ABSTRACT

Title of dissertation: FUNDAMENTAL LIMITS IN MULTIMEDIA FORENSICS AND ANTI-FORENSICS

Xiaoyu Chu, Doctor of Philosophy, 2015

Dissertation directed by: Professor K. J. Ray Liu
Department of Electrical and Computer Engineering

As the use of multimedia editing tools increases, people become questioning the authenticity of multimedia content. This is specially a big concern for authorities, such as law enforcement, news reporter and government, who constantly use multimedia evidence to make critical decisions. To verify the authenticity of multimedia content, many forensic techniques have been proposed to identify the processing history of multimedia content under question. However, as new technologies emerge and more complicated scenarios are considered, the limitation of multimedia forensics has been gradually realized by forensic researchers. It is the inevitable trend in multimedia forensics to explore the fundamental limits. In this dissertation, we propose several theoretical frameworks to study the fundamental limits in various forensic problems.

Specifically, we begin by developing empirical forensic techniques to deal with the limitation of existing techniques due to the emergence of new technology, compressive sensing. Then, we go one step further to explore the fundamental limit

of forensic performance. Two types of forensic problems have been examined. In operation forensics, we propose an information theoretical framework and define forensicability as the maximum information features contain about hypotheses of processing histories. Based on this framework, we have found the maximum number of JPEG compressions one can detect. In order forensics, an information theoretical criterion is proposed to determine when we can and cannot detect the order of manipulation operations that have been applied on multimedia content.

Additionally, we have examined the fundamental tradeoffs in multimedia anti-forensics, where attacking techniques are developed by forgers to conceal manipulation fingerprints and confuse forensic investigations. In this field, we have defined concealability as the effectiveness of anti-forensics concealing manipulation fingerprints. Then, a tradeoff between concealability, rate and distortion is proposed and characterized for compression anti-forensics, which provides us valuable insights of how forgers may behave under their best strategy.

FUNDAMENTAL LIMITS IN MULTIMEDIA FORENSICS AND ANTI-FORENSICS

by

Xiaoyu Chu

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2015

Advisory Committee:
Professor K. J. Ray Liu, Chair/Advisor
Professor Min Wu
Professor Gang Qu
Professor Piya Pal
Professor Lawrence C. Washington

Dedication

To my parents.

Acknowledgments

First of all, I would like to express the deepest gratitude to my advisor, Prof. K. J. Ray Liu. I can still remember the excitement I had when I first received a reply from someone I admire. His approval brought back my confidence in my darkest time and made me feel that I had finally found someone who can really appreciate my talent. At that moment, I decided to myself that I will follow this person no matter what. Over the passed five years, I have had distractions, immaturity and doubt. But he never gave up on me. His consideration, encouragement and guidance have made me accomplish so much that I could never foresee when I first came to this state. Without him, this dissertation would never be possible.

I would like to thank other members of my dissertation committee, Prof. Min Wu, Prof. Qu Gang, Prof. Piya Pal, and Prof. Lawrence Washington, for their precious time and effort serving on my committee. I would like to specially thank Prof. Min Wu for her guidance when I was her teaching assistant and her reference during my job applications.

Thanks should also be given to all members in Signal and Information Group for their friendship and assistance. I would like to specially thank Prof. Matthew Stamm, Dr. Yan Chen, and Dr. Wan-Yi Lin for their enormous help in both research and life. I also wish to give my thanks to Feng Han, Wei Guan, Yang Gao, and Zhuang-Han Wu, with whom I feel like having a family here.

Most of all, I would like to give my greatest appreciation to my parents. They are always there for me and unconditionally provide the strongest support to my

study and research. My persistence and enthusiasm for new challenges are directly due to them. To my dearest parents I dedicate this dissertation.

Table of Contents

List of Tables viii

List of Figures ix

1 Introduction 1
1.1 Motivation . . . 1
1.2 Dissertation Outline . . . 5
1.2.1 Compressive Sensing Forensics (Chapter 2) . . . 5
1.2.2 Fundamental Limits in Operation Forensics (Chapter 3) . . . 6
1.2.3 Fundamental Limits in Order Forensics (Chapter 4) . . . 7
1.2.4 Fundamental Tradeoffs in Compression Anti-forensics (Chapter 5) . . . 8

2 Compressive Sensing Forensics 10
2.1 System Model . . . 14
2.1.1 Compressive Sensing Overview . . . 15
2.1.2 Signal Model . . . 16
2.2 Compressive Sensing Fingerprints . . . 20
2.3 Compressive Sensing Detection . . . 27
2.3.1 Zero Ratio Detection Scheme . . . 27
2.3.2 Distribution-based Detection Scheme . . . 30
2.4 Detecting Compressive Sensing in Digital Images . . . 34
2.4.1 Compressive Sensing Fingerprints in Digital Images . . . 35
2.4.2 DWT Coefficient Distribution Models . . . 39
2.4.3 Compressive Sensing Detection . . . 40
2.5 Measurement Number Estimation . . . 45
2.6 Simulations and Results . . . 48
2.6.1 Sparse Signals in the Presence of Noise . . . 49
2.6.2 Nearly Sparse Signals and Nearly Sparse Signals in the Presence of Noise . . . 54
2.6.3 Images . . . 57
2.6.4 Estimator of the Number of Compressive Measurements . . . 59
2.7 Summary . . . 62

3 Fundamental Limits in Operation Forensics 64
3.1 Information Theoretical Framework . . . 67
3.1.1 Channel between Multimedia States and Features . . . 67
3.1.2 Forensicability . . . 71
3.1.3 Expected Perfect Detection . . . 75
3.2 Information Theoretical Modeling for JPEG Compression Forensics . . . 78
3.2.1 Background on JPEG Compression Forensics . . . 78
3.2.2 DCT Coefficients Feature Model . . . 80
3.2.3 Forensicability for JPEG Compression Forensics . . . 83

3.3 Data-Driven Results and Analysis . . . 87
3.3.1 Verification of Observation Noise Model . . . 89
3.3.2 Forensicability Calculation . . . 91
3.3.3 Estimation Error Probability Lower Bound . . . 93
3.3.4 Maximum Number of Detectable Compressions . . . 98
3.3.5 Quality Factor Patterns having the Highest and Lowest Forensicabilities . . . 100
3.3.6 Optimal Strategies for Forgers and Investigators . . . 104
3.3.7 Forensicabilities for Image Outliers . . . 107
3.4 Summary . . . 111

