Revisiting WS estimator

Beside the ML estimator (9.16), it is also proposed to revisit the WS estimator [128]. Technically, the WS estimator [128] can be given in the following manner

$$\hat{R} \propto \sum_i w_i (D_i - \bar{D}_i) D_i, \quad (9.17)$$

where the weight w_i can be defined as

$$w_i \propto \frac{\mathbf{1}_{D_i \notin \{0,1\}}}{1 + \mathcal{V}(\mathbf{D})_i}. \quad (9.18)$$

Here $\mathbf{1}_E$ denotes the indicator function and $\mathcal{V}(\mathbf{D})_i$ denotes the local empirical variance with respect to the coefficient D_i . To calculate this local empirical variance, the authors in [128] employ the DCT coefficients at the same frequency in four adjacent blocks (excluding the center block) as

$$\begin{aligned} \mathcal{V}(\mathbf{D})_{i,j} = & \frac{1}{4} (D^2(i-8, j) + D^2(i+8, j) + D^2(i, j-8) + D^2(i, j+8)) \\ & - \left[\frac{1}{4} (D(i-8, j) + D(i+8, j) + D(i, j-8) + D(i, j+8)) \right]^2. \end{aligned} \quad (9.19)$$

As discussed in [95, 103, 104], an appropriate choice of linear filter and weight vector can remarkably improve the performance of WS estimator. Here rather than employing the DCT coefficients at the same frequency in four adjacent blocks, we propose to employ two coefficients at the same frequency in two adjacent blocks and two other coefficients at the adjacent frequencies in the same block

$$\begin{aligned} \mathcal{V}(\mathbf{D})_{i,j} = & \frac{1}{4} (D^2(i-8, j) + D^2(i+8, j) + D^2(i, j-1) + D^2(i, j+1)) \\ & - \left[\frac{1}{4} (D(i-8, j) + D(i+8, j) + D(i, j-1) + D(i, j+1)) \right]^2. \end{aligned} \quad (9.20)$$

9.4 Numerical Experiments

9.4.1 Detection Performance of the proposed LRT

To illustrate the detection performance of the test δ_{Jst}^* based on the proposed model, the reference BOSSBase database [171] containing 10000 grayscale images of size 512×512 in PGM format is chosen to conduct experiments. The embedding rate R is set at 0.05 for the Jsteg algorithm. The hidden bits are drawn from a binomial distribution $\mathcal{B}(1, 1/2)$, i.e. each hidden bit can be 0 or 1 with the same probability. The coefficients in which hidden bits are embedded are randomly chosen. All the PGM images are converted to JPEG format using *imagemagick* with quality factor

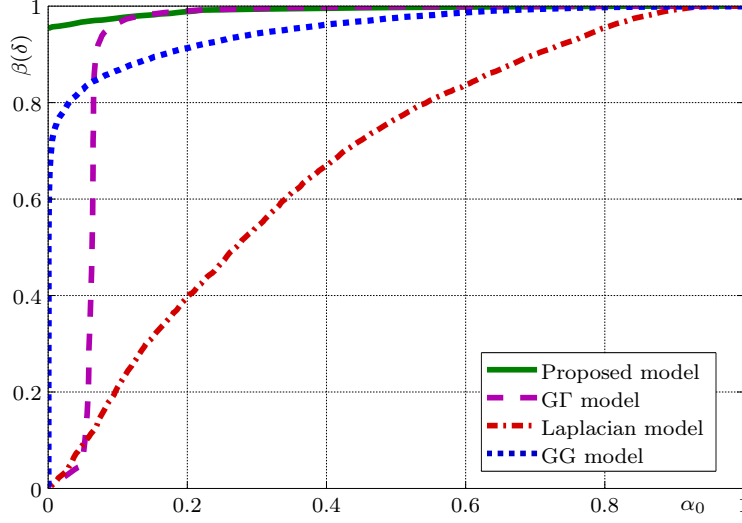


Figure 9.2: Detection performance of the test δ_{jst}^* based on the quantized Laplacian, quantized GG, quantized GF, and proposed model on the BOSSBase with embedding rate $R = 0.05$.

of 70. The 63 vectors of AC coefficients are extracted from every image. The quantization matrix is given in the header of each image file. The parameters $\theta_k = (\alpha_k, \beta_k)$ are estimated on each frequency of each image based on the ML approach. The estimates $\hat{\theta}_k$ are used to generate each vector D_k such that the simulated data involves the same model parameters as the ones estimated from real images. Therefore, the simulated data perfectly follows the proposed model and there is no correlation between simulated DCT coefficients. The test δ_{jst}^* is performed on 10000 simulated images and 10000 real JPEG images to evaluate the loss of power of the proposed test δ_{jst}^* in the practical context. The power functions on simulated images and real images are shown in Figure 9.1. A small loss of power is obviously revealed between the two power functions, which may be caused by the following reasons. Firstly, the accuracy of the proposed model may be affected by simplistic assumptions (e.g. the pixels are identically distributed within a 8×8 block). Secondly, DCT coefficients exhibit correlation in a real JPEG image. Finally, it can be caused by the estimation of model parameters, which are numerically provided by the optimization method.

The proposed test δ_{jst}^* can be used with any cover image model. The more accurate the cover image model is, the better the detection performance is. The Figure 9.2 illustrates the detection performance of the test δ_{jst}^* based on quantized Laplacian, quantized GG, quantized GF and proposed model on 10000 real JPEG images. The test δ_{jst}^* based on the proposed model shows a higher correct detection probability than the quantized Laplacian-based, quantized GG-based and quantized GF-based. In this experiment, the proposed detector can provide an important power with zero false alarm. Moreover, it appears that the quantized GF model fails on a subset of about 1000 images, which leads to its very low power function for $0 \leq \alpha_0 \leq 0.1$ while the proposed model still shows a high detection performance. The

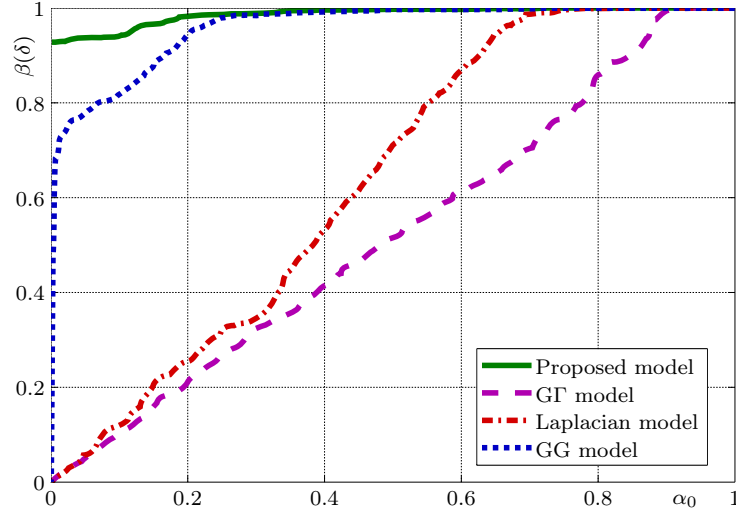


Figure 9.3: Detection performance of the test δ_{jst}^* based on the quantized Laplacian, quantized GG, quantized GF, and proposed model on the subset of 1000 images from the BOSSBase with embedding rate $R = 0.05$.

detection performance on this subset is illustrated in Figure 9.3. These results also show that the proposed model is more robust and accurate for the DCT coefficients. The term "robust" means the accuracy of the model on a wide range of images. Besides, the test δ_{jst}^* based on the proposed model is nearly perfect for embedding rate $R = 0.1$, i.e. $\beta(\delta_{jst}^*) \cong 1$, for any false alarm probability α_0 .

Potentially, there are many detectors in the literature could be compared with the proposed test. The ZP detector [122] was known as the first quantitative attack on Jsteg. The well-known WS detector [128, Eq. (9)] is also included in the comparison because of its efficiency and low computational complexity. The recent quantitative structural detector ZMH-Sym [126] based on the ZMH framework and the exploitation of the natural symmetry of DCT coefficients in the cover image was shown as the best detector among histogram-based attacks. The quantized Laplacian-based test [129] is also performed because it is based on the same framework of hypothesis testing theory. On the contrary to the support vector regression-based detector [119] that needs an expensive training phase, all above detectors, including the proposed one, work solely on an image-by-image basis. The Figure 9.4 shows the comparison of the proposed test δ_{jst}^* with other detectors. We have also performed the ZMH-Cat structural detector [126] but do not report it in Figure 9.4 because its power function is considerably worse than the detector ZMH-Sym's one, see Figure 9.5. Obviously, the proposed test outperforms other detectors, whatever the false alarm probability. It should particularly be noted that for very low-false alarm rate the proposed test performs much better than the others, which is the most important in practice since the false-alarm probability must be set very low.

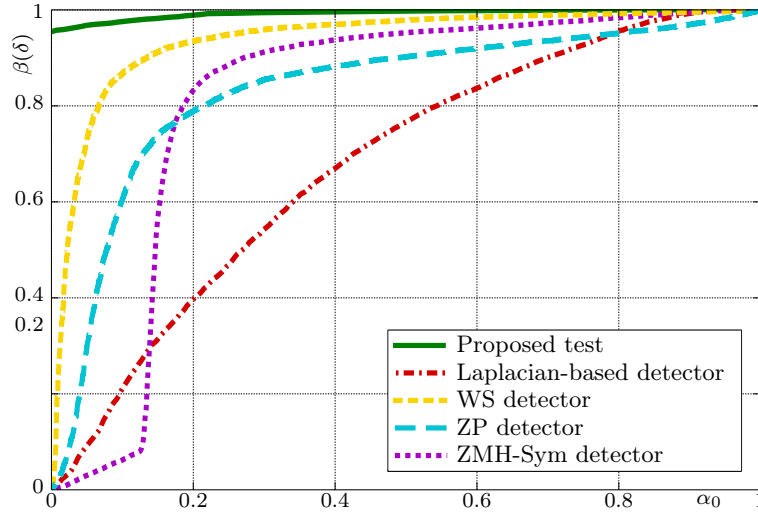


Figure 9.4: Comparison between the proposed test δ_{jst}^* , ZMH-Sym detector, ZP detector, WS detector and quantized Laplacian-based test.

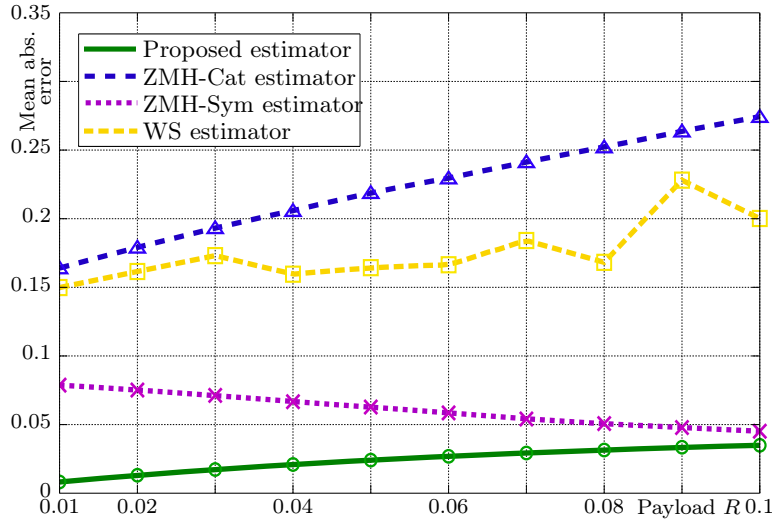


Figure 9.5: Mean absolute error for all estimators.

9.4.2 Accuracy of the Proposed Estimator

In terms of embedding rate estimation, the accuracy of all estimators is evaluated for embedding rate R ranging from 0.01 to 0.1 using the Mean Absolute Error (MAE) criteria. The Figure 9.5 shows the MAE of all estimators. The proposed ML estimator (9.16) outperforms other estimators. The ZMH-Sym estimator has a comparable accuracy to the proposed ML estimator but it has more outliers, which leads to the degradation of the ROC curve. The high detection performance of the test δ_{jst}^* (Figure 9.4) and high accuracy of embedding rate estimation (Figure 9.5) also emphasize the accuracy of the proposed model of quantized DCT coefficients.

The modified WS estimator is also performed on the BOSSBase. Figure 9.6 shows the MAE for the proposed ML estimator, standard WS estimator and modi-

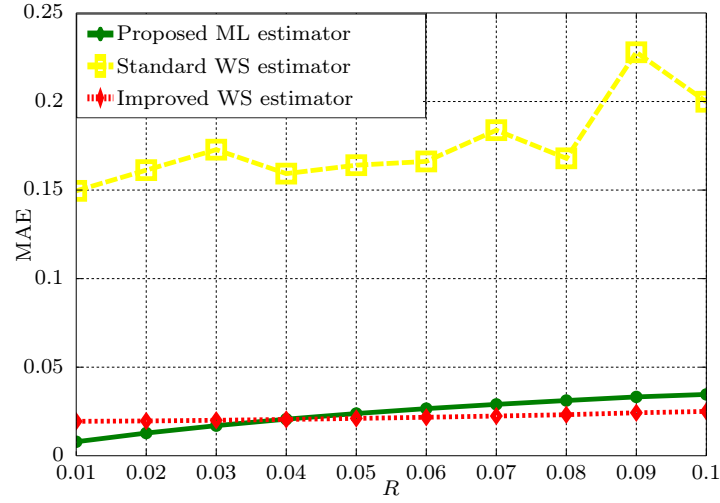


Figure 9.6: Mean absolute error for proposed ML estimator, standard WS estimator and improved WS estimator.

fied WS estimator. It can be noted that the accuracy of the modified WS estimator is considerably improved and very competitive with the proposed ML estimator. Moreover, its performance is quite stable for embedding rate ranging from 0.01 to 0.1. Meanwhile, the detection performance of the modified WS detector is only slightly improved compared with the standard WS detector, see Figure 9.7. It should be noted that the modified WS estimator employs the coefficients at the same frequency in two vertical adjacent blocks and the coefficients at two horizontal adjacent frequencies in the same block for the calculation of weights. Other directions also result in the same performance. The considerable performance improvement of this modified WS estimator can be justified due to the fact that the proposed calculation of weights indirectly takes into account intra-block and inter-block correlations of DCT coefficients.