4 Fundamental Limits in Order Forensics 112
4.1 System Model . . . 115
4.1.1 Order of Operations May Not be Detectable . . . 115
4.1.2 Information Theoretical Model for Multiple Hypotheses Estimation Problems . . . 119
4.2 Information Theoretical Criteria . . . 122
4.2.1 Mutual Information Criterion to Obtain the Best Estimator . 123
4.2.2 Information Theoretical Criteria for Multiple Hypotheses Estimation Problems . . . 127
4.3 Detecting the Order of Resizing and Blurring . . . 132
4.4 Simulation Results . . . 140
4.4.1 Detect Double JPEG Compression . . . 141
4.4.2 Detect the Order of Resizing and Contrast Enhancement . . . 145
4.4.3 Detect the Order of Resizing and Blurring . . . 150
4.5 Summary . . . 153

5 Fundamental Tradeoffs in Compression Anti-forensics 156
5.1 Background . . . 160
5.1.1 JPEG Compression . . . 161
5.1.2 Double JPEG Compression Fingerprints . . . 162
5.1.3 Double JPEG Compression Detection . . . 164
5.1.4 JPEG Compression Anti-Forensics . . . 165
5.2 Concealability-Rate-Distortion Tradeoff . . . 167
5.3 Flexible Anti-Forensic Dither . . . 173
5.4 Anti-Forensic Transcoder . . . 180
5.5 Simulation Results and Analysis . . . 184
5.5.1 Two C-R-D Tradeoffs Revealed From Simulation . . . 185
5.5.2 C-R-D Tradeoff for Lower Secondary Quality Factors . . . 189
5.5.3 C-R-D Tradeoff for Higher Secondary Quality Factors . . . 192
5.6 Summary . . . 196

6 Conclusions and Future Work 198
6.1 Conclusions . . . 198
6.2 Future Work . . . 201

Bibliography 203

List of Tables

2.1 Relative error of estimating compressive measurements for images. . . 62

3.1 $\min_{\mathcal{Q}_M} P_e^0$ for different M. 98
3.2 $\min_{\mathcal{Q}_M} P_e^0$ for different M when (a) $\lambda = 0.02$ and (b) $\lambda = 0.7$. 109

5.1 Numbers of images in (a) training database and (b) testing database that were used in our experiment. 186

List of Figures

2.1 Fingerprints of compressive sensing for sparse signals in the presence of measurement noise or environment noise. The upper row shows the observed signals from (a) traditional sensing, (b) compressive sensing corrupted with measurement noise and (c) compressive sensing corrupted with environment noise. The bottom row shows the corresponding noise histograms of the observed signals above. 21

2.2 Example showing the fingerprints of compressive sensing in a nearly sparse signal with and without the presence of noise. The top row shows the histograms observed from a nearly sparse signal after (a) traditional sensing and (b) compressive sensing. The bottom row shows the histograms observed from a nearly sparse signal in the presence of noise after (c) traditional sensing and (d) compressive sensing. 23

2.3 (a) A hyperspectral image taken from [9] with dimension 1024×1024 pixel. (b) It's monochromatic image (obtained from raw data) corresponding to wavelength of 400nm. (c) The same monochromatic image obtained by compressive sensing and reconstructed from $1024^2 \times 50\%$ compressive measurements. (d) and (e) Histograms of DWT subband 3 coefficients from (d) the traditionally sensed image and (e) the compressively sensed image. 24

2.4 An example showing compressive sensing fingerprints in the (a) 'mug' image captured by a single pixel camera [79]. (b) The histogram of pixel variations (magnitude of the gradient) for the 'mug' image captured by a traditional digital camera. (c) The pixel variation histogram for the compressively sensed image of the same scene acquired using the single pixel camera. 26

2.5 Fitting the histogram of the observed signal to the estimated signal distribution. The Laplace distribution was used to generate each sample of the nearly sparse signal. The left figure shows the fitting result when this signal was obtained by traditional sensing, while the right one shows the result for a when the signal was compressively sensed. 31

2.6 Histograms of DWT coefficients taken from uncompressed Lena (left), the same image after JPEG 2000 compression (right), and the reconstructed compressively sensed Lena (center). 36

2.7 ROC curves obtained by using the image compression detection technique in [58] to identify JPEG 2000 compression in a set of unaltered and JPEG 2000 compressed images (left) and a set of unaltered and compressively sensed images (right). In the right figure "false alarms" correspond only to unaltered images misclassified as JPEG 2000 compressed. Since there is no JPEG 2000 compressed image in the seconde test set, the results in the right figure demonstrate that compressive sensing can be easily misidentified as JPEG 2000 compression. 37
2.8 ROC curves obtained by using the proposed scheme in Section 2.3.2 to identify compressively sensed images from traditionally sensed images (left) and to identify compressively sensed images from traditionally sensed but JPEG 2000 compressed images (right). 39
2.9 Fit the coefficient histogram of compressively sensed Lena with both Laplace model and Laplace mixture model. Coefficients are taken from the third subband after 6-level DWT decomposition with wavelet basis 'bior4.4'. 41
2.10 ROC curves of zero ratio detector and distribution-based detector on signals modeled as sparse signals in the presence of noise for (a) M/N=0.1, (b) M/N=0.4 and (c) M/N=0.9. 'Msure' is short for measurement and 'Envron' is short for environment. 'ZR' denotes the zero ratio detector and 'DB' denotes the distribution-based detector. . 51
2.11 ROC curves of zero ratio detector and distribution-based detector on signals modeled as sparse signals in the presence of noise when different reconstruction algorithms were used. 53
2.12 ROC curves of distribution-based detection on nearly sparse signals and nearly sparse signals in the presence of noise for (a) M/N=0.1, (b) M/N=0.4 and (c) M/N=0.9. 'Msure' is short for measurement and 'Envron' is short for environment. 56
2.13 ROC curves of the first (left) and second (right) step detections on each DWT sub-band coefficients. $M/N = 0.25$ is used in compressive sensing. 59
2.14 ROC curves of the first (left) and second (right) step detections on coefficients of DWT sub-band 3 under different compression ratios of compressive sensing. 60
2.15 Estimated $\hat{M}$ versus the real M for (a) sparse signals in the presence of noise, (b) nearly sparse signals, (c) nearly sparse signals in the presence of noise. 61

3.1 Typical process that a multimedia signal may go through when considering forensics. 67
3.2 Abstract channel model in our information theoretical framework. . . 68
3.3 Channel model for the example of multiple compression detection forensics. 69

3.4 An illustration of the mapping between multimedia states and features in the example of multiple compression detection. 70
3.5 Abstract channel between multimedia states and features in the information theoretical framework for operation forensics. 71
3.6 Abstract channel inner structure for the model in Fig. 3.3. 82
3.7 Normalized histograms of observation noise and their estimated Gaussian distributions (plotted in red lines) on different histogram bins for (a) single compressed images with quantization step size of 6 in the examined subband and (b) doubly compressed images with quantization step size of 6 then 7 in the examined subband. Bin i means that the observation noise on normalized histogram bin $B(iq_{last})$ is examined, where q_{last} denotes the last quantization step size. The mean square error of each estimation is also shown in the subfigure. . 88
3.8 Variance of observation noise versus histogram bin index for (a) single compressed images with quantization step size of 6 in the examined subband; and (b) doubly compressed images with quantization step size of 6 then 7 in the examined subband. 90
3.9 The reachable forensicabilities of different compression quality factors $\mathcal{Q}_M$ and the upper bound of forensicability for different M's. 92
3.10 Experimental error probabilities of several estimators comparing with the theoretical lower bound of error probabilities, where two randomly selected $\mathcal{Q}_{20}$'s are taken as examples: (a) $\mathcal{Q}_{20} = \{..., 8, 11, 13, 6, 5\}$ and (b) $\mathcal{Q}_{20} = \{..., 11, 9, 7, 8, 13\}$. Estimators used in experiments are, in order of displayed legends, maximum likelihood estimator using DCT coefficient histogram on UCID database, Dresden databases, and synthetic data; support vector machine using first significant digit of DCT coefficients on UCID database. 95
3.11 Patterns of $\mathcal{Q}_M$ yielding the highest and lowest forensicabilities. . . . 100
3.12 The best 9 DCT subbands (shown as blue cells) for detection, which yield the highest forensicabilities for (a) $M = 2$, (b) $M = 3$, (c) $M = 4$ and (d) $M = 5$. Numbers 1 through 9 represent the order of these subbands regarding their forensicabilities from the highest to the lowest. 106
3.13 Histogram of λ in subband (2,3) of images from UCID and Dresden databases. 108
3.14 Representative image outliers in UCID and Dresden databases with (a) $\lambda \cong 0.02$ and (b) $\lambda \geq 0.7$. 110

4.1 Fingerprints for detecting the order of resizing and blurring. (a) and (b) are the original image and the DFT of its p-map, respectively. (c) - (f) show the DFT of the p-map of (c) the resized image, (d) the blurred image, (e) the blurred then resized image, and (f) the resized then blurred image. Resizing factor is 1.5 (upscaling). Gaussian blur is used with variance 1. Regions of interests are highlighted by dotted squares and circles. 116

4.2 A confusing example that we may not be able to detect the order. Plotted are DFTs of the p-map of (a) the blurred image, (b) the blurred then resized image, and (c) the resized then blurred image when resizing factor is 1.5 and the variance of Gaussian blur is 0.7. Regions of interests are highlighted by dotted squares and circles. . . 119
4.3 A typical process of estimating the hypotheses. 120
4.4 Compare a simple hypothesis channel and a ROC curve. 124
4.5 The central horizontal line of the DFT of the p-map of (a) an unaltered image, (b) a resized image, (c) a blurred image, (d) a blurred then resized image, and (e) a resized then blurred image. 135
4.6 The process of how to calculate the PSNR from the central horizontal line of the DFT of a p-map. Take Fig. 4.5(d) as an example. 137
4.7 The noise energy pattern signal (dotted blue lines) extracted from the DFT of the p-map and their polynomial fitting curves (solid red lines) for (a) an unaltered image, (b) a resized image, (c) a blurred image, (d) a blurred then resized image, and (e) a resized then blurred image. 139
4.8 Distinguishability test results of detecting double JPEG compression by applying our information theoretical framework and criteria. (a) priors are known and uniform. (b) priors are unknown. 144
4.9 Distinguishability test results of detecting the order of resizing and contrast enhancement by applying our information theoretical framework and criteria. (a) Priors are known and uniform. (b) Priors are unknown. 148
4.10 Distinguishability test results of detecting the order of resizing and blurring by applying our information theoretical framework and criteria. (a) Priors are known and uniform. (b) Priors are unknown. . . 152
4.11 The DFT of the p-map of (a) a single JPEG compressed image with compression quality factor 75, and (b)-(e) double JPEG compressed images with compression quality factors 75 then 85 and interleaved by (b) reszing, (c) blurring, (d) blurring then resizing, and (e) resizing then blurring. The same image in Fig. 4.1(a) is examined in this example. Resizing factor is 1.5 and the variance of Gaussian blur is 1. Regions of interests are highlighted by dotted rectangles. 154

5.1 Histograms of DCT coefficients subtracted from sub-band (0,2) of a natural image been (a) single compressed with specific quantization step 5, (b) doubly compressed with quantization step 3 followed by 5, and (c) doubly compressed with quantization step 7 followed by 5. 162
5.2 The system model considered in this chapter. 167