9.5 Conclusion

This chapter designs the most powerful LRT for the steganalysis of Jsteg algorithm based on the very accurate model of quantized DCT coefficients. The proposed accurate model allows us to detect small changes in the cover image due to secret message embedding. The LRT can warrant a prescribed false alarm probability and maximizes the correct detection probability. An estimator of embedding rate is also designed within ML framework based on the proposed model of DCT coefficients. Moreover, the WS estimator is revisited using a different technique of calculation of weights such that the estimation accuracy of the modified WS estimator is considerably improved. Numerical results on simulated database and large image database highlight the relevance of the proposed approaches. Future researches could look into a test taking into account the image content as nuisance parameters. A first

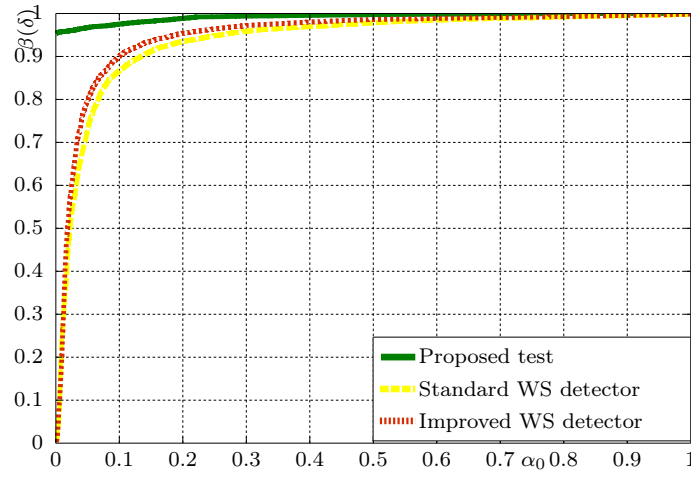


Figure 9.7: Comparison between the proposed test δ_{jst}^* , standard WS detector and improved WS detector.

step has been proposed in [192].

Conclusions and Perspectives

10.1 Conclusions

The evolution of digital imaging technology has raised a number of information security challenges. Digital images, which have nowadays become ubiquitous, can be easily manipulated because of a large availability of low-cost image editing tools. To restore the trustworthiness of digital images, the field of digital image forensics has been emerged.

The main goal of this thesis is to address the problem of image origin identification and hidden data detection in the field of digital image forensics. The state of the art shows that hypothesis testing theory is of limited exploitation by most of existing detectors compared with supervised classification. Besides, the empirical image models employed by those detectors do not take into account the image structure, nor the non-stationarity of acquisition noise, resulting in overall poor detection performance. Therefore, the proposed approach aims to design a statistical test within hypothesis testing framework based on an accurate parametric image model, which is also established in this thesis.

The thesis starts from establishing a statistical model that can accurately characterize a natural image acquired by a digital camera. This is accomplished by modeling the main steps of image processing pipeline and studying image statistics during these stages. Chapter 4 performs the study of image statistics in the spatial domain and DCT domain, from RAW format to JPEG format. By modeling various noise sources that corrupt a RAW image during acquisition stage, the heteroscedastic noise model is provided, in which the RAW pixel is normally distributed and its variance is linearly dependent on its expectation. This noise model is more relevant to characterize a RAW image than the conventional AWGN since the former take into account the contribution of Poisson noise in the acquisition stage. The study of image statistics is extended to TIFF image by starting from the heteroscedastic noise model and taking into account the non-linear effect of gamma correction, which leads to the so-called generalized noise model. The latter model gives the pixel's variance as a non-linear function of its expectation. It is shown that the generalized noise model is also relevant for JPEG images with moderate-to-high quality factors. One specificity is that the generalized noise model is proposed to combine with LLMMSE filter to design an efficient method for image denoising. The study of image statistics in the DCT domain is performed by modeling the distribution of DCT coefficients based on the doubly stochastic model. The proposed model of DCT coefficients outperforms other models, e.g. Laplacian, Generalized Gaussian

and Generalized Gamma. Another contribution is that algorithms for estimation of model parameters are designed in Chapter 4. In general, accurate image models and parameter estimation provided in this chapter build up a solid foundation for applications of hypothesis testing theory in the field of digital image forensics.

Based on the proposed statistical image models, the problems of camera model identification and hidden data detection are formally stated in the framework of hypothesis testing theory. The objective in this thesis is to design a test that can meet two criteria of optimality simultaneously: the warranting of a prescribed false alarm probability and maximization of the correct detection probability. In the theoretical context where all model parameters are known, or the hypotheses are simple, such optimal test is given by the LRT based on the Neyman-Pearson lemma. However, this scenario is unrealistic because the model parameters are unknown in practice. The GLRT is proposed to deal with the difficulty of unknown parameters, by replacing unknown parameters by their ML estimates in the LR. The GLRT almost coincides with the UMP test under large conditions. Numerical results show that the loss of power of the GLRTs designed in this thesis compared with the LRT is negligible. The main strength of the proposed GLRTs is that they can warrant a prescribed false alarm probability as well as achieve a high detection performance on a large database. In general, the proposed GLRTs outperform prior-art detectors, thanks to the relevant image model and the accurate estimation algorithm of model parameters. The list of statistical tests studied in this thesis is given in Table 10.1.

In this thesis, the problem of camera model identification is approached by many directions, depending on the image model. In particular, the heteroscedastic noise model is firstly exploited and its parameters (a, b) are considered as camera fingerprint for camera model identification from RAW images in Chapter 5. Two main limitations of this methodology are the unavailability of RAW images in practice and the dependence of parameters (a, b) on ISO sensitivity. Since most cameras export images in JPEG format, two more methodologies are proposed to complete these limitations by relying on the generalized noise model and the proposed model of DCT coefficients in Chapter 6 and 7, respectively. The parameters $(\tilde{a}, \tilde{b}, \gamma)$ of the generalized noise model and the parameters $(\tilde{c}, \tilde{d})$ characterizing the relation of α and β are proposed as camera fingerprint to identify camera models from JPEG images. These fingerprints are invariant to image content and camera settings, and robust to non-linear operations.

Although the problem of hidden data detection has been stated in the hypothesis testing framework, prior-art approaches are based on image models that can not totally characterize a natural image, e.g. the piece-wise polynomial model for steganalysis of LSB replacement. Therefore, a loss of power is revealed even on a simulated database. By contrast, the proposed approach relies on a more accurate image model that is already established in this thesis. The steganalysis of LSB replacement in a natural RAW image based on the heteroscedastic noise model is studied in Chapter 8. The proposed tests do not only show no loss of optimality, but also can warrant a prescribed false alarm probability, which the AUMP test based on the piece-wise polynomial model fails in practice. Furthermore, the specificity in

δ_{het}^*	LRT based on the heteroscedastic noise model when the image parameters μ_i and camera parameters (a_j, b_j) are known.
$\bar{\delta}_{\text{het}}^*$	GLRT based on the heteroscedastic noise model to deal with unknown image parameters μ_i when the camera parameters (a_j, b_j) are known.
$\tilde{\delta}_{\text{het}}^*$	GLRT based on the heteroscedastic noise model to deal with unknown parameters (μ_i, a_1, b_1) .
δ_{gen}^*	LRT based on the generalized noise model when the image parameters $\tilde{\mu}_i$ and camera parameters $(\tilde{a}_j, \tilde{b}_j, \gamma_j)$ are known.
$\bar{\delta}_{\text{gen}}^*$	GLRT based on the generalized noise model to deal with unknown image parameters $\tilde{\mu}_i$ when the camera parameters $(\tilde{a}_j, \tilde{b}_j, \gamma_j)$ are known.
$\tilde{\delta}_{\text{gen}}^*$	GLRT based on the generalized noise model to deal with unknown parameters $(\tilde{\mu}_i, \tilde{a}_1, \tilde{b}_1, \gamma_1)$.
δ_{dct}^*	LRT based on the DCT coefficient model when the parameters α_k and camera parameters $(\tilde{c}_{k,j}, \tilde{d}_{k,j})$ are known.
$\bar{\delta}_{\text{dct}}^*$	GLRT based on the DCT coefficient model to deal with unknown parameters α_k when the camera parameters $(\tilde{c}_{k,j}, \tilde{d}_{k,j})$ are known.
$\tilde{\delta}_{\text{dct}}^*$	GLRT based on the DCT coefficient model to deal with unknown parameters $(\alpha_k, \tilde{c}_{k,1}, \tilde{d}_{k,1})$.
$\hat{\delta}_{\text{ncl}}^*$	GLRT based on the heteroscedastic noise model for steganalysis of LSB replacement in a non-clipped RAW image.
$\hat{\delta}_{\text{cl}}^*$	GLRT based on the heteroscedastic noise model for steganalysis of LSB replacement in a clipped RAW image.
δ_{jst}^*	LRT based on the DCT coefficient model for steganalysis of Jsteg algorithm.

Table 10.1: List of proposed statistical tests.

Chapter 8 is that the clipping phenomenon is taken into account in the design of GLRT, which prior-art detectors can not tolerate this kind of degradation. Using clipped pixels can significantly improve the detection performance. The steganalysis of Jsteg algorithm based on the state-of-the-art model of DCT coefficients is studied in Chapter 9. Another contribution is for quantitative steganalysis, a ML estimator based on the proposed model of DCT coefficients is derived and the WS estimator is improved by employing the coefficients at the same frequency in two vertical adjacent blocks and the coefficients at two horizontal adjacent frequencies in the same block for the calculation of weights. Numerical results do not only show the better performance of the proposed detector compared with prior-art ones, but also emphasize the high accuracy of the proposed model of DCT coefficients.

Overall, this thesis shows significant contribution to the field of digital image forensics and statistical image modeling.

10.2 Perspectives

The thesis highlights several aspects of potential interest for future researches, according to three following axis: digital forensics, statistical image modeling, and statistical hypothesis testing theory.

10.2.1 Perspectives to Digital Forensics

10.2.1.1 Image Forgery Detection

A direct extension of the proposed methodologies is for image forgery detection problem. As shown by the state of the art, the image forgery detection problem has not been formally stated in the hypothesis testing framework yet. Consequently, an optimal detector that can warrant a prescribed false alarm probability has never been studied in the literature. In fact, based on the image models proposed in this thesis, the image forgery identification can be stated similarly as the steganalysis problem:

$$\begin{cases} \mathcal{H}_0 & : \mathbf{Z} \sim P_0 \\ \mathcal{H}_1 & : \mathbf{Z} \sim P_1, \end{cases} \quad (10.1)$$

where P_0 is the probability distribution of the original image and P_1 is the probability distribution of the falsified image due to a certain manipulation. It can be noted that any kind of manipulation that modifies the statistical properties of the original image, e.g. data hiding or contrast enhancement, can be formulated as (10.1) and studied.

The formulation (10.1) can be also performed for checking inconsistencies due to splicing in the suspect region of the image. Based on the proposed fingerprints, if the suspect region contains the fingerprint issued from a different camera model compared with the fingerprint of the rest of the image, the image under investigation could be declared to be inauthentic. This research could be performed in the supervised or unsupervised mode. The supervised mode implies the prior knowledge of the suspect region's location, which implies that forensic analysts only need to check the inconsistency of the suspect region. The unsupervised mode involves a search over the whole image by checking inconsistency of each possible region.

The JPEG compression history [55–57] and double-compression detection problem [62–65] have been considerably studied in the literature. However, these researches are either based on the Laplacian model of DCT coefficient [55, 56] or a simplistic model [64, 65]. Using the proposed model of DCT coefficients is expected to achieve better performance in terms of accuracy of quantization matrix estimation or detection performance of double compression.

10.2.1.2 Video Forensics

It can be noted that research activities in the field of digital forensics mainly focus on analyzing still images. Compared with a variety of fingerprints and techniques

proposed for image forensics, researches about video forensics are generally moderate, see [193] and references therein for an overview of recent video forensic methods. Nowadays, with the advances of remarkable editing tools that make ease of modifying video content. A wide range of falsification operations can be operated on video content, including all possible operations that can be applied to images. The falsification on video content is more difficult to detect not only due to the large amount of possible operations but also due to the complexity of video formation pipeline. The falsification can be applied on each frame of a video sequence, which resemble to an image, as well as can be performed in the temporal dimension. This increases the number of degrees of freedom in the falsification. Moreover, video content is always available in a lossy compression format. Compression techniques may alter or erase fingerprints left by previous processing operations (e.g. PRNU or CFA interpolations) or by falsification operations. Therefore, it is urgent for law enforcement agencies and scientific researches to more focus on video forensics and build up a set of efficient techniques. This thesis has established a complete framework of studying image statistics in the whole image processing pipeline, especially providing a highly accurate image model in spatial domain and DCT domain. A similar framework could be performed for video forensics because of the same characteristics of acquisition device. The proposed model of DCT coefficients is of potential interest since video compression techniques are similar to JPEG compression.

10.2.1.3 Anti-forensics and Counter Anti-forensics

Identifying appropriate intrinsic fingerprints is crucial to perform forensic analysis. However, their credibility can be questioned since a skillful adversary has also knowledge of these fingerprints, thus he can develop techniques [194–196] to hide evidence of image origin and forgery and fool forensic techniques. These techniques are called *anti-forensics*. Like other types of manipulation, anti-forensic techniques may leave behind fingerprints so that forensic analysts can develop countermeasures to detect the use of anti-forensics [197–199]. Moreover, apart from designing a statistical test for digital forensics isolatedly, it is desirable in future researches to take into account actions of adversaries and forensic analysts into the same framework, for instance game-theoretic framework [200] or information-theoretic framework [201], in order to provide an optimal strategy for forensic analysts.

Although anti-forensic techniques are designed to fool forensic ones, they also show some important insights. On the one hand, vulnerabilities in existing forensic techniques help forensic analysts assess the reliability of their forensic techniques and reinforce them to minimize the false alarm probability. On the other hand, anti-forensics can be used by camera manufacturers to protect against reverse-engineering [202] for intellectual property purposes since forensic techniques can estimate image processing operations used by manufacturers inside the camera.