5.3 Examples of concealabilities related to ROC curves. When the detector achieves perfect detection, the forger has concealability of the fingerprints as 0. When the ROC curve is at or below the random decision line, we say that the forger has achieved concealability as 1. Then for those ROC curves between perfect detection and random decision, the concealability ranges from 0 to 1 and depends on a certain false alarm rate. 171
5.4 An illustration of how to determine $S_0^{(k)}$ and $S_1^{(k)}$ for a certain value of $Y = kq_1$. The vertical arrows denote the position of a certain quantized bin in the coefficient histogram. The horizontal line segment at the bottom of each arrow represents the quantization interval where all values within this range will be mapped into the quantized bin indicated by the arrow. lq_2 is the quantized bin that kq_1 will be mapped into during the recompression. According to different positions of lq_2 and its quantization intervals, there are four cases for $S_0^{(k)}$, while $S_1^{(k)}$ keeps the same for the same kq_1. 177
5.5 Histograms of DCT coefficients of an anti-forensically modified and double compressed image with anti-forensic strength (a) $\alpha = 0$, (b) $\alpha = 0.4$, and (c) $\alpha = 1$. 181
5.6 Concealability, rate, and distortion triples for all tested anti-forensic strengths and secondary quality factors with distortion defined based on (a) MSSIM in (5.11) and (b) MSE. 187
5.7 Tradeoff of concealability, rate, and distortion for the case where the second quality factor is smaller than the first one. (a) plots the reachable (C,R,D) points, where the points with the same marker and color are those who have the same secondary compression quality factor but have been applied different anti-forensic strengths. The higher the concealability, the more the anti-forensic strength. (b) is the polynominal fitting surface of (a). 190
5.8 Rate changes with anti-forensic strength for lower secondary quality factor case. 191
5.9 Tradeoff of concealability, rate, and distortion for the higher secondary quality factor case. (a) plots the R-D-C points. Points with the same marker and same color are those obtained by using the same secondary quality factor but different anti-forensic strengths. (b) is the polynominal fitting surfaces of (a). 193

Chapter 1

Introduction

1.1 Motivation

Nowadays, multimedia has played an important role in recording and conveying message. Many critical decisions and statements made by governments, news reporters and law enforcement are based on multimedia evidence. However, the increase of easily accessible editing software and online tools has made multimedia content untrustworthy. People begin questioning the origin of given multimedia content and how it was processed. To verify the authenticity of multimedia content, the field of multimedia forensics has been developed. Researchers in multimedia forensics aim to study and develop forensic techniques to identify the origin of multimedia content and how it was processed after capture.

In the past decade, many forensic techniques have been developed to verify the authenticity of multimedia content [88]. For example, by disassembling a digital camera into separated components and appropriately modeling each of them, forensic researchers can estimate the parameters of camera components and thus identify the camera model that was used to capture an image under question [94]. Color filter arrays and sensor pattern noise can also used to identify digital cameras [59, 78]. Given that most cameras automatically compress the captured image using their prescribed compression parameters, identifying the type of source en-

coder and its parameters can help us find the digital camera that was used to generate a compressed image [58]. Besides identifying the source of a given multimedia file, forensic techniques also enable us to detect post-processing manipulations applied on the multimedia content. For example, given current techniques, we can detect global and local contrast enhancement [86], resizing [77], single and double compression [31, 76], median filtering [50], blurring [96], and so on.

To take care of new emerged technologies and deal with more complex problems, forensic researchers have never stopped improving their forensic techniques and also proposing new algorithms. Specifically, new features have been found for detecting the same manipulation operation [19, 55, 76]. State of the art technologies have been used in forensics, such as machine learning [55, 74] and deep learning [60]. With so many forensic techniques proposed for one forensic problem, such as double compression detection, fusion algorithms have been proposed to effectively combine these techniques to achieve better detection performance [3]. In addition, given that most manipulation detectors can only detect single manipulation operations while making a forgery often involves the application of multiple operations, more complicated manipulation processes have been studied by forensic researchers recently. For example, multiple JPEG compressions have been examined in forensics to identify the number of applied compressions [66]. Double JPEG compressions interleaved with resizing or contrast enhancement in between have also been studied to evaluate the effect of intermediate operations in double compression detections [6, 36].

While forensic researchers strive to provide new solutions for more realistic forensic scenarios, it is also of key interest to understand the limitations of mul-

timedia forensics. This limitation may due to the constraint of existing forensic techniques when dealing the emergence of new technologies. For example, compressive sensing technology has been proposed recently. It is a promising acquisition technology that has been widely used in various fields of digital signal processing. However, there has been no forensic techniques designed to consider this acquisition method. Furthermore, existing forensic techniques identifying the source of an image may easily be confused by images acquired by compressive sensing.

Another limitation of forensics comes from the limited information contained in multimedia content. In typical processes of forensic algorithms, features are extracted from multimedia content to make the estimation of the process history happened on this multimedia content. The limited statistics contained by certain features must result in the limitation of forensic information these features can convey about the multimedia's process history. Forensic researchers have begun to notice these limitations as they attempt to identify more complex processing histories. For example, when detecting multiple JPEG compressions, the detection performance is largely degenerated as investigators try to detect four times of compressions [66]. In addition, if multiple different operations are applied on multimedia content, the interplay between these operations may also results in the failure of identifying the complete processing history.

Noticed of these constraints, one would wonder what are the fundamental limits in multimedia forensics? How much information that we can extract from multimedia content towards identifying its processing history? A. Swaminathan *et al.* proposed an estimation framework and a pattern classification framework

for component forensics and explored the fundamental limits towards identifying cameras [92, 93]. However, there is no work considering the fundamental limits of multimedia forensics towards estimating the processing history of multimedia content after capture. Different from proposing empirical forensic techniques, which answers "what can we do", exploring the fundamental limits of forensics answers the question of "what cannot we do". It enables us to acknowledge the capability of forensic investigators and know how far forensic techniques can be improved.

With the development of multimedia forensics, anti-forensic schemes have also been studied. In anti-forensics, researchers stand on forgers' side and develop attacking techniques to conceal fingerprints of manipulations and thus fool forensic techniques. Specifically, we now have anti-forensic schemes to conceal the fingerprints of compression [87], contrast enhancement [15], resizing [49], median filtering [101], and so on. Studying anti-forensics enables us to understand the behavior of forgers and their possible attacks so that forensic investigators can find the weakness of their techniques and improve them accordingly. When applying anti-forensics, forgers mainly concern about the effectiveness of their anti-forensic techniques. Meanwhile, there are certain constraints on applied anti-forensics to make sure the modified multimedia content is not too distorted that forensic investigators can immediately tell the manipulation. Limited by these factors, the behavior of forgers is dependent on the fundamental tradeoffs in anti-forensics. Finding and characterizing these tradeoffs can help us better predicting the behavior of forgers and preparing corresponding reactions.