10.2.1.4 Steganalysis

The high detection performance of statistical tests based on the heteroscedastic noise model and the proposed model of DCT coefficients encourages an extension to other image formats and target to other steganographic schemes. For example, future researches could rely on the generalized noise model to design a statistical test for steganalysis of LSB replacement or LSB matching in a natural image in TIFF or JPEG format, similar to [115, 116]. Moreover, the proposed model of DCT coefficients can be applied to steganalysis of Outguess or F5 techniques.

10.2.2 Perspectives to Statistical Image Modeling

The statistical image modeling performed in this thesis could be potentially improved by taking into account the block structure and correlation between DCT coefficients to derive a more generalized model as proposed in Chapter 4, yet at the expense of more complex expressions and extra computational cost.

It should be noted that the algorithm proposed for estimation of parameters $(\tilde{a}, \tilde{b}, \gamma)$ in the generalized noise model only works with a sub-image (i.e. 1/64 of the whole image). This implies a limitation that the proposed algorithm can not work as the inspected image is already cropped to a smaller size. It is desirable to establish a joint distribution of pixels taking into account the spatial correlation in a natural image caused by the demosaicing. This will allow to model a natural image as a whole and extract more meaningful informations for applications in digital forensics. This path of statistical image modeling will be of more interest for detection of forgery in a suspect region with small size.

Beside JPEG compression scheme, JPEG2000 scheme is another influential compression standard. The JPEG2000 scheme is based on Discrete Wavelet Transform (DWT). Therefore, it is also possible to study the distribution of DWT coefficients based on the similar framework proposed in Chapter 4.

10.2.3 Perspectives to Statistical Hypothesis Testing Theory Applied for Digital Forensics

The main framework of our researches in this thesis is binary hypothesis testing. It could be possible for future researches to extend to multiple hypothesis testing for a scenario that includes N sources of images or N types of falsification. Obviously, the criteria of optimality should be redefined. Besides, when the whole image is appropriately modeled, the theory of change detection [203] could be applied to detect any change in statistical properties at the pixel-level due to a certain falsification.

Statistical Hypothesis Testing Theory

A.1 Introduction

The origin identification problem and the steganalysis problem stated in Chapter 2 are concerned with making decisions between statistical hypotheses, i.e. either accepting or rejecting a hypothesis. Typically, a statistical hypothesis represents a specific structure of a population of observations that are associated with stochastic errors. Compared with classification framework, statistical hypothesis testing theory not only provides a rational decision rule for forensic analysts, but also offers many advantages:

- The decision rule is given following a certain criterion of optimality.
- A prescribed false alarm rate can be guaranteed, which is important in the operational context.
- The randomness of the decision can be characterized. Errors probabilities are explicitly and analytically given.
- Stochastic errors associated with empirical observations are taken into account in decision-making process.

The most challenging and difficult task is to determine the specific structure of the underlying observations, i.e. the probability distribution P_{θ} from which data is drawn. The probability distribution P_{θ} is specified by a parameter vector θ that may be entirely or partly unknown. This task is more difficult because we have to deal with high-dimensional natural images in which pixels are not i.i.d and take into account heteroscedasticity of noise. An accurate image model can provide much meaningful information and allows us to design a powerful statistical test.

The whole work presented in this thesis is involved in the application of hypothesis testing theory in image origin identification and hidden data detection in natural images. Therefore this chapter aims to briefly recall some basis aspects of statistical hypothesis testing theory. The chapter is organized as follows. Section A.2 provides basic concepts of statistical hypothesis testing theory. Section A.3 presents some criteria of optimality and statistical tests between two simple hypotheses. Next, Section A.4 presents an approach to deal with a more difficult scenario in which one of two hypotheses becomes composite. Finally, Section A.5 concludes the chapter.

A.2 Basic Concepts

A set of observations obtained in practice can be interfered by stochastic errors (e.g. imperfections of manufacturing process, random noise). Statistical hypothesis testing theory allows us to take into account those stochastic errors to give a decision rule that can meet some criteria of optimality. Therefore, an image $\mathbf{Z} = (z_1, \dots, z_N)^T$ is considered as realizations of a random vector Z following a known probability distribution $Z \sim P_{\boldsymbol{\theta}}$ where the parameter vector $\boldsymbol{\theta}$ may be partly or entirely unknown.

It is assumed that the image space $\mathcal{Z}^N$ can be partitioned into two disjoint subsets $\mathcal{S}_0$ and $\mathcal{S}_1$, i.e.

$$\mathcal{Z}^N = \mathcal{S}_0 \cup \mathcal{S}_1, \quad \text{and} \quad \mathcal{S}_0 \cap \mathcal{S}_1 = \emptyset. \quad (\text{A.1})$$

The behavior of observations in the subset $\mathcal{S}_j$, $j = \{0, 1\}$, is characterized by a class of probability distributions $\{P_{\boldsymbol{\theta}} | \boldsymbol{\theta} \in \boldsymbol{\Theta}_j\}$. Here, the subset $\boldsymbol{\Theta}_j$ contains parameter vectors $\boldsymbol{\theta}$ that specify the probability distribution of observations over $\mathcal{S}_j$. For obvious reasons, we also assume that

$$\boldsymbol{\theta}_1 \neq \boldsymbol{\theta}_2 \Leftrightarrow P_{\boldsymbol{\theta}_1} \neq P_{\boldsymbol{\theta}_2}. \quad (\text{A.2})$$

Without loss of generality, it is assumed that the observations are continuous, and the probability distribution $P_{\boldsymbol{\theta}}$ admits a pdf $f_{\boldsymbol{\theta}} : \mathcal{Z}^N \rightarrow \mathbb{R}$. The subsets $\boldsymbol{\Theta}_0$ and $\boldsymbol{\Theta}_1$ mutually form the parameter space $\boldsymbol{\Theta}$ that contains all possible parameter vectors $\boldsymbol{\theta}$. The dimension of the parameter space $\boldsymbol{\Theta}$ is very small compared with the dimension of the image space $\mathcal{Z}^N$. In the framework of statistical hypothesis testing, the goal is to decide one subset which the parameter vector $\boldsymbol{\theta}$ belongs to. The information of the parameter vector $\boldsymbol{\theta}$ is obtained from realizations $\mathbf{Z}$. For the sake of clarity, only the binary hypothesis testing is considered in this thesis. The extension to multi-hypothesis testing is straightforward.

Definition A.1. (*Statistical hypothesis*). A (parametric) *statistical hypothesis* refers to a set $\boldsymbol{\Theta}$ that contains all possible parameter vector $\boldsymbol{\theta}$ specifying the probability distribution $P_{\boldsymbol{\theta}}$

$$\mathcal{H}_j = \left\{ Z \sim P_{\boldsymbol{\theta}}, \boldsymbol{\theta} \in \boldsymbol{\Theta}_j \right\}, \quad j = \{0, 1\}.$$

The hypothesis $\mathcal{H}_j$ is called *simple* when the set $\boldsymbol{\Theta}_j$ is a singleton. In other words, the simple hypothesis selects a unique point in the parameter space $\boldsymbol{\Theta}$ and defines uniquely the probability distribution $P_{\boldsymbol{\theta}}$. On the contrary, the hypothesis $\mathcal{H}_j$ is called *composite*. Furthermore, by terminology, the hypothesis $\mathcal{H}_0$ is called *null hypothesis* and the hypothesis $\mathcal{H}_1$ is *alternative hypothesis*.

Definition A.2. (*Statistical test*). A *statistical test* δ is a surjective and measurable application from the image space $\mathcal{Z}^N$ to the set of hypotheses $\delta : \mathcal{Z}^N \rightarrow \{\mathcal{H}_0, \mathcal{H}_1\}$. In other words, the test δ gives a decision rule of either accepting or rejecting a hypothesis.

The ultimate goal of statistical hypothesis testing theory is to design a test that is optimal towards some criteria. Mono-criteria approach, e.g. Bayesian test and minimax test, involves designing a test such that it satisfies only one criterion while a test designed following bi-criteria approach must satisfy two criteria of optimality simultaneously. The readers are referred to [20, 204, 205] for more details of mono-criteria approach. The bi-criteria approach is discussed in this chapter.

A.3 Test between Two Simple Hypotheses

This section considers the simplest scenario in which two simple hypotheses $\mathcal{H}_0$ and $\mathcal{H}_1$ are defined as follows

$$\begin{cases} \mathcal{H}_0 &= \{\boldsymbol{\theta} = \boldsymbol{\theta}_0\} \\ \mathcal{H}_1 &= \{\boldsymbol{\theta} = \boldsymbol{\theta}_1\}, \end{cases} \quad (\text{A.3})$$

where $\boldsymbol{\theta}_0$ and $\boldsymbol{\theta}_1$ are known in advance. When performing a test δ between two simple hypotheses $\mathcal{H}_0$ and $\mathcal{H}_1$, we may arrive at the correct decision, or we may commit one of two errors:

1. We may wrongly reject the null hypothesis $\mathcal{H}_0$ when it is true (Type I error).
2. We may wrongly accept $\mathcal{H}_0$ when it is false (Type II error).

Definition A.3. (*Error probabilities*). Each error is measured by a probability

$$\mathbb{P}_{\mathcal{H}_j}[\delta(\mathbf{Z}) \neq \mathcal{H}_j] \triangleq \alpha_j(\delta), \quad j = \{0, 1\},$$

where $\mathbb{P}_{\mathcal{H}_j}$ is the probability of occurrence under hypothesis $\mathcal{H}_j$. By terminology, $\alpha_0(\delta)$ and $\alpha_1(\delta)$ are called *false alarm probability* and *miss-detection probability* of the test δ , respectively.

Definition A.4. (*Power of a test*). The *power* of the test δ , denoted $\beta(\delta)$, is defined by the correct detection probability

$$\beta(\delta) \triangleq \mathbb{P}_{\mathcal{H}_1}[\delta(\mathbf{Z}) = \mathcal{H}_1] = 1 - \alpha_1(\delta).$$

Two criteria that are required to optimize are the false alarm probability $\alpha_0(\delta)$ and the test power $\beta(\delta)$. The idea is to set an acceptable maximum false alarm probability and to maximize the correct detection probability. Therefore, let define a class of tests $\mathcal{K}_{\alpha_0}$ as follows

$$\mathcal{K}_{\alpha_0} = \left\{ \delta : \alpha_0(\delta) = \mathbb{P}_{\mathcal{H}_0}[\delta(\mathbf{Z}) = \mathcal{H}_1] \leq \alpha_0 \right\}. \quad (\text{A.4})$$

The class $\mathcal{K}_{\alpha_0}$ consists of all the tests whose false alarm probability is upper-bounded by a prescribed rate α_0 . For different levels α_0 we can achieve different test powers $\beta(\delta)$. The behavior of the test power $\beta(\delta)$ is characterized by a function of α_0 , which is called Receiver Operating Characteristic (ROC) curve. The detection performance of the test δ is evaluated via ROC curve.

Definition A.5. (*Most powerful test*). A test δ^* is called *Most Powerful* (MP) in the class $\mathcal{K}_{\alpha_0}$ if

$$\forall \delta \in \mathcal{K}_{\alpha_0}, \quad \beta(\delta^*) \geq \beta(\delta).$$

The MP test achieves two goals simultaneously: warranting a prescribed false alarm rate and maximizing the correct detection probability. The MP test is given by the Neyman-Pearson (NP) lemma [20, Theorem 3.2.1].

Definition A.6. (*Likelihood Ratio*). *Likelihood Ratio* (LR) is the quantity

$$\Lambda(\mathbf{Z}) = \frac{f_{\boldsymbol{\theta}_1}(\mathbf{Z})}{f_{\boldsymbol{\theta}_0}(\mathbf{Z})}.$$

Lemma A.1. (Neyman-Pearson lemma). *Given a prescribed rate $\alpha_0 \in]0; 1[$ and two probability distributions $P_{\boldsymbol{\theta}_0}$ and $P_{\boldsymbol{\theta}_1}$, when one performs a test of null hypothesis $\mathcal{H}_0 = \{\boldsymbol{\theta} = \boldsymbol{\theta}_0\}$ against alternative hypothesis $\mathcal{H}_1 = \{\boldsymbol{\theta} = \boldsymbol{\theta}_1\}$, the MP test in the class $\mathcal{K}_{\alpha_0}$ is the Likelihood Ratio Test (LRT) given by*

$$\delta^*(\mathbf{Z}) = \begin{cases} \mathcal{H}_0 & \text{if } \Lambda(\mathbf{Z}) = \frac{f_{\boldsymbol{\theta}_1}(\mathbf{Z})}{f_{\boldsymbol{\theta}_0}(\mathbf{Z})} < \tau^* \\ \mathcal{H}_1 & \text{if } \Lambda(\mathbf{Z}) = \frac{f_{\boldsymbol{\theta}_1}(\mathbf{Z})}{f_{\boldsymbol{\theta}_0}(\mathbf{Z})} \geq \tau^*, \end{cases}$$

where, to ensure that the LRT δ^* is in the class $\mathcal{K}_{\alpha_0}$, the optimal decision threshold τ^* is the solution of the equation

$$\mathbb{P}_{\mathcal{H}_0}[\Lambda(\mathbf{Z}) \geq \tau^*] = \alpha_0.$$

When the random variables z_i , $i = \{1, 2, \dots, N\}$, are statistically independent, the Maximum Likelihood (ML) function $f_{\boldsymbol{\theta}_j}(\mathbf{Z})$, $j = \{0, 1\}$, can be rewritten as $f_{\boldsymbol{\theta}_j}(\mathbf{Z}) = \prod_{i=1}^N f_{\boldsymbol{\theta}_j}(z_i)$. Accordingly, the LR $\Lambda(\mathbf{Z})$ can be given by

$$\Lambda(\mathbf{Z}) = \prod_{i=1}^N \frac{f_{\boldsymbol{\theta}_1}(z_i)}{f_{\boldsymbol{\theta}_0}(z_i)} = \prod_{i=1}^N \Lambda(z_i). \quad (\text{A.5})$$

Taking logarithm on both sides of (A.5), we obtain the log-LR

$$\log(\Lambda(\mathbf{Z})) = \sum_{i=1}^N \log(\Lambda(z_i)) = \sum_{i=1}^N \log\left(\frac{f_{\boldsymbol{\theta}_1}(z_i)}{f_{\boldsymbol{\theta}_0}(z_i)}\right). \quad (\text{A.6})$$

Since the logarithm function is monotone, taking logarithm of the LR $\Lambda(\mathbf{Z})$ does not change the decision given by the LRT δ^* . Only the decision threshold is changed according to the log-LR $\log(\Lambda(\mathbf{Z}))$,¹ the test power $\beta(\delta^*)$ remains unchanged. In

¹Accordingly, the decision threshold τ^* is replaced by $\log(\tau^*)$.

practice, the log-LR is more utilized than the LR. For the sake of clarity, in this thesis the logarithm is usually taken in the calculation of the LR $\Lambda(\mathbf{Z})$, i.e.

$$\Lambda(\mathbf{Z}) = \sum_{i=1}^N \Lambda(z_i) = \sum_{i=1}^N \log \left(\frac{f_{\theta_1}(z_i)}{f_{\theta_0}(z_i)} \right). \quad (\text{A.7})$$

One advantage of statistical hypothesis testing theory is that it allows to establish analytically the statistical performance of a test. This is accomplished by studying the statistical distribution of the LR $\Lambda(\mathbf{Z})$. However, providing analytically an exact distribution of a sum of random variables is a difficult task, particularly when the distribution of each random variable $\Lambda(z_i)$ can not be easily defined in practice. To overcome this difficulty, the asymptotic theory can be exploited based on the fact that the number of pixels is very large in natural images. To this end, it is proposed to rely on the Lindeberg CLT [20, Theorem 11.2.5]. One advantage of the Lindeberg CLT is that the random variables are not required to be identically distributed. It can deal with the difficult scenario where pixels in a natural image are not identically distributed.