1.2 Dissertation Outline

From the discussion above, we can clearly see the necessity of acknowledging and exploring the fundamental limits in multimedia forensics and anti-forensics. In this dissertation, both theoretical and empirical methods have been proposed to explore the fundamental problems in forensics and anti-forensics. Specifically, to solve the limitation problem of existing forensic techniques when dealing with newly emerged technologies, a set of empirical algorithms have been developed. Furthermore, to explore the fundamental limits of estimating multimedia content's processing history, we propose several theoretical frameworks for different forensic scenarios. In addition, the fundamental tradeoffs in multimedia anti-forensics are proposed and characterized. The rest of this dissertation is organized as follows.

1.2.1 Compressive Sensing Forensics (Chapter 2)

Compressive sensing, as a new signal acquisition technology known for its sub-Nyquist sensing rate, has seen increased popularity in recent years. However, current forensic techniques identifying a signal's acquisition history do not account for the possibility that a signal could be compressively sensed. In this chapter, we propose a set of forensic techniques to identify signals acquired by compressive sensing. We do this by first identifying the fingerprints left in a signal by compressive sensing. We then propose two compressive sensing detection techniques that can operate on a broad class of signals. Since compressive sensing fingerprints can be confused with fingerprints left by traditional image compression techniques, we propose a

forensic technique specifically designed to identify compressive sensing in digital images. Additionally, we propose a technique to forensically estimate the number of compressive measurements used to acquire a signal. Through a series of experiments, we demonstrate that each of our proposed techniques can perform reliably under realistic conditions. Simulation results show that both our zero ratio detector and distribution-based detector yield perfect detections for all reasonable conditions that compressive sensing is used in applications, and the specific two-step detector for images can at least achieve probability of detection of 90% for probability of false alarm less than 10%. Additionally, our estimator for the number of compressive measurements can well reflect the real number.

1.2.2 Fundamental Limits in Operation Forensics (Chapter 3)

While more and more forensic techniques have been proposed to detect the processing history of multimedia content, one starts to wonder if there exists a fundamental limit on the capability of forensics. In other words, besides keeping on searching what investigators can do, it is also important to find out the limit of their capability and what they cannot do. In this chapter, we explore the fundamental limit of operation forensics by proposing an information theoretical framework. Specifically, we consider a general forensic system of estimating operations' hypotheses based on extracted features from the multimedia content. In this system, forensicability is defined as the maximum forensic information that features contain about operations. Then, due to its conceptual similarity with mutual infor-

mation in information theory, forensicability is measured as the mutual information between features and operations' hypotheses. Such a measurement gives the error probability lower bound of all practical estimators which use these features to detect the operations' hypotheses. Furthermore, it can determine the maximum number of hypotheses that we can theoretically detect. To demonstrate the effectiveness of our proposed information theoretical framework, we apply this framework on a forensic example of detecting the number of JPEG compressions based on DCT coefficient histograms. We conclude that, under typical settings of forensic analysis, the maximum number of JPEG compressions that we can perfectly detect using DCT coefficient histogram features is 4. Furthermore, we obtain the optimal strategies for investigators and forgers based on the fundamental measurement of forensicability.

1.2.3 Fundamental Limits in Order Forensics (Chapter 4)

When multiple manipulation operations are applied on multimedia content, investigators not only need to identify the use of each operation, but also need to detect the order of these operations. By detecting the order of operations, investigators can know the complete processing history of the multimedia content. Furthermore, detecting the order of operations may also provide information about when the multimedia content was manipulated and who manipulated it. However, when multiple operations are involved in the analysis, the interplay among operations may affect the fingerprints of earlier applied operations and make it difficult to detect the order of operations. This leads to a fundamental question of when we can

and cannot detect the order of operations. In this work, we propose an information theoretical framework by using mutual information based criteria to determine the detectability of the order of operations regarding certain features and estimators. A case study of detecting the order of resizing and blurring has been examined to demonstrate the effectiveness of the proposed framework and criteria. In addition, two known forensic problems are considered in the simulations to show that the results obtained from the proposed framework and criteria match those of existing works.

1.2.4 Fundamental Tradeoffs in Compression Anti-forensics (Chapter 5)

To conceal fingerprints of manipulation operations, anti-forensics has been used by forgers to fool forensic detectors. However, when anti-forensic techniques are applied to multimedia content, distortion may be introduced, or the data size may be increased. Furthermore, when compressing an anti-forensically modified forgery, a tradeoff between the rate and distortion is introduced into the system. As a result, a forger must balance three factors: how much the fingerprints can be forensically concealed, the data rate, and the distortion, are interrelated to form a three dimensional tradeoff. In this paper, we characterize this tradeoff by defining concealability and using it to measure the effectivenss of an anti-forensic attack. Then, to demonstrate this tradeoff in a realistic scenario, we examine the concealability-rate-distortion (C-R-D) tradeoff in double JPEG compression anti-forensics. To

evaluate this tradeoff, we propose flexible anti-forensic dither as an attack in which the forger can vary the strength of anti-forensics. To reduce the time and computational complexity associated with decoding a JPEG file, applying anti-forensics, and recompressing, we propose an anti-forensic transcoder to efficiently complete these tasks in one step. Through simulation, two surprising results are revealed. One is that if a forger uses a lower quality factor in the second compression, applying anti-forensics can both increase concealability and decrease the data rate. The other is that for any pairing of concealability and distortion values, achieved by using a higher secondary quality factor, can also be achieved by using a lower secondary quality factor at a lower data rate. As a result, the forger has an incentive to always recompress using a lower secondary quality factor.

Chapter 2

Compressive Sensing Forensics

Since the initial development of digital multimedia forensics, researchers have sought to identify how different digital signals were captured and stored. Information about how a signal was acquired can be used to both identify the specific device used to capture the signal and to verify the signal's authenticity. Furthermore, knowledge of how a signal was captured can be used to help trace its processing history. As a result, determining how a signal was acquired has become an important forensic problem.

Typically, forensic algorithms determine how a signal was acquired by identifying imperceptible traces introduced into a digital signal during the acquisition process. These traces, which are known as fingerprints, arise due to properties of the sensor used to capture the signal or as a result of the signal processing operations used to form the digital signal. Existing forensic algorithms capable of identifying a signal's acquisition history are focused almost exclusively on images and videos [16, 59, 78, 88, 94]. While each of these specifically designed techniques performs strongly, it is necessary to develop forensic algorithms capable of identifying the acquisition history of a broader class of signals.