Definition A.7. (*Convergence in distribution*). A sequence of random variables $\{X_n\}$ with cdf $\{F_{X_n}(x)\}$ is said to *converge in distribution* to a random variable X with cdf $F_X(x)$ if

$$F_{X_n}(x) \rightarrow F_X(x)$$

at every point $x \in \mathbb{R}$. This convergence is denoted $X_n \xrightarrow{d} X$.

Definition A.8. (*Convergence in probability*). A sequence of random variables $\{X_n\}$ *converges in probability* to a random variable X if, for every $\epsilon > 0$,

$$\lim_{n \rightarrow \infty} \mathbb{P}[|X_n - X| > \epsilon] = 0.$$

This convergence is denoted $X_n \xrightarrow{p} X$. Convergence in probability implies convergence in distribution; the converse is false in general. However, if X_n converges in distribution to a distribution assigning probability one to a constant c , then X_n converges in probability to c , and conversely.

Theorem A.1. (Lindeberg CLT). Suppose a sequence of independent random variables $\{X_n\}$, each with finite mean μ_n and finite variance σ_n^2 . Let $m_n = \sum_{i=1}^n \mu_i$ and $s_n^2 = \sum_{i=1}^n \sigma_i^2$. If this sequence satisfies the Lindeberg's condition, i.e. for every $\epsilon > 0$,

$$\lim_{n \rightarrow \infty} \frac{1}{s_n^2} \sum_{i=1}^n \mathbb{E} \left[(X_i - \mu_i)^2 \cdot \mathbf{1}_{|X_i - \mu_i| > \epsilon \cdot s_n} \right] = 0,$$

where $\mathbf{1}_{(\cdot)}$ is the indicator function, then

$$\frac{\sum_{i=1}^n X_i - m_n}{s_n} \xrightarrow{d} \mathcal{N}(0, 1),$$

or equivalently,

$$\sum_{i=1}^n X_i \xrightarrow{d} \mathcal{N}(m_n, s_n^2).$$

Therefore, in order to determine the asymptotic distribution of the LR $\Lambda(\mathbf{Z})$ in virtue of the Lindeberg CLT, it is sufficient to calculate mean and variance of each random variable $\Lambda(z_i)$. Here we present the Delta methods that can facilitate the calculation of the first two moments of a random variable.

Lemma A.2. (The first Delta method [20, theorem 11.2.14]) Suppose $X_1, X_2, \dots$ and X are random vectors in $\mathbb{R}^k$. Assume $\tau_n(X_n - \mu) \xrightarrow{d} X$ where μ is a constant vector and $\{\tau_n\}$ is a sequence of constants $\tau_n \rightarrow \infty$. Suppose h is a function from $\mathbb{R}^k$ to $\mathbb{R}$ which is differentiable at μ with gradient of dimension $1 \times k$ at μ equal to $\dot{h}(\mu)$. Then

$$\tau_n(h(X_n) - h(\mu)) \xrightarrow{d} \dot{h}(\mu)X.$$

Corollary A.1. Suppose X a random variable with mean μ and variance σ^2 . Suppose h is function from $\mathbb{R}$ to $\mathbb{R}$ which is differentiable at μ . The first two moments of the random variable $h(X)$ can be approximated as

$$\mathbb{E}[h(X)] = h(\mu) \quad \text{and} \quad \mathbb{V}\text{ar}[h(X)] = (h'(\mu))^2 \sigma^2.$$

Lemma A.3. (The second Delta method [189]) Suppose two random variables X and Y with mean μ_X, μ_Y and variance σ_X^2, σ_Y^2 , respectively. The covariance between X and Y is denoted as σ_{XY} . Consider the quotient $T = \frac{X}{Y}$. The first two moments of the random variable T can be approximated as

$$\begin{aligned} \mathbb{E}\left[\frac{X}{Y}\right] &= \frac{\mu_X}{\mu_Y} \\ \mathbb{V}\text{ar}\left[\frac{X}{Y}\right] &= \frac{\sigma_X^2}{\mu_Y^2} - 2\frac{\mu_X}{\mu_Y^3}\sigma_{XY} + \frac{\mu_X^2}{\mu_Y^4}\sigma_Y^2. \end{aligned}$$

Example A.1. Given an observation $\mathbf{Z}$ and two probability distributions $P_{\theta_0}, P_{\theta_1}$, suppose the expectation and variance of the LR $\Lambda(z_i)$ are known in advance. In virtue of Lindeberg CLT, the LR $\Lambda(\mathbf{Z})$ can be asymptotically distributed following a Gaussian distribution under each hypothesis $\mathcal{H}_j, j = \{0, 1\}$,

$$\Lambda(\mathbf{Z}) \xrightarrow{d} \mathcal{N}(\mu_j, \sigma_j^2), \tag{A.8}$$

where the mean μ_j and variance σ_j^2 are given by

$$\mu_j = \sum_{i=1}^N \mathbb{E}_{\mathcal{H}_j}[\Lambda(z_i)] \tag{A.9}$$

$$\sigma_j^2 = \sum_{i=1}^N \mathbb{V}\text{ar}_{\mathcal{H}_j}[\Lambda(z_i)]. \tag{A.10}$$

Here, $\mathbb{E}_{\mathcal{H}_j}$ and $\text{Var}_{\mathcal{H}_j}$ respectively denote the mathematical expectation and variance under hypothesis $\mathcal{H}_j$. Then the optimal decision threshold τ^* and the test power $\beta(\delta^*)$ are easily expressed as

$$\tau^* = \mu_0 + \sigma_0 \cdot \Phi^{-1}(1 - \alpha_0), \quad (\text{A.11})$$

$$\beta(\delta^*) = 1 - \Phi\left(\frac{\tau^* - \mu_1}{\sigma_1}\right). \quad (\text{A.12})$$

Proof. The decision threshold τ^* is the solution of equation

$$\alpha_0 = \mathbb{P}_{\mathcal{H}_0}[\Lambda^*(\mathbf{Z}) \geq \tau^*] = 1 - \Phi\left(\frac{\tau^* - \mu_0}{\sigma_0}\right),$$

thus we derive $\tau^* = \mu_0 + \sigma_0 \cdot \Phi^{-1}(1 - \alpha_0)$. Meanwhile, the power $\beta(\delta^*)$ is defined as the correct detection probability

$$\beta(\delta^*) = \mathbb{P}_{\mathcal{H}_0}[\Lambda^*(\mathbf{Z}) \geq \tau^*] = 1 - \Phi\left(\frac{\tau^* - \mu_1}{\sigma_1}\right).$$

□

Remark A.1. It can be noted that the decision threshold τ^* depends on image content. Since a natural image is heterogeneous, it is proposed to normalize the LR $\Lambda(\mathbf{Z})$ in order to set the decision threshold independently of the image content. The normalized LR is defined as follows

$$\Lambda^*(\mathbf{Z}) = \frac{\Lambda(\mathbf{Z}) - \mu_0}{\sigma_0}. \quad (\text{A.13})$$

Therefore, the normalized LR $\Lambda^*(\mathbf{Z})$ differs from the LR $\Lambda(\mathbf{Z})$ only by an additive constant and a multiplicative constant, which does not change the decision rule given by the LRT δ^* . In doing so, the normalized LR $\Lambda^*(\mathbf{Z})$ tends to be distributed following the standard Gaussian distribution under hypothesis $\mathcal{H}_0$. The decision threshold and the power of the LRT based on the normalized LR $\Lambda^*(\mathbf{Z})$ are given by

$$\tau^* = \Phi^{-1}(1 - \alpha_0), \quad (\text{A.14})$$

$$\beta(\delta^*) = 1 - \Phi\left(\frac{\mu_0 - \mu_1 + \tau^* \sigma_0}{\sigma_1}\right). \quad (\text{A.15})$$

A.4 Test between Two Composite Hypotheses

The scenario where both null hypothesis and alternative hypothesis are simple is mainly of theoretical interest, since problems arising in practical applications typically involve a parametric family of distributions specified by unknown parameters. Suppose that two hypotheses $\mathcal{H}_0$ and $\mathcal{H}_1$ are now defined as

$$\begin{cases} \mathcal{H}_0 &= \{P_{\theta} | \theta \in \Theta_0\} \\ \mathcal{H}_1 &= \{P_{\theta} | \theta \in \Theta_1\}. \end{cases} \quad (\text{A.16})$$

In this scenario, the NP lemma is no longer valid since for each parameter $\theta_0 \in \Theta_0$ and $\theta_1 \in \Theta_1$ we derive a different MP test. Therefore, it is important to design a test that is universally optimal. Firstly, let us redefine the false alarm probability and the power of a test δ between two composites hypotheses.

Definition A.9. (*False alarm probability*). The *false alarm probability* of a test δ between two composites hypotheses is the maximum probability of rejecting null hypothesis $\mathcal{H}_0$ when it is true actually

$$\alpha_0(\delta) = \sup_{\theta_0 \in \Theta_0} \mathbb{P}_{\mathcal{H}_0} [\delta(\mathbf{Z}) = \mathcal{H}_1].$$

Consequently, the class of tests $\mathcal{K}_{\alpha_0}$ is redefined as follows

$$\mathcal{K}_{\alpha_0} = \left\{ \delta : \alpha_0(\delta) = \sup_{\theta_0 \in \Theta_0} \mathbb{P}_{\mathcal{H}_0} [\delta(\mathbf{Z}) = \mathcal{H}_1] \leq \alpha_0 \right\}. \quad (\text{A.17})$$

This means that the class $\mathcal{K}_{\alpha_0}$ comprises all the tests whose false alarm probability is upper-bounded by a prescribed rate α_0 for any parameter $\theta_0 \in \Theta_0$.

Definition A.10. (*Power of a test*). The *power* of a test δ of null hypothesis $\mathcal{H}_0$ against alternative hypothesis $\mathcal{H}_1$ for each parameter $\theta_1 \in \Theta_1$ is the probability of accepting alternative hypothesis $\mathcal{H}_1$ when it is true

$$\beta(\theta_1; \delta) = \mathbb{P}_{\mathcal{H}_1} [\delta(\mathbf{Z}) = \mathcal{H}_1].$$

Following the bi-criteria approach, the goal is to design a test in the class $\mathcal{K}_{\alpha_0}$ that can warrant a prescribed false alarm rate α_0 and maximizes the test power $\beta(\theta_1; \delta)$ in the domain Θ_1 .

Definition A.11. (*Uniformly most powerful test*). A test δ^* is said to be Uniformly Most Powerful (UMP) over the class $\mathcal{K}_{\alpha_0}$ if

$$\forall \delta \in \mathcal{K}_{\alpha_0}, \quad \forall \theta_1 \in \Theta_1, \quad \beta(\theta_1; \delta) \leq \beta(\theta_1; \delta^*),$$

or equivalently,

$$\forall \delta \in \mathcal{K}_{\alpha_0}, \quad \sup_{\theta_1 \in \Theta_1} [\beta(\theta_1; \delta) - \beta(\theta_1; \delta^*)] \leq 0.$$

Unfortunately, such UMP test might scarcely exist in practice. In fact, the existence of UMP tests is restricted to the case involved in univariate family of densities f_{θ} admitting a monotone likelihood ratio [20, Theorem 3.4.1]. An alternative solution is to design a test that asymptotically coincides with a UMP test in virtue of a large number of pixels in a natural image.

Definition A.12. (*Asymptotically Uniformly Most Powerful test*). The test δ^* is *Asymptotically Uniformly Most Powerful* (AUMP) in the class $\mathcal{K}_{\alpha_0}^{\infty}$, given by

$$\mathcal{K}_{\alpha_0}^{\infty} = \left\{ \delta : \limsup_{N \rightarrow +\infty} \sup_{\theta_0 \in \Theta_0} \mathbb{P}_{\mathcal{H}_0} [\delta(\mathbf{Z}) = \mathcal{H}_1] \leq \alpha_0 \right\},$$

if $\delta^* \in \mathcal{K}_{\alpha_0}^\infty$ and

$$\forall \delta \in \mathcal{K}_{\alpha_0}^\infty, \quad \limsup_{N \rightarrow +\infty} \sup_{\boldsymbol{\theta}_1 \in \boldsymbol{\Theta}_1} [\beta(\boldsymbol{\theta}_1; \delta) - \beta(\boldsymbol{\theta}_1; \delta^*)] \leq 0.$$

In general, the state of the art of statistical hypothesis testing theory shows that it is difficult to design a single test that is optimal in any scenario (e.g. univariate or multivariate family, simple or composite hypotheses ...). Nevertheless, there exists a test that coincides with the optimal tests under large conditions, namely Generalized Likelihood Ratio Test (GLRT) [206, 207]. The GLRT is based on a modified version of Neyman-Pearson approach.