Recently, a new method of capturing signals known as *compressive sensing* has gained considerable attention. Compressive sensing is a signal processing technique

capable of acquiring sparse signals at sampling rates below the Nyquist rate [29]. Rather than measuring the signal's value at a series of uniformly spaced points, each compressive measurement corresponds to a randomly weighted summation of the entire signal. The sparse signal can then be reconstructed using l_1 minimization from much fewer measurements than are needed by traditional uniform sampling [11]. Furthermore, many real signals that are not ideally sparse can be modeled as either sparse signals in the presence of noise or signals that are 'nearly sparse'. Compressive sensing can be used to acquire these signals with low amounts of reconstruction error [12].

Due to the effectiveness of compressive sensing's sub-Nyquist acquisition rate, researchers in various signal processing fields have applied compressive sensing techniques to many signal acquisition systems. These applicable fields include but not limited to magnetic resonance imaging [63], photoacoustic imaging [80], astronomical imaging [8], radar [71], electrocardiography [2], networked data [41], and speech and audio [100].

While acquisition schemes based on compressive sensing principles are widely studied in the realm of research, the impact of compressive sensing has led people to design and build real devices based on this technique. Single pixel or single sensor acquisition devices have been developed for capturing conventional images [1] and hyperspectral images [91]. In these applications, compressive sensing not only reduced the acquisition power but also solved the 'out of focus' problem encountered in traditional cameras [1]. Moreover, due to the power consumption of billions of A-to-D conversion in video acquisition, a custom CMOS chip was designed by

adopting compressive sensing technology to slash energy consumption by a factor of 15 [83]. Devices that apply compressive sensing to other applicable signals have also been developed and built [47]. Researchers from Rice University have even started a company, called InView, to develop low cost shortwave infrared cameras using compressive sensing [27].

While an increasing number of technologies have begun to make use of compressive sensing, there are currently no existing forensic techniques capable of differentiating between signals captured using compressive sensing and those captured by traditional uniform sampling. This has important consequences for the forensics community.

As the number of devices that incorporate compressive sensing into their signal processing pipeline increases, detecting the use of compressive sensing will become an important part of forensically identifying a signal's origin. A motivating example can be seen in hyperspectral imaging, which is used in many critical applications such as surveillance drones and environmental monitoring. Compressive sensing has been recently used to capture and store hyperspectral images [45]. Detecting evidence of compressive sensing in a hyperspectral image can help forensic investigators identify the device. Furthermore, there may be scenarios where our government is presented with an image captured by another government's surveillance drone. In this scenario, we may want to analyze the image to 1) verify the validity of the image and 2) understand the capabilities of the other government's surveillance drone. Similarly, hyperspectral images of landscapes may potentially be used in court cases related to environmental contamination or mineral rights.

Additionally, the use of compressive sensing can affect the output of existing forensic algorithms. For example, compressive sensing may also be used to acquire, compress, and store certain types of images [45]. However, existing compression detection schemes in [58] and [61] may misidentify a compressively sensed image as an image that has been captured by a standard digital camera, then subsequently compressed. Thus, it is necessary to design a specific forensic scheme for compressive sensing detection to solve such confusions. In summary, it is clear that the identification of compressively sensed signals is an important forensic problem.

In this chapter, we propose a new forensic technique capable of identifying signals that have been acquired by compressive sensing. We begin by identifying the fingerprints that compressive sensing introduces into a signal. Because virtually no compressively sensed signal is truly sparse, we show that the reconstruction error introduced into compressively sensed signals has certain characteristics. We use these characteristics as compressive sensing's fingerprints and examine these fingerprints under three models commonly applied to compressively sensed signals: sparse signals in the presence of noise, nearly sparse signals, and nearly sparse signals in the presence of noise. We then propose a set of forensic techniques to identify compressively sensed signals that fit each of these models. Furthermore, we develop a forensic technique specifically designed to identify compressively sensed images and differentiate them from images that have undergone traditional lossy compression. Additionally, we propose a technique to forensically estimate the number of compressive measurements used to acquire a signal.

The remainder of this paper is organized as follows. In Section 2.1, we provide

a brief review of compressive sensing and present three different models of compressively sensed signals. In Section 2.2, we identify and analyze the fingerprints left in a signal by compressive sensing. Using these fingerprints, we propose two different compressive sensing detection techniques in Section 2.3. To address specific challenges encountered when identifying compressive sensing in digital images, we present a two step compressive sensing detection technique that can discriminate between images that have been compressed using wavelet-based coders and images that have been compressively sensed in Section 2.4. In Section 2.5, we propose an estimator for the number of compressive measurements used to acquire a signal. A series of experimental results are presented in Section 5.5 that demonstrate the effectiveness of our proposed forensic techniques. Finally, in Section 2.7 we conclude this paper.

2.1 System Model

We begin this section by providing a brief overview of compressive sensing. We then discuss the three different models used for real world signals that are compressively sensed. Throughout this paper, we will use $\underline{s}$ and $\underline{x}$ to denote the original signal and the observed signal, respectively. Given the observed signal may be obtained by either traditional sensing or compressive sensing, it will correspondingly equal to the direct, maybe noisy, observation of the original signal, or the reconstructed one from compressive measurements.

2.1.1 Compressive Sensing Overview

Traditionally, a discretely indexed signal is formed from a continuously indexed signal through uniform sampling. During uniform sampling, observations of the continuously indexed signal are performed at uniformly spaced intervals over a fixed duration. As a result, each entry s_i in a discretely indexed signal $\underline{s} = (s_1, s_2, \ldots, s_n)^T$ corresponds to a single, direct measurement of the continuously indexed signal, and we directly observe these measurements in traditional sensing. Thus, if we use $\underline{x}$ to denote the observed signal in such case, then $\underline{x} = \underline{s}$.

The recent development of compressive sensing has allowed sparse signals, which have only a few nonzero entries, to be captured with far fewer observations than traditional sampling. During compressive sensing, each compressive measurement corresponds to a linear combination of the continuously indexed signal's values at all the locations that would be observed during uniform sampling. Defining the weighting vector for the i^{th} compressive measurement as $\underline{\varphi}_i$, then each compressive measurement y_i can be written as

$$y_i = \underline{\varphi}_i^T \underline{s}. \tag{2.1}$$

If $m (m \ll n)$ compressive measurements are collected, the transpose of the set of weighting vectors can be vertically concatenated to form the observation matrix Φ. As a result, the measurement vector $\underline{y} = (y_1, y_2, \ldots, y_m)^T$ containing each compressive measurement can be written as

$$\underline{y} = \Phi \underline{s}. \tag{2.2}$$

Typically, random matrices are used for observation matrices Φ in order to satisfy

the restricted isometry property for later reconstruction [12]. In this work, we use Gaussian distribution with zero mean and unit variance to generate matrix Φ.