Definition A.13. (*Generalized Likelihood Ratio Test*). Given two composite hypotheses $\mathcal{H}_0$ and $\mathcal{H}_1$, the Generalized Likelihood Ratio (GLR) between these hypotheses is defined by

$$\hat{\Lambda}(\mathbf{Z}) = \log \frac{\sup_{\boldsymbol{\theta}_1 \in \boldsymbol{\Theta}_1} f_{\boldsymbol{\theta}_1}(\mathbf{Z})}{\sup_{\boldsymbol{\theta}_0 \in \boldsymbol{\Theta}_0} f_{\boldsymbol{\theta}_0}(\mathbf{Z})}.$$

Consequently, the GLRT $\hat{\delta}$ is defined by the following decision rule

$$\hat{\delta}(\mathbf{Z}) = \begin{cases} \mathcal{H}_0 & \text{if } \hat{\Lambda}(\mathbf{Z}) < \hat{\tau} \\ \mathcal{H}_1 & \text{if } \hat{\Lambda}(\mathbf{Z}) \geq \hat{\tau}, \end{cases}$$

where the decision threshold $\hat{\tau}$ is the solution of the equation

$$\sup_{\boldsymbol{\theta}_0 \in \boldsymbol{\Theta}_0} \mathbb{P}_{\mathcal{H}_0} [\hat{\Lambda}(\mathbf{Z}) \geq \hat{\tau}] = \alpha_0,$$

to ensure that the GLRT $\hat{\delta}$ is in the class $\mathcal{K}_{\alpha_0}$.

The idea of GLRT is to select the parameters the most likely characterizing the distribution of a given observation $\mathbf{Z}$. This is accomplished by estimating unknown parameters $\boldsymbol{\theta}$ by the ML approach. In doing so, the parameter space reduces to a singleton. The Neyman-Pearson can be applied straightforwardly.

Definition A.14. (*Maximum Likelihood estimation*). Given a sample $\mathbf{x} = \{x_1, \dots, x_n\}$ of n i.i.d observations coming from a probability distribution with a pdf $f_{\boldsymbol{\theta}}$, the ML estimate $\hat{\boldsymbol{\theta}}^{\text{ML}}$ is defined as the solution maximizing the likelihood function $\mathcal{L}(\boldsymbol{\theta}; \mathbf{x})$

$$\hat{\boldsymbol{\theta}}^{\text{ML}} = \arg \max_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \mathcal{L}(\boldsymbol{\theta}; \mathbf{x}) = \arg \max_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \prod_{i=1}^n f_{\boldsymbol{\theta}}(x_i; \boldsymbol{\theta}), \quad (\text{A.18})$$

or equivalently

$$\hat{\boldsymbol{\theta}}^{\text{ML}} = \arg \max_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \sum_{i=1}^n \log (f_{\boldsymbol{\theta}}(x_i; \boldsymbol{\theta})). \quad (\text{A.19})$$

For some probability distributions such as Normal distribution, the ML estimate can be provided analytically as an explicit function of the observations $\mathbf{x}$. However, in many cases in practice, it is delicate to provide it analytically. In this scenario, it is proposed to resolve the maximization problem (A.19) numerically by an optimization method, e.g. Nelder-Mead method [157].

Under some regularity conditions, as the sample size n tends to infinity, the ML estimate $\hat{\boldsymbol{\theta}}^{\text{ML}}$ has following properties [189]:

- Asymptotic consistency: $\hat{\boldsymbol{\theta}}^{\text{ML}} \xrightarrow{p} \boldsymbol{\theta}_0$ where $\boldsymbol{\theta}_0$ denotes the true value of the parameter $\boldsymbol{\theta}$.
- Asymptotic normality: $\hat{\boldsymbol{\theta}}^{\text{ML}} \xrightarrow{d} \mathcal{N}\left(\boldsymbol{\theta}_0, \frac{\mathbf{F}^{-1}(\boldsymbol{\theta})}{n}\right)$ where $\mathbf{F}$ denotes the Fisher information matrix defined as

$$\begin{aligned} \mathbf{F}_{i,j}(\boldsymbol{\theta}) &= \mathbb{E} \left[\left(\frac{\partial}{\partial \theta_i} \log(f_{\boldsymbol{\theta}}(\mathbf{x}; \boldsymbol{\theta})) \right) \left(\frac{\partial}{\partial \theta_j} \log(f_{\boldsymbol{\theta}}(\mathbf{x}; \boldsymbol{\theta})) \right) \right] \\ &= -\mathbb{E} \left[\frac{\partial^2}{\partial \theta_i \partial \theta_j} \log(f_{\boldsymbol{\theta}}(\mathbf{x}; \boldsymbol{\theta})) \right]. \end{aligned} \quad (\text{A.20})$$

These two properties of ML estimate are of importance in the design of a GLRT since the fact of replacing unknown parameters by their ML estimates might not be sufficient. It is also desirable to take into account the variability of ML estimates in the establishment of statistical performances of the GLRT. Therefore, the detection performance of a GLRT in practice depends on two crucial conditions: the relevance of the probability distribution describing the underlying observation and the accuracy estimation of unknown parameters.

In Remark A.1, the fact of defining the normalized LR as (A.13) requires the prior knowledge of the expectation μ_0 and variance σ_0^2 that depend on image content. Since this latter is unknown in practice, the expectation μ_0 and variance σ_0^2 can not be defined. A solution in this scenario is to exploit the consistent estimates that was employed to design the GLRT in order to provide an estimate of μ_0 and σ_0^2 , denoted $\hat{\mu}_0$ and $\hat{\sigma}_0^2$. Consequently, the normalized GLR can be defined as

$$\hat{\Lambda}^*(\mathbf{Z}) = \frac{\hat{\Lambda}(\mathbf{Z}) - \hat{\mu}_0}{\hat{\sigma}_0} \quad (\text{A.21})$$

In order to determine the statistical distribution of the normalized GLR $\hat{\Lambda}^*(\mathbf{Z})$, it is proposed to rely on the Slutsky's theorem.

Theorem A.2. (Slutsky's theorem [20, theorem 11.2.11]). *Suppose $\{X_n\}$ is a sequence of real-valued random variables such that $X_n \xrightarrow{d} X$. Further, suppose $\{A_n\}$ and $\{B_n\}$ satisfy $A_n \xrightarrow{p} a$ and $B_n \xrightarrow{p} b$, where a and b are constants. Then*

$$A_n X_n + B_n \xrightarrow{d} aX + b. \quad (\text{A.22})$$

Remark A.2. Suppose that the statistical distribution of the GLR $\hat{\Lambda}(\mathbf{Z})$ under hypothesis $\mathcal{H}_j$ can be determined in virtue of Lindeberg CLT, say $\hat{\Lambda}(\mathbf{Z}) \xrightarrow{d} \mathcal{N}(\mu_j, \sigma_j^2)$, from Slutsky's theorem, the normalized GLR $\hat{\Lambda}^*(\mathbf{Z})$ follows the Gaussian distribution as

$$\begin{cases} \hat{\Lambda}^*(\mathbf{Z}) \xrightarrow{d} \mathcal{N}(0, 1) & \text{under } \mathcal{H}_0, \\ \hat{\Lambda}^*(\mathbf{Z}) \xrightarrow{d} \mathcal{N}\left(\frac{\mu_1 - \mu_0}{\sigma_0}, \frac{\sigma_1^2}{\sigma_0^2}\right) & \text{under } \mathcal{H}_1. \end{cases} \quad (\text{A.23})$$

Consequently, the decision threshold and the power function of the GLRT $\hat{\delta}$ can be accordingly defined as in Remark A.1.

Remark A.3. In many applications, the parameter vector $\boldsymbol{\theta}$ consists of informative parameters and nuisance parameters. Informative parameter vector is assumed to be constant within the observation $\mathbf{Z}$. For instance, in the problem of hidden data detection, the only one informative parameter is the embedding rate R , other parameters such as image content, image size, camera settings are nuisance parameters since they do not contain any information about the presence of secret message. Nuisance parameters are not of interest but they also characterize the family of distribution $P_{\boldsymbol{\theta}}$, thus they must be taken into account in the design of statistical tests. This becomes difficult since the dimension of nuisance parameter space may be large and nuisance parameters are generally unknown in advance. The GLRT allows us to deal with the difficulty of unknown nuisance parameters by replacing them by ML estimates. Besides, another possible approach is to use invariance principle [20] to eliminate nuisance parameters.

A.5 Conclusion

In this chapter, several aspects about statistical hypothesis testing theory are briefly discussed. Hypothesis testing theory allows us to warrant a prescribed false alarm probability and analytically establish the performance of statistical tests, which is of importance in the operational context. In a simplest scenario where two hypotheses are simple, an optimal detector is given by the LRT. The statistical decision problem becomes more difficult when two hypotheses are composites. A UMP test is mainly of theoretical interest and scarcely exists. An AUMP test is presented in virtue of a large number of pixels in a natural image. In practice, a GLRT is designed by replacing unknown parameters by their ML estimates. The GLRT almost coincides with the UMP test under large conditions. Moreover, the GLRT also allows us to deal with the difficulty of unknown nuisance parameters. To achieve high detection performance for the GLRT, an accurate statistical model of natural images and an efficient method of parameter estimation are crucial.

Bibliography

- [1] C.-T. Hsu and J.-L. Wu, “Hidden digital watermarks in images,” *IEEE Transactions on Image Processing*, vol. 8, no. 1, pp. 58–68, Jan. 1999. (Cited on page 11.)
- [2] L. Me and G. R. Arce, “A class of authentication digital watermarks for secure multimedia communication,” *IEEE Transactions on Image Processing*, vol. 10, no. 11, pp. 1754–1764, Nov. 2001. (Cited on page 11.)
- [3] C. I. Podilchuk and E. J. Delp, “Digital watermarking: algorithms and applications,” *IEEE Signal Processing Magazine*, vol. 18, no. 4, pp. 33–46, Jul. 2001. (Cited on page 11.)
- [4] P. W. Wong and N. Memon, “Secret and public key image watermarking schemes for image authentication and ownership verification,” *IEEE Transactions on Image Processing*, vol. 10, no. 10, pp. 1593–1601, Oct. 2001. (Cited on page 11.)
- [5] L. Xie and G. R. Arce, “A class of authentication digital watermarks for secure multimedia communication,” *IEEE Transactions on Image Processing*, vol. 10, no. 11, pp. 1754–1764, Nov. 2001. (Cited on page 11.)
- [6] Z. Guojuan and L. Dianji, “An overview of digital watermarking in image forensics,” in *International Joint Conference on Computational Sciences and Optimization*, Avr. 2011, pp. 332–335. (Cited on page 11.)
- [7] G. L. Friedman, “The trustworthy digital camera: restoring credibility to the photographic image,” *IEEE Transactions on Consumer Electronics*, vol. 39, no. 4, pp. 905–910, Nov. 1993. (Cited on page 11.)
- [8] M. Schneider and S.-F. Chang, “A robust content based digital signature for image authentication,” in *IEEE International Conference on Image Processing*, vol. 3, Sep. 1996, pp. 227–230. (Cited on page 11.)
- [9] C.-Y. Lin and S.-F. Chang, “A robust image authentication method distinguishing JPEG compression from malicious manipulations,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 11, no. 2, pp. 153–168, Feb. 2001. (Cited on page 11.)
- [10] C.-S. Lu and H.-K. M. Liao, “Structural digital signature for image authentication: an incidental distortion resistant scheme,” *IEEE Transactions on Multimedia*, vol. 5, no. 2, pp. 161–173, Jun. 2003. (Cited on page 11.)
- [11] R. Ramanath, W. E. Snyder, Y. Yoo, and M. S. Drew, “Color image processing pipeline,” *IEEE Signal Processing Magazine*, vol. 22, no. 1, pp. 34–43, Jan. 2005. (Cited on pages 12, 15 and 16.)

- [12] H. T. Sencar and H. Menon, *Digital Image Forensics: There is More to a Picture than Meets the Eye*. New York: Springer, 2013. (Cited on pages 13 and 16.)
- [13] J. Janesick, T. Elliott, S. Collins, M. Blouke, and J. Freeman, *Scientific Charge-Coupled Devices*. SPIE Publications, 2001. (Cited on page 14.)
- [14] R. Ramanath, W. E. Snyder, and G. L. Bilbro, "Demosaicking methods for Bayer color arrays," *Journal of Electronic Imaging*, vol. 11, no. 3, pp. 306–615, Jul. 2002. (Cited on page 15.)
- [15] W. Pennebaker and J. Mitchell, *JPEG Still Image Compression Data*. Springer, 1992. (Cited on page 17.)
- [16] K. Lee, D. S. Kim, and T. Kim, "Regression-based prediction for blocking artifact reduction in JPEG-compressed images," *IEEE Transactions on Image Processing*, vol. 14, no. 1, pp. 36–48, Jan. 2005. (Cited on page 19.)
- [17] T. Mitchell, *Machine Learning*. McGraw Hill, 1997. (Cited on page 20.)
- [18] V. N. Vapnik, *The Nature of Statistical Learning Theory*, 2nd ed. Springer Verlag, 2000. (Cited on pages 20 and 21.)
- [19] 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 pages 20 and 21.)
- [20] E. L. Lehmann and J. P. Romano, *Testing Statistical Hypotheses*, 3rd ed. New York: Springer, 2005. (Cited on pages 20, 37, 55, 83, 85, 123, 183, 184, 185, 186, 188, 190 and 191.)
- [21] C.-C. Chang and C.-J. Lin, "LIBSVM: a library for support vector machines. Available at <http://www.csie.ntu.edu.tw/~cjlin/svm>," 2001. (Cited on page 21.)
- [22] C. Scott, "Performance measures for Neyman-Pearson classification," *IEEE Transactions on Information Theory*, vol. 53, no. 8, pp. 2852–2863, Aug. 2007. (Cited on page 21.)
- [23] K. S. Choi, E. Y. Lam, and K. Wong, "Source camera identification using footprints from lens aberration," in *SPIE Digital Photography*, vol. 6069, Feb. 2006, pp. 172–179. (Cited on pages 22 and 23.)
- [24] A. Swaminathan, M. Wu, and K. J. R. Liu, "Nonintrusive component forensics of visual sensors using output images," *IEEE Transactions on Information Forensics and Security*, vol. 2, no. 1, pp. 91–106, Mar. 2007. (Cited on pages 22, 25, 26 and 93.)