After the compressive measurements are obtained, the discretely indexed signal $\underline{x}$, which we will observe from compressive sensing, is reconstructed from the compressive measurements. This is done by solving the following constrained l_1 minimization problem

$$\min_{\tilde{\underline{x}}} ||\tilde{\underline{x}}||_{l_1}, \quad s.t. \ \Phi\tilde{\underline{x}} = \underline{y}. \tag{2.3}$$

If $\underline{s}$ is sparse, then given enough compressive measurements, $O(k \log n)$, where k and n are the sparsity and length of $\underline{s}$ respectively, the signal can be perfectly reconstructed, i.e. $\underline{x} = \underline{s}$ [11].

Compressive sensing forensics, however, is a reverse engineering problem of compressive sensing, which starts from the reconstructed signal and tries to reveal how the signal was acquired. Forensic investigators only observe a reconstructed signal $\underline{x}$. Then, based on the fingerprints extracted from this signal, they identify whether the observed signal was traditionally sensed or compressively sensed and reconstructed. Furthermore, forensic investigators can also estimate the number of compressive measurements m solely based on the reconstructed signal.

2.1.2 Signal Model

In theory, if a truly sparse signal is compressively sensed, it can be perfectly reconstructed [11]. In practice, however, this is rarely the case. Often, the compressive measurements of a truly sparse signal will be corrupted by noise. This can occur

due to sensing in a noisy environment or due to noise within the sensors themselves. Furthermore, it is often the case that signals of interest are not truly sparse, but rather nearly sparse or 'compressible'. While non-sparse but compressible signals cannot be perfectly reconstructed, a bound can be placed on the reconstruction error [12]. If enough compressive measurements are captured, the reconstruction error can be made sufficiently small.

Here, we discuss several commonly used models applied to signals that are compressively sensed in real world scenarios. In subsequent sections, we will exploit the effects of these nonideal conditions to identify the use of compressive sensing.

Sparse Signals in the Presence of Noise

There are many scenarios in which a true signal has only a few nonzero coefficients (i.e., nonzero entries s_i in $\underline{s}$), but the signal is corrupted by noise during sensing. These signals can be modeled as sparse signals in the presence of noise. For example, in radar signal analysis the time-frequency plane is discretized into a grid where the number of grid cells is much larger than the total number of targets. The radar coefficients under this time-frequency shift operator basis are modeled as sparse signals in the presence of noise [43].

Under this model, let $\underline{s}$ represent a sparse signal to be sensed. If $\underline{s}$ is sensed using traditional uniform sampling, the observed signal $\underline{x}$ is given by

$$\underline{x} = \underline{s} + \underline{\eta}. \tag{2.4}$$

where $\underline{\eta}$ is a vector containing i.i.d. noise. Regardless of whether the noise originates in the sensor or is due to an environmental source, a unique noise measurement

occurs at each signal observation x_i.

If $\underline{s}$ is compressively sensed, however, noise can be introduced into the compressive measurements. Under some scenarios, additive noise directly corrupts each compressive measurement [43]. This is equivalent to sensing using a noisy sensor. We refer to this type of noise as measurement noise, and model the compressive measurements as

$$\underline{y} = \Phi\, \underline{s} + \underline{\eta}^m, \tag{2.5}$$

where $\underline{\eta}^m$ is i.i.d. noise. In other scenarios, the sparse signal directly mixes with some noise process while it is being sensed [100]. We refer to this type of noise as environment noise. We model compressive measurements in the presence of i.i.d. environment noise $\underline{\eta}^e$ as

$$\underline{y} = \Phi(\underline{s} + \underline{\eta}^e). \tag{2.6}$$

If the compressive measurements are corrupted by either measurement or environment noise, the sparse signal is no longer reconstructed using (2.3). Instead, the reconstructed signal $\underline{x}$ is obtained by solving

$$\min_{\tilde{\underline{x}}} ||\tilde{\underline{x}}||_{l_1}, \;\; s.t. \;\; ||\underline{y} - \Phi\tilde{\underline{x}}||_{l_2}^2 \leq \epsilon \tag{2.7}$$

where ϵ is a parameter that depends on the noise power [17]. We note that in this equation, the constraint present in (2.3) is replaced with the inequality $||\underline{y} - \Phi\tilde{\underline{x}}||_{l_2}^2 \leq \epsilon$.

Nearly Sparse Signals

While many important types of signals are not truly sparse, they satisfy certain conditions allowing them to be well approximated by sparse signals. These signals

are known as nearly sparse or compressible signals. The discrete wavelet transform coefficients of a digital image corresponding to a natural scene are a widely used example of a nearly sparse signal [65]. Gabor coefficients of certain classes of oscillatory signals can also be modeled as nearly sparse signals [34]. Though nearly sparse signals cannot be perfectly reconstructed if they are compressively sensed, they can be reconstructed with little error if enough compressive measurements are obtained.

To formally define nearly sparse signals, we first sort the entries of the signal $\underline{s}$ in descending order $s_{(1)}, s_{(2)}, \ldots, s_{(n)}$, such that $|s_{(1)}| \geq |s_{(2)}| \geq \ldots \geq |s_{(n)}|$. The signal $\underline{s}$ is compressible if and only if its sorted coefficients fall inside a weak l_p ball of radius R for some $0 < p < \infty$ [12], i.e.

$$|s_{(i)}| \leq R \cdot i^{-1/p}, \;\; i = 1, 2, \ldots, n. \tag{2.8}$$

We model nearly sparse signals as compressible signals whose entries are i.i.d. random variables. Signals drawn from many commonly occurring distributions such as the Laplace and Gaussian distributions are compressible [12].