- [25] —, “Digital image forensics via intrinsic fingerprints,” *IEEE Transactions on Information Forensics and Security*, vol. 3, no. 1, pp. 101–117, Mar. 2008. (Cited on pages 22 and 27.)
- [26] H. Cao and A. C. Kot, “Accurate detection of demosaicing regularity for digital image forensics,” *IEEE Transactions on Information Forensics and Security*, vol. 4, no. 4, pp. 899–910, Dec. 2009. (Cited on pages 22 and 26.)
- [27] K. S. Choi, E. Y. Lam, and K. K. Y. Wong, “Source camera identification by JPEG compression statistics for image forensics,” in *IEEE Region Conference TENCON*, Nov. 2006, pp. 1–4. (Cited on page 22.)
- [28] G. Xu, S. Gao, Y. Q. Shi, R. Hu, and W. Su, “Camera model identification using markovian transition probability matrix,” in *Digital Watermarking*, vol. 5703. Springer, Aug. 2009, pp. 294–307. (Cited on pages 22, 26 and 27.)
- [29] J. Lukas, J. Fridrich, and M. Goljan, “Digital camera identification from sensor pattern noise,” *IEEE Transactions on Information Forensics and Security*, vol. 1, no. 2, pp. 205–214, Jun. 2006. (Cited on pages 22, 62 and 126.)
- [30] M. Chen, J. Fridrich, M. Goljan, and J. Lukas, “Determining image origin and integrity using sensor noise,” *IEEE Transactions on Information Forensics and Security*, vol. 3, no. 1, pp. 74–90, Mar. 2008. (Cited on pages 22, 24, 27, 41, 93 and 94.)
- [31] T. Filler, J. Fridrich, and M. Goljan, “Using sensor pattern noise for camera model identification,” in *IEEE International Conference on Image Processing*, Oct. 2008, pp. 1296–1299. (Cited on pages 22, 25 and 93.)
- [32] J. Fridrich, “Digital image forensics,” *IEEE Signal Processing Magazine*, vol. 26, no. 2, pp. 26–37, Mar. 2009. (Cited on pages 22, 24, 27, 41, 93, 94, 111 and 131.)
- [33] I. Amerini, R. Caldelli, V. Cappellini, F. Picchioni, and A. Piva, “Analysis of denoising filters for photo response non uniformity noise extraction in source camera identification,” in *International Conference on Digital Signal Processing*, Jul. 2009, pp. 1–7. (Cited on pages 22 and 25.)
- [34] C.-T. Li, “Source camera identification using enhanced sensor pattern noise,” *IEEE Transactions on Information Forensics and Security*, vol. 5, no. 2, pp. 280–287, Jun. 2010. (Cited on pages 22 and 25.)
- [35] X. Kang, Y. Li, Z. Qu, and J. Huang, “Enhancing source camera identification performance with a camera reference phase sensor pattern noise,” *IEEE Transactions on Information Forensics and Security*, vol. 7, no. 2, pp. 393–402, Apr. 2012. (Cited on pages 22 and 25.)

- [36] C.-T. Li and Y. Li, "Color-decoupled photo response non-uniformity for digital image forensics," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 22, no. 2, pp. 260–271, Feb. 2012. (Cited on pages 22 and 25.)
- [37] F. Devernay and O. Faugeras, "Automatic calibration and removal of distortion from scenes of structured environments," in *SPIE Investigate and Trial Image Processing*, vol. 2567, 1995, pp. 62–67. (Cited on page 23.)
- [38] P. D. Kovesi, "Matlab and octave functions for computer vision and image processing. Available at <http://www.csse.uwa.edu.au/~pk/research/matlabfns>." (Cited on page 23.)
- [39] K. Kurosawa, K. Kuroki, and N. Saitoh, "CCD fingerprint method-identification of a video camera from videotaped images," in *IEEE International Conference on Image Processing*, vol. 3, Oct. 1999, pp. 537–540. (Cited on page 23.)
- [40] C. R. Holt, "Two-channel detectors for arbitrary linear channel distortion," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 35, no. 3, pp. 267–273, Mar. 1987. (Cited on page 24.)
- [41] S. Bayram, H. Sencar, N. Memon, and I. Avcibas, "Source camera identification based on cfa interpolation," in *IEEE International Conference on Image Processing*, vol. 3, Sep. 2005, pp. 69–72. (Cited on page 26.)
- [42] H. Farid and S. Lyu, "Higher-order wavelet statistics and their application to digital forensics," in *IEEE Conference on Computer Vision and Pattern Recognition*, vol. 8, Jun. 2003. (Cited on page 27.)
- [43] S. Bayram, I. Avcibas, B. Sankur, and N. Menon, "Image manipulation detection with binary similarity measures," in *European Signal Processing Conference*, 2005, pp. 752–755. (Cited on page 27.)
- [44] —, "Image manipulation detection with binary similarity measures," *Journal of Electronic Imaging*, vol. 15, no. 4, 2006. (Cited on page 27.)
- [45] S. Lyu and H. Farid, "Steganalysis using higher-order image statistics," *IEEE Transactions on Information Forensics and Security*, vol. 1, no. 1, pp. 111–119, Mar. 2006. (Cited on pages 27 and 37.)
- [46] M. K. Johnson and H. Farid, "Exposing digital forgeries through chromatic aberration," in *ACM Workshop on Multimedia and Security*, 2006, pp. 48–55. (Cited on page 27.)
- [47] A. C. Popescu and H. Farid, "Exposing digital forgeries in color filter array interpolated images," *IEEE Transactions on Signal Processing*, vol. 53, no. 10, pp. 3948–3959, Oct. 2005. (Cited on page 27.)

- [48] A. C. Gallagher and T. Chen, "Image authentication by detecting traces of demosaicing," in *IEEE Conference on Computer Vision and Pattern Recognition*, Jun. 2008, pp. 1–8. (Cited on page 27.)
- [49] A. E. Dirik and N. Memon, "Image tamper detection based on demosaicing artifacts," in *IEEE International Conference on Image Processing*, Nov. 2009, pp. 1497–1500. (Cited on page 27.)
- [50] H. Farid, "Blind inverse gamma correction," *IEEE Transactions on Image Processing*, vol. 10, no. 10, pp. 1428–1433, Oct. 2001. (Cited on pages 27, 44, 59 and 64.)
- [51] A. C. Popescu and H. Farid, "Statistical tools for digital forensics," in *Information Hiding*, 2004, pp. 128–147. (Cited on pages 27 and 28.)
- [52] Y.-F. Hsu and S.-F. Chang, "Image splicing detection using camera response function consistency and automatic segmentation," in *IEEE International Conference on Multimedia and Expo*, Jul. 2007, pp. 28–31. (Cited on page 27.)
- [53] Z. Lin, R. Wang, X. Tang, and H.-V. Shum, "Detecting doctored images using camera response normality and consistency," in *IEEE Conference on Computer Vision and Pattern Recognition*, Jun. 2005, pp. 1087–1092. (Cited on page 27.)
- [54] Y.-F. Hsu and S.-F. Chang, "Camera response function for image forensics: an automatic algorithm for splicing detection," *IEEE Transactions on Information Forensics and Security*, vol. 5, no. 4, pp. 816–825, Dec. 2010. (Cited on page 27.)
- [55] Z. Fan and R. Queroz, "Identification of bitmap compression history: JPEG detection and quantizer estimation," *IEEE Transactions on Image Processing*, vol. 12, no. 2, pp. 230–235, Feb. 2003. (Cited on pages 28 and 178.)
- [56] R. Neelamani, R. Queroz, Z. Fan, S. Dash, and R. Baraniuk, "JPEG compression history estimation for color images," *IEEE Transactions on Image Processing*, vol. 15, no. 6, pp. 1365–1378, Jun. 2006. (Cited on pages 28 and 178.)
- [57] W. S. Lin, S. K. Tjoa, H. V. Zhao, and K. J. R. Liu, "Digital image source coder forensics via intrinsic fingerprints," *IEEE Transactions on Information Forensics and Security*, vol. 4, no. 3, pp. 460–475, Sep. 2009. (Cited on pages 28 and 178.)
- [58] J. Lukas and J. Fridrich, "Estimation of primary quantization matrix in double compressed JPEG images," in *Digital Forensic Research Workshop*, Aug. 2003, pp. 5–8. (Cited on page 28.)
- [59] T. Pevny and J. Fridrich, "Detection of double compression in JPEG images for applications in steganography," *IEEE Transactions on Information Forensics and Security*, vol. 3, no. 2, pp. 247–258, Jun. 2008. (Cited on page 28.)

- [60] D. Fu, Y. Q. Shi, and W. Su, "A generalized Benford's law for JPEG coefficients and its applications in image forensics," in *SPIE Security, Steganography and Watermarking of Multimedia Contents*, vol. 6505, Jan. 2007, pp. 1–11. (Cited on page 28.)
- [61] B. Li, Y. Q. Shi, and J. Huang, "Detecting doubly compressed JPEG images by using mode based first digit features," in *IEEE International Workshop on Multimedia Signal Processing*, Oct. 2008, pp. 730–735. (Cited on page 28.)
- [62] 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, Nov. 2009. (Cited on pages 28 and 178.)
- [63] Y.-L. Chen and C.-T. Hsu, "Detecting recompression of JPEG images via periodicity analysis of compression artifacts for tampering detection," *IEEE Transactions on Information Forensics and Security*, vol. 6, no. 2, pp. 396–406, Jun. 2011. (Cited on pages 28 and 178.)
- [64] T. Bianchi and A. Piva, "Detection of non aligned double JPEG compression based on integer periodicity maps," *IEEE Transactions on Information Forensics and Security*, vol. 7, no. 2, pp. 842–848, Apr. 2012. (Cited on pages 28 and 178.)
- [65] —, "Image forgery localization via block-grained analysis of JPEG artifacts," *IEEE Transactions on Information Forensics and Security*, vol. 7, no. 3, pp. 1003–1017, Jun. 2012. (Cited on pages 28 and 178.)
- [66] 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, Sep. 2009. (Cited on page 28.)
- [67] M. Barni, A. Costanzo, and L. Sabatini, "Identification of cut and paste tampering by means of double-JPEG detection and image segmentation," in *IEEE International Symposium on Circuits and Systems*, Jun. 2010, pp. 1687–1690. (Cited on page 28.)
- [68] A. C. Popescu and H. Farid, "Exposing digital forgeries by detecting traces of resampling," *IEEE Transactions on Signal Processing*, vol. 53, no. 2, pp. 758–767, Feb. 2005. (Cited on page 28.)
- [69] M. Kirchner, "Fast and reliable resampling detection by spectral analysis of fixed linear predictor residue," in *ACM Workshop on Multimedia and Security*, 2008, pp. 11–20. (Cited on page 28.)
- [70] A. C. Gallagher, "Detection of linear and cubic interpolation in jpeg compressed images," in *Canadian Conference on Computer and Robot Vision*, May 2005, pp. 65–72. (Cited on page 28.)

- [71] B. Mahdian and S. Saic, “Blind authentication using periodic properties of interpolation,” *IEEE Transactions on Information Forensics and Security*, vol. 3, no. 3, pp. 529–538, Sep. 2008. (Cited on page 28.)
- [72] M. C. Stamm and K. J. R. Liu, “Forensic detection of image manipulation using statistical intrinsic fingerprints,” *IEEE Transactions on Information Forensics and Security*, vol. 5, no. 3, pp. 492–506, Sep. 2010. (Cited on page 28.)
- [73] A. Bovik, “Streaking in median filtered images,” *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 35, no. 4, pp. 493–503, Apr. 1987. (Cited on page 28.)
- [74] M. Kirchner and J. Fridrich, “On detection of median filtering in digital images,” in *SPIE Security, Steganography and Watermarking of Multimedia Contents*, vol. 7541, Feb. 2010, pp. 1–6. (Cited on page 28.)
- [75] G. Cao, Y. Zhao, R. Ni, L. Yu, and H. Tian, “Forensic detection of median filtering in digital images,” in *IEEE International Conference on Multimedia and Expo*, Jul. 2010, pp. 89–94. (Cited on page 28.)
- [76] T. Pevny, P. Bas, and J. Fridrich, “Steganalysis by subtractive pixel adjacency matrix,” *IEEE Transactions on Information Forensics and Security*, vol. 5, no. 2, pp. 215–224, Jun. 2010. (Cited on pages 28 and 37.)
- [77] H.-D. Yuan, “Blind forensics of median filtering in digital images,” *IEEE Transactions on Information Forensics and Security*, vol. 6, no. 4, pp. 1335–1345, Dec. 2011. (Cited on page 28.)
- [78] T.-T. Ng and S.-F. Chang, “A model for image splicing,” in *IEEE International Conference on Image Processing*, Sep. 2004, pp. 1169–1172. (Cited on page 28.)
- [79] M. K. Johnson and H. Farid, “Exposing digital forgeries by detecting inconsistencies in lighting,” in *ACM Workshop on Multimedia and Security*, 2005, pp. 1–10. (Cited on pages 28 and 29.)
- [80] —, “Exposing digital forgeries through specular highlights on the eye,” in *Information Hiding*, vol. 4567, 2007, pp. 311–325. (Cited on page 28.)
- [81] —, “Exposing digital forgeries in complex lighting environments,” *IEEE Transactions on Information Forensics and Security*, vol. 2, no. 3, pp. 450–461, Sep. 2007. (Cited on page 29.)
- [82] G. J. Simmons, “Prisoners’ problem and the subliminal channel,” in *Proceedings of CRYPTO*, 1984, pp. 51–67. (Cited on page 29.)
- [83] A. Kerckhoffs, “La cryptographie militaire,” *Journal des Sciences Militaires*, pp. 161–191, 1883. (Cited on page 30.)

- [84] J. Fridrich, M. Goljan, and D. Soukal, "Perturbed quantization steganography using wet paper codes," in *ACM Workshop on Multimedia and Security*, Sep. 2004, pp. 4–15. (Cited on page 31.)
- [85] L. Marvel, J. Boncelet, and C. Retter, "Spread spectrum image steganography," *IEEE Transactions on Image Processing*, vol. 8, no. 8, pp. 1075–1083, Aug. 1999. (Cited on page 31.)
- [86] B. Chen and G. Wornell, "Quantization index modulation: a class of provably good methods for digital watermarking and information embedding," *IEEE Transactions on Information Theory*, vol. 47, no. 4, pp. 1423–1443, May 2001. (Cited on page 31.)
- [87] P. Sallee, "Model-based steganography," in *Digital Watermarking*, vol. 2939, 2004, pp. 154–167. (Cited on pages 31 and 168.)
- [88] N. Provos, "Defending against statistical steganalysis," in *USENIX Security Symposium*, Aug. 2001, pp. 323–335. (Cited on page 31.)
- [89] Y. Kim, Z. Duric, and D. Richards, "Modified matrix encoding technique for minimal distortion steganography," in *Information Hiding*, vol. 4437, Jul. 2006, pp. 314–327. (Cited on page 31.)
- [90] T. Pevny, T. Filler, and P. Bas, "Using high-dimensional image models to perform highly undetectable steganography," in *Information Hiding*, vol. 6387, Jun. 2010, pp. 161–177. (Cited on page 31.)
- [91] T. Sharp, "An implementation of key-based digital signal steganography," in *Information Hiding*, vol. 2137, Jun. 2001, pp. 13–26. (Cited on page 31.)
- [92] D. Upham, "Steganographic algorithm Jsteg [online]. Available at <http://zoooid.org/paul/crypto/jsteg>," 1993. (Cited on page 31.)
- [93] S. Katzenbeisser and F. Peticolas, *Information Hiding Techniques for Steganography and Digital Watermarking*. London: Artech house, 1999. (Cited on page 31.)
- [94] I. Cox, M. L. Miller, J. A. Bloom, J. Fridrich, and T. Kalker, *Digital Watermarking and Steganography*. San Francisco, CA: Morgan Kaufmann, 2007. (Cited on page 31.)
- [95] R. Böhme, *Advanced Statistical Steganalysis*. New York: Springer, 2010. (Cited on pages 31, 34, 35 and 169.)
- [96] J. Fridrich, M. Goljan, and R. Du, "Detecting LSB steganography in color and gray-scale images," *IEEE Multimedia*, vol. 8, no. 4, pp. 22–28, 2001. (Cited on pages 33 and 38.)

- [97] S. Dumitrescu, X. Wu, and Z. Wang, "Detection of LSB steganography via sample pair analysis," *IEEE Transactions on Signal Processing*, vol. 51, no. 7, pp. 1995–2007, Jul. 2003. (Cited on page 33.)
- [98] P. Lu, X. Luo, Q. Tang, and L. Shen, "An improved sample pairs method for detection of LSB embedding," in *Information Hiding*, vol. 3200, May 2004, pp. 116–127. (Cited on pages 33 and 35.)
- [99] A. D. Ker, "A fusion of maximum likelihood and structural steganalysis," in *Information Hiding*, vol. 4567, Jun. 2007, pp. 204–219. (Cited on pages 33, 35, 38 and 168.)
- [100] —, "A general framework for the structural steganalysis of LSB replacement," in *Information Hiding*, vol. 3727, Jun. 2005, pp. 296–311. (Cited on pages 33 and 35.)
- [101] —, "Fourth-order structural steganalysis and analysis of cover assumptions," in *SPIE Security, Steganography and Watermarking of Multimedia Contents*, vol. 6072, Feb. 2006, pp. 25–38. (Cited on pages 33, 34 and 35.)
- [102] J. Fridrich and M. Goljan, "On estimation of secret message length in LSB steganography in spatial domain," in *SPIE Security, Steganography and Watermarking of Multimedia Contents*, vol. 5306, Jun. 2004, pp. 23–34. (Cited on pages 35 and 37.)
- [103] A. D. Ker and R. Böhme, "Revisiting weighted stego-image steganalysis," in *SPIE Security, Steganography and Watermarking of Multimedia Contents*, May 2008, pp. 501–517. (Cited on pages 35, 37, 151 and 169.)
- [104] A. D. Ker, "Optimally weighted least squares steganalysis," in *SPIE Security, Steganography and Watermarking of Multimedia Contents*, vol. 6505, Feb. 2007, pp. 06–16. (Cited on pages 35, 37, 38 and 169.)
- [105] A. Westfeld and P. Andreas, "Attacks on steganographic systems," in *Information Hiding*, Oct. 1999, pp. 61–76. (Cited on page 36.)
- [106] R. Chandramouli and N. Memon, "Analysis of LSB based image steganography techniques," in *IEEE International Conference on Image Processing*, Oct. 2001, pp. 1019–1022. (Cited on page 36.)
- [107] O. Dabeer, K. Sullivan, U. Madhow, S. Chandrasekaran, and B. Manjunath, "Detection of hiding in the least significant bit," *IEEE Transactions on Signal Processing*, vol. 52, no. 10, pp. 3046–3058, Oct. 2004. (Cited on pages 36, 37, 41 and 140.)
- [108] W. Hoeffding, "Asymptotically optimal tests for multinomial distributions," *The Annals of Mathematical Statistics*, vol. 36, no. 2, pp. 369–401, Apr. 1965. (Cited on page 37.)

- [109] L. Fillatre, “Adaptive steganalysis of least significant bit replacement in grayscale natural images,” *IEEE Transactions on Signal Processing*, vol. 60, no. 2, pp. 556–569, Feb. 2012. (Cited on pages 37, 140, 151, 152, 154 and 155.)
- [110] R. Cogranne, C. Zitzmann, L. Fillatre, F. Retraint, I. Nikiforov, and P. Cornu, “Statistical decision by using quantized observations,” in *IEEE International Symposium on Information Theory*, Aug. 2011, pp. 1210–1214. (Cited on pages 37, 140, 141 and 155.)
- [111] R. Cogranne, C. Zitzmann, L. Fillatre, I. Nikiforov, F. Retraint, and P. Cornu, “A cover image model for reliable steganalysis,” in *Information Hiding*, vol. 6958, May 2011, pp. 178–192. (Cited on pages 37, 140 and 155.)
- [112] —, “Reliable detection of hidden information based on a non-linear local model,” in *IEEE Workshop on Statistical Signal Processing*, Jun. 2011, pp. 493–496. (Cited on pages 37, 140 and 155.)
- [113] C. Zitzmann, R. Cogranne, F. Retraint, I. Nikiforov, L. Fillatre, and P. Cornu, “Statistical decision methods in hidden information detection,” in *Information Hiding*, vol. 6958, May 2011, pp. 163–177. (Cited on pages 37, 140, 143 and 155.)
- [114] 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, Jan. 2014. (Cited on pages 37 and 140.)
- [115] R. Cogranne and F. Retraint, “Application of hypothesis testing theory for optimal detection of LSB matching data hiding,” *Signal Processing*, vol. 93, no. 7, pp. 1724–1737, Jul. 2013. (Cited on pages 37 and 180.)
- [116] —, “An asymptotically uniformly most powerful test for LSB matching detection,” *IEEE Transactions on Information Forensics and Security*, vol. 8, no. 3, pp. 464–476, Mar. 2013. (Cited on pages 37 and 180.)
- [117] Y. Wang and P. Moulin, “Optimized feature extraction for learning-based image steganalysis,” *IEEE Transactions on Information Forensics and Security*, vol. 2, no. 1, pp. 31–45, Mar. 2007. (Cited on page 37.)
- [118] J. Kodovsky, J. Fridrich, and V. Holub, “Ensemble classifiers for steganalysis of digital media,” *IEEE Transactions on Information Forensics and Security*, vol. 7, no. 2, pp. 432–444, Apr. 2012. (Cited on page 37.)
- [119] T. Pevny, J. Fridrich, and A. D. Ker, “From blind to quantitative steganalysis,” *IEEE Transactions on Information Forensics and Security*, vol. 7, no. 2, pp. 445–454, Apr. 2012. (Cited on pages 37 and 171.)

- [120] J. Fridrich and J. Kodovsky, "Rich models for steganalysis of digital images," *IEEE Transactions on Information Forensics and Security*, vol. 7, no. 3, pp. 868–882, Jun. 2012. (Cited on page 37.)
- [121] A. D. Ker, P. Bas, R. Böhme, R. Cogranne, S. Craver, T. Filler, J. Fridrich, and T. Pevny, "Moving steganography and steganalysis from the laboratory into the real world," in *ACM Workshop on Information Hiding and Multimedia Security*, Jun. 2013, pp. 45–58. (Cited on page 38.)
- [122] T. Zhang and X. Ping, "A fast and effective steganalytic technique against jsteg-like algorithms," in *ACM Symposium on Applied Computing*, 2003, pp. 307–311. (Cited on pages 38, 163 and 171.)
- [123] X. Yu, Y. Wang, and T. Tan, "Model based steganalysis," in *IEEE International Conference on Image Processing*, vol. 4, Oct. 2004, pp. 2625–2628. (Cited on pages 38 and 163.)
- [124] K. Lee, A. Westfeld, and S. Lee, "Category attack for LSB embedding of JPEG images," in *Digital Watermarking*, vol. 4283, Nov. 2006, pp. 35–48. (Cited on pages 38 and 163.)
- [125] K. Lee and A. Westfeld, "Generalized category attack: Improving histogram-based attack on JPEG LSB embedding," in *Information Hiding*, vol. 4567, Jun. 2007, pp. 378–392. (Cited on pages 38 and 163.)
- [126] J. Kodovsky and J. Fridrich, "Quantitative structural steganalysis of Jsteg," *IEEE Transactions on Information Forensics and Security*, vol. 5, no. 4, pp. 681–693, Dec. 2010. (Cited on pages 38, 46, 73, 163, 168 and 171.)
- [127] A. Westfeld, "Generic adoption of spatial steganalysis to transformed domain," in *Information Hiding*, vol. 5284, May 2007, pp. 161–177. (Cited on pages 39 and 163.)
- [128] R. Böhme, "Weighted stego-image steganalysis for JPEG covers," in *Information Hiding*, vol. 5284. Springer, May 2007, pp. 178–194. (Cited on pages 39, 163, 164, 169 and 171.)
- [129] C. Zitzmann, R. Cogranne, L. Fillatre, I. Nikiforov, F. Retraint, and P. Cornu, "Hidden information detection based on quantized laplacian distribution," in *IEEE International Conference on Acoustics, Speech and Signal Processing*, Mar. 2012, pp. 1793–1796. (Cited on pages 39, 41, 163 and 171.)
- [130] D. T. Kuan, A. A. Sawchuk, T. C. Strand, and P. Chavel, "Adaptive noise smoothing filter for images with signal-dependent noise," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 7, no. 2, pp. 165–177, Mar. 1985. (Cited on pages 42 and 61.)

- [131] F. Luisier, T. Blu, and M. Unser, “Image denoising in mixed Poisson-Gaussian noise,” *IEEE Transactions on Image Processing*, vol. 20, no. 3, pp. 696–708, Mar. 2011. (Cited on pages 42 and 43.)
- [132] M. Makitalo and A. Foi, “Optimal inversion of the generalized Anscombe transformation for Poisson-Gaussian noise,” *IEEE Transactions on Image Processing*, vol. 22, no. 1, pp. 91–103, Jan. 2013. (Cited on pages 42 and 43.)
- [133] A. Foi, M. Trimeche, V. Katkovnik, and K. Egiazarian, “Practical Poissonian-Gaussian noise modeling and fitting for single-image raw-data,” *IEEE Transactions on Image Processing*, vol. 17, no. 10, pp. 1737–1754, Oct. 2008. (Cited on pages 42, 43, 51, 52, 53, 55, 60, 83, 89, 143 and 153.)
- [134] T. H. Thai, R. Cogranne, and F. Retraint, “Camera model identification based on the heteroscedastic noise model,” *IEEE Transactions on Image Processing*, vol. 23, no. 1, pp. 250–263, Jan. 2014. (Cited on pages 42, 43, 50, 51, 52, 73, 78 and 99.)
- [135] H. Faraji and W. J. MacLean, “CCD noise removal in digital images,” *IEEE Transactions on Image Processing*, vol. 15, no. 9, pp. 2676–2685, Sep. 2006. (Cited on pages 42, 44, 52 and 59.)
- [136] G. E. Healey and R. Kondepudy, “Radiometric CCD camera calibration and noise estimation,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 16, no. 3, pp. 267–276, Mar. 1994. (Cited on pages 42, 43 and 51.)
- [137] M. D. Grossberg and S. K. Nayar, “What can be known about the radiometric response from images ?” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 26, no. 10, pp. 1272–1282, Oct. 2004. (Cited on page 44.)
- [138] T.-T. Ng, S.-F. Chang, and M.-P. Tsui, “Using geometry invariants for camera response function estimation,” in *IEEE Conference on Computer Vision and Pattern Recognition*, Jun. 2007, pp. 1–8. (Cited on page 44.)
- [139] C. Liu, R. Szeliski, S. B. K., C. L. Zitnick, and W. T. Freeman, “Automatic estimation and removal of noise from a single image,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 30, no. 2, pp. 299–314, Feb. 2008. (Cited on pages 44 and 52.)
- [140] S. J. Kim, H. T. Lin, Z. Lu, S. Su, S. Lin, and M. S. Brown, “A new in-camera imaging model for color computer vision and its application,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 12, pp. 2289–2302, Dec. 2012. (Cited on page 44.)
- [141] W. K. Pratt, *Digital Image Processing*. New york: Wiley, 1978. (Cited on page 45.)

- [142] R. Reininger and J. Gibson, "Distributions of the two-dimensional DCT coefficients for images," *IEEE Transaction on Communications*, vol. 31, no. 6, pp. 835–839, Jun. 1983. (Cited on page 45.)
- [143] T. Eude, R. Grisel, H. Cherifi, and R. Debie, "On the distribution of the DCT coefficients," in *IEEE International Conference on Image Processing*, Apr. 1994, pp. 365–368. (Cited on page 45.)
- [144] J. E. Eggerton and M. D. Srinath, "Statistical distributions of image DCT coefficients," *Computers and Electrical Engineering*, vol. 12, no. 3-4, pp. 137–145, Jan. 1986. (Cited on page 45.)
- [145] F. Muller, "Distribution shape of two-dimensional DCT coefficients of natural images," *Electronics Letters*, vol. 29, no. 22, pp. 1935–1936, Oct. 1993. (Cited on page 45.)
- [146] J.-H. Chang, J.-W. Shin, N. S. Kim, and S. K. Mitra, "Image probability distribution based on generalized gamma function," *IEEE Signal Processing Letters*, vol. 12, no. 4, pp. 325–328, Apr. 2005. (Cited on pages 45, 71 and 134.)
- [147] E. Y. Lam and J. W. Goodman, "A mathematical analysis of the DCT coefficient distributions for images," *IEEE Transactions on Image Processing*, vol. 9, no. 10, pp. 1661–1666, Oct. 2000. (Cited on pages 45, 46 and 67.)
- [148] A. Hjørungnes, J. M. Lervik, and T. A. Ramstad, "Entropy coding of composite sources modeled by infinite gaussian mixture distributions," in *IEEE Digital Signal Processing Workshop*, 1996, pp. 235–238. (Cited on page 45.)
- [149] F. Perez-Gonzalez, G. L. Heileman, and C. T. Abdallah, "Benford's law in image processing," in *IEEE International Conference on Image Processing*, vol. 1, 2007, pp. 405–408. (Cited on page 45.)
- [150] M. Blum, "On the central limit theorem for correlated random variables," *Proceedings of the IEEE*, vol. 52, no. 3, pp. 308–309, Mar. 1964. (Cited on pages 46 and 65.)
- [151] C. Tu and T. D. Tran, "Context-based entropy coding of block transform coefficients for image compression," *IEEE Transactions on Image Processing*, vol. 11, no. 11, pp. 1271–1283, Nov. 2002. (Cited on pages 46 and 72.)
- [152] J. Fridrich, "Feature-based steganalysis for JPEG images and its implications for future design of steganographic schemes," in *Information Hiding*, vol. 3200, May 2004, pp. 67–81. (Cited on pages 46 and 73.)
- [153] T. H. Thai, F. Retraint, and R. Cogranne, "Generalized signal-dependent noise model and parameter estimation for natural images," *submitted to IEEE Signal Processing Letters*, 2014. (Cited on pages 50 and 73.)

- [154] T. H. Thai, R. Cogranne, and F. Retraint, "Statistical model of quantized DCT coefficients: application in the steganalysis of jsteg algorithm," *IEEE Transactions on Image Processing*, vol. 23, no. 5, pp. 1980–1993, May 2014. (Cited on pages 50, 73 and 164.)
- [155] A. Foi, "Clipped noisy images : Heteroskedastic modeling and practical denoising," *Signal Processing*, vol. 89, no. 12, pp. 2609–2629, Dec. 2009. (Cited on pages 51, 140, 142, 144, 148 and 153.)
- [156] X. Liu, M. Tanaka, and M. Tokunomi, "Estimation of signal dependent noise parameters from a single image," in *IEEE International Conference on Image Processing*, Sep. 2013, pp. 79–82. (Cited on pages 52 and 59.)
- [157] J. A. Nelder and R. Mead, "A simplex method for function minimization," *The Computer Journal*, vol. 7, pp. 308–313, 1965. (Cited on pages 52, 61, 69, 160, 169 and 190.)
- [158] R. J. Carroll, "Adapting for heteroscedasticity in linear models," *The Annals of Statistics*, vol. 10, no. 4, pp. 1224–1233, Dec. 1982. (Cited on pages 52, 54 and 86.)
- [159] J. Shao, "Asymptotic distribution of the weighted least squares estimator," *Annals of the Institute of Statistical Mathematics*, vol. 41, no. 2, pp. 365–382, 1989. (Cited on pages 52, 54 and 86.)
- [160] N. L. Johnson, S. Kotz, and N. Balakrishnan, *Continuous univariate distributions*, 2nd ed. New york: Wiley, 1994. (Cited on pages 53, 134, 142 and 160.)
- [161] C. R. Rao and H. Toutenburg, *Linear models : Least Squares and Alternatives*, 2nd ed. New york: Springer, 1999. (Cited on pages 54, 118, 119 and 124.)
- [162] 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, Apr. 1996. (Cited on pages 58 and 70.)
- [163] R. P. Kleihorst, R. L. Lagendiik, and J. Biemond, "An adaptive order-statistic noise filter for gamma-corrected image sequences," *IEEE Transactions on Image Processing*, vol. 6, no. 10, pp. 1442–1446, Oct. 1997. (Cited on page 59.)
- [164] P. J. Rousseeuw and C. Croux, "Alternatives to the median absolute deviation," *J. Amer. Statist. Assoc.*, vol. 88, no. 424, pp. 1273–1283, Dec. 1993. (Cited on page 60.)
- [165] M. K. Mihçak, I. Kozintsev, and K. Ramchandran, "Spatially adaptive statistical modeling of wavelet image coefficients and its application to denoising," in *IEEE International Conference on Acoustics, Speech and Signal Processing*, vol. 6, Mar. 1999, pp. 3253–3256. (Cited on pages 62 and 126.)

- [166] M. A. Robertson and R. L. Stevenson, “DCT quantization noise in compressed images,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 15, no. 1, pp. 27–38, Jan. 2005. (Cited on pages 62 and 101.)
- [167] N. Ponomarenko, V. Lukin, A. Zelensky, K. Egiazarian, M. Carli, and F. Battisti, “TID2008 - a database for evaluation of full-reference visual quality assessment metrics,” *Advances Modern Radioelectronics*, vol. 10, pp. 30–45, 2009. (Cited on page 63.)
- [168] T. Gloe and R. Bohme, “The Dresden image database for benchmarking digital image forensics,” in *ACM Symposium on Applied Computing*, vol. 2, 2010, pp. 1585–1591. (Cited on pages 63, 72, 89, 91, 92, 108, 110, 127, 128, 130 and 151.)
- [169] M.-S. Alouini, A. Abdi, and M. Kaveh, “Sum of gamma variates and performance of wireless communication systems over nakagami-fading channels,” *IEEE Transactions on Vehicular Technology*, vol. 50, no. 6, pp. 1471–1480, Nov. 2001. (Cited on page 66.)
- [170] I. M. Ryzhik and I. S. Gradshteyn, *Tables of Integrals, Series, and Products*. United Kingdom: Elsevier, 2007. (Cited on pages 67, 68 and 122.)
- [171] P. Bas, T. Filler, and T. Pevný, “Break our steganographic system — the ins and outs of organizing boss,” in *Information Hiding*, May 2011. (Cited on pages 67, 91, 92, 131, 154 and 169.)
- [172] J. Fan and I. Gijbels, *Local Polynomial Modelling and Its Applications*. Chapman & Hall, 1996. (Cited on page 72.)
- [173] X. Pan, X. Zhang, and S. Lyu, “Exposing image forgery with blind noise estimation,” in *Proceedings of ACM Workshop on Multimedia and Security*, Sep. 2011, pp. 15–20. (Cited on page 78.)
- [174] —, “Exposing image splicing with inconsistent local noise variances,” in *IEEE International Conference on Computational Photography*, Apr. 2012, pp. 1–10. (Cited on page 78.)
- [175] F. R. Hampel, “The influence curve and its role in robust estimation,” *Journal of the American Statistical Association*, no. 69, pp. 382–393, Jun. 1974. (Cited on pages 90 and 109.)
- [176] T. H. Thai, F. Retraint, and R. Cogranne, “Camera model identification based on the generalized noise model in natural images,” *submitted to IEEE Transactions on Image Processing*, 2014. (Cited on page 100.)
- [177] T. H. Thai, R. Cogranne, and F. Retraint, “Système d’identification d’un modèle d’appareil photographique associé à une image compressée au format jpeg, procédé, utilisations and applications associés,” PS/B52545/FR, 2014. (Cited on page 100.)

- [178] —, “Camera model identification based on DCT coefficients statistics,” *submitted to IEEE Transactions on Information Forensics and Security*, 2014. (Cited on page 116.)
- [179] —, “Système d’identification d’un modèle d’appareil photographique associé à une image compressée au format jpeg, procédé, utilisations and applications associés,” PS/B52546/FR, 2014. (Cited on page 116.)
- [180] L. Tierney and J. B. Kadane, “Accurate approximations for posterior moments and marginal densities,” *Journal of American Statistical Association*, vol. 81, no. 393, pp. 82–86, Mar. 1986. (Cited on pages 121 and 133.)
- [181] R. Butler and A. Wood, “Laplace approximations for hypergeometric functions with matrix argument,” *The Annals of Statistics*, vol. 30, no. 4, pp. 1155–1177, 2002. (Cited on pages 121 and 133.)
- [182] T. H. Thai, F. Retraint, and R. Cogranne, “Statistical detection of data hidden in least significant bits of clipped images,” *Signal Processing*, vol. 98, pp. 263–274, May 2014. (Cited on page 140.)
- [183] L. Le Cam, *Asymptotics Methods in Statistical Decision Theory*. New york: Series in Statistics, Springer, 1986. (Cited on page 143.)
- [184] A. Foi, V. Katkovnik, and K. Egiazarian, “Signal-dependent noise removal in pointwise shape-adaptive DCT domain with locally adaptive variance,” in *European Signal Processing Conference*, Sep. 2007, pp. 2159–2163. (Cited on page 144.)
- [185] K. Hirakawa and T. W. Parks, “Image denoising for signal-dependent noise,” in *IEEE International Conference on Acoustics, Speech and Signal Processing*, vol. 2, Mar. 2005, pp. 29–32. (Cited on page 144.)
- [186] P. R. Prucnal and B. Saleh, “Transformation of image-signal-dependent noise into image-signal-independent noise,” *Optics Letters*, vol. 6, no. 7, pp. 316–318, Jul. 1981. (Cited on pages 144 and 155.)
- [187] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, “Image denoising by sparse 3-d transform-domain collaborative filtering,” *IEEE Transactions on Image Processing*, vol. 16, no. 8, pp. 2080–2095, Aug. 2007. (Cited on page 144.)
- [188] J. Portilla, V. Strela, M. J. Wainwright, and E. P. Simoncelli, “Image denoising using scale mixtures of Gaussians in the wavelet domain,” *IEEE Transactions on Image Processing*, vol. 12, no. 11, pp. 1338–1351, Nov. 2003. (Cited on page 144.)
- [189] A. Stuart and J. K. Ord, *Kendall’s Advanced Theory of Statistics*, 6th ed. London, UK: Arnold, 1994, vol. 1. (Cited on pages 160, 186 and 190.)

- [190] T. H. Thai, R. Cogranne, and F. Retraint, “Optimal detection of outguess using an accurate model of dct coefficients,” 2014. (Cited on page 164.)
- [191] A. Wald, “Some generalizations of the theory of cumulative sums of random variable,” *The Annals of Mathematical Statistics*, pp. 287–293, Sep. 1945. (Cited on page 167.)
- [192] 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 *to be published in Proceedings of ACM Information Hiding and Multimedia Security Workshop*, Jun. 2014. (Cited on page 174.)
- [193] P. Bestagini, M. Fontani, S. Milani, M. Barni, A. Piva, M. Tagliasacchi, and S. Tubaro, “An overview on video forensics,” in *European Signal Processing Conference*, Aug. 2012, pp. 1229–1233. (Cited on page 179.)
- [194] T. Gloe, M. Kirchner, A. Winkler, and R. Böhme, “Can we trust digital image forensics?” in *ACM International Conference on Multimedia*, 2007, pp. 78–86. (Cited on page 179.)
- [195] M. Kirchner and R. Böhme, “Hiding traces of resampling in digital images,” *IEEE Transactions on Information Forensics and Security*, vol. 3, no. 4, pp. 582–592, Dec. 2008. (Cited on page 179.)
- [196] M. C. Stamm and K. J. R. Liu, “Anti-forensics of digital image compression,” *IEEE Transactions on Information Forensics and Security*, vol. 6, no. 3, pp. 1050–1065, Sep. 2011. (Cited on page 179.)
- [197] M. Goljan, J. Fridrich, and M. Chen, “Defending against fingerprint-copy attack in sensor-based camera identification,” *IEEE Transactions on Information Forensics and Security*, vol. 6, no. 1, pp. 227–236, Mar. 2011. (Cited on page 179.)
- [198] S. Lai and R. Böhme, “Countering counter-forensics: the case of JPEG compression,” in *Information Hiding*, vol. 6958, May 2011, pp. 285–298. (Cited on page 179.)
- [199] G. Valenzise, M. Tagliasacchi, and S. Tubaro, “Revealing the traces of JPEG compression anti-forensics,” *IEEE Transactions on Information Forensics and Security*, vol. 8, no. 2, pp. 335–349, Feb. 2013. (Cited on page 179.)
- [200] M. C. Stamm, W. S. Lin, and K. J. R. Liu, “Forensics vs. anti-forensics: A decision and game theoretic framework,” in *IEEE International Conference on Acoustics, Speech and Signal Processing*, Mar. 2012, pp. 1749–1752. (Cited on page 179.)
- [201] M. Barni and B. Tondi, “The source identification game: An information-theoretic perspective,” *IEEE Transactions on Information Forensics and Security*, vol. 8, no. 3, pp. 450–463, Mar. 2013. (Cited on page 179.)

-
- [202] M. C. Stamm and K. J. R. Liu, “Protection against reverse engineering in digital cameras,” in *IEEE International Conference on Acoustics, Speech and Signal Processing*, May 2013, pp. 8702–8706. (Cited on page 179.)
 - [203] M. Basseville and I. Nikoforov, *Detection of Abrupt Changes: Theory and Application*. Prentice Hall. (Cited on page 180.)
 - [204] A. A. Borovkov, *Mathematical Statistics*. Amsterdam: Gordon and Breach Sciences, 1998. (Cited on page 183.)
 - [205] T. Ferguson, *Mathematical Statistics: A Decision Theoretic Approach*. Academic press, 1967. (Cited on page 183.)
 - [206] D. Birkes, “Generalized likelihood ratio tests and uniformly most powerful tests,” *The American Statistician*, vol. 44, no. 2, pp. 163–166, 1990. (Cited on page 189.)
 - [207] O. Zeitouni, J. Ziv, and N. Merhav, “When is the generalized likelihood ratio test optimal?” *IEEE Transactions on Information Theory*, vol. 38, no. 5, pp. 1597–1602, Sep. 1992. (Cited on page 189.)