Nearly Sparse Signals in the Presence of Noise

In some real world scenarios, a nearly sparse signal may be compressively sensed in a noisy environment. As a result, we adopt nearly sparse signals in the presence of noise as a third signal model. These signals can be viewed as a combination of the previous two models. Provided that the noise power is sufficiently small, nearly sparse signals will remain compressible when corrupted by noise. As a result, we will see that detecting compressive sensing in signals that fit this models

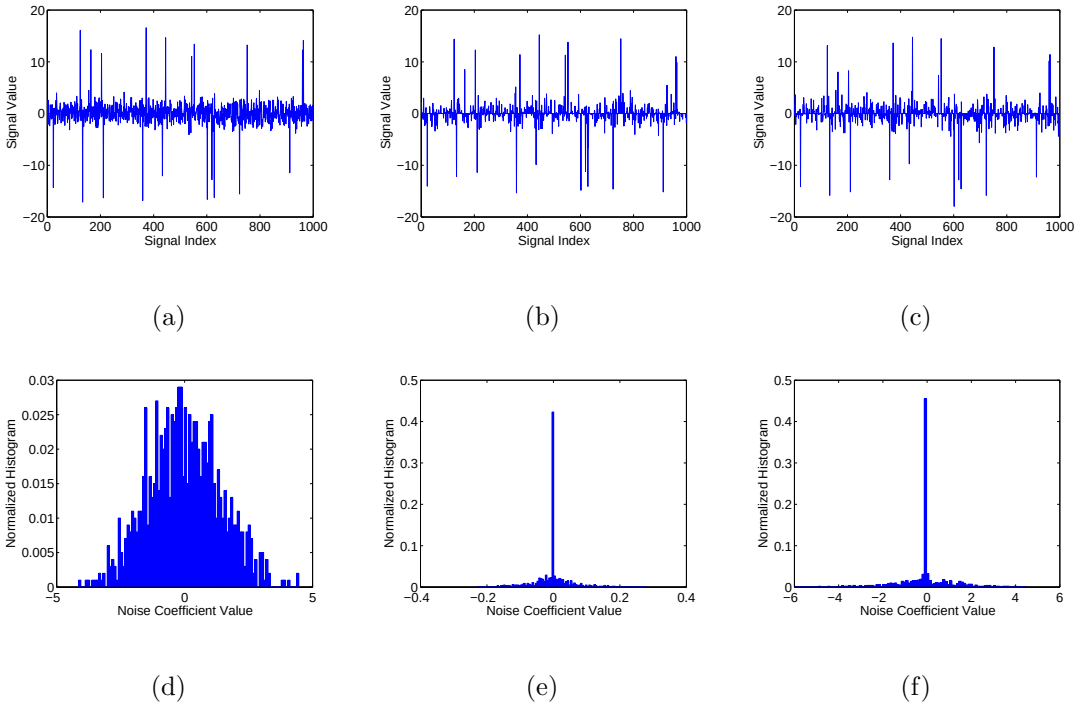
is similar to detecting compressive sensing in nearly sparse signals.

2.2 Compressive Sensing Fingerprints

To identify the fingerprints left by compressive sensing, we first examine sparse signals in the presence of noise, then examine nearly sparse signals.

Consider a signal $\underline{x}$ formed by sensing a sparse signal $\underline{s}$ in the presence of noise. Assuming that the locations of the nonzero components of $\underline{s}$ are known, the entries of $\underline{x}$ that do not correspond to nonzero values can be gathered together to form the vector $\underline{x}^n$. If $\underline{x}$ was acquired using traditional uniform sampling, each entry in $\underline{x}^n$ will directly correspond to a single noise observation. As a result, the normalized histogram of $\underline{x}^n$ approximates the distribution of the noise source. This can be seen in Fig. 2.1(d).

This is not the case, however, if $\underline{x}$ was acquired via compressive sensing. If measurement noise is encountered during sensing, the noise affects each compressive measurement. During reconstruction, no single value of $\underline{x}$ will correspond to a single noise observation. If environment noise is present during compressive sensing, both the sparse signal and the noise will be captured during the measurement process. Reconstructing the signal by solving (2.7), however, ensures that $\underline{x}$ will accurately reconstruct the $\underline{s}$ but not the noise. As a result, if $\underline{x}$ was captured using compressive sensing, the normalized histogram of $\underline{x}^n$ will not match the distribution of the noise source. In fact, because $\underline{x}$ was chosen to maximize the sparsity of the reconstructed signal, a significant number of entries in $\underline{x}^n$ will be zero or near zero. This will result



(a) (b) (c)

(d) (e) (f)

Figure 2.1: Fingerprints of compressive sensing for sparse signals in the presence of measurement noise or environment noise. The upper row shows the observed signals from (a) traditional sensing, (b) compressive sensing corrupted with measurement noise and (c) compressive sensing corrupted with environment noise. The bottom row shows the corresponding noise histograms of the observed signals above.

in the presence of an impulsive peak at zero in the normalized histogram of $\underline{x}^n$ as can be seen in Fig.s 2.1(e) and (f). This peak is the fingerprints left by compressive sensing for sparse signals in the presence of noise.

A similar effect can be observed if $\underline{x}$ was formed by sensing a nearly sparse signal. As it is shown in Fig. 2.2(a), the normalized histogram of traditionally sensed signal $\underline{x}$ will closely match the distribution of the nearly sparse signal being sensed. However, the use of compressive sensing will greatly increase the histogram's kurtosis and result in a big concentration at zero as can be seen in Fig. 2.2(b). Furthermore, this result holds true for nearly sparse signals in the presence of noise, as can be seen in Fig. 2.2(c) and (d).

To show the effectiveness of compressive sensing fingerprints in real applications, we take a hyperspectral image, which is shown in Fig. 2.3(a), as an example. Hyperspectral images are composed of many sub-images in different spectrum bands, each of which can be obtained by compressive sensing [91]. Therefore, in this example, we take one sub-image out to examine. Comparing the traditionally sensed sub-image in Fig. 2.3(b) and the compressively sensed image in Fig. 2.3(c), we can hardly tell the difference. However, the histogram of transform domain coefficients from compressively sensed image, as it is shown in Fig. 2.3(d), has a much higher kurtosis at zero than that from the traditionally sensed image, which is shown in Fig. 2.3(e).

Furthermore, in order to show that such fingerprints also exist in real compressive sensing devices, we examine a single pixel camera captured image and an image of the same scene but being captured by a traditional digital camera [79]. The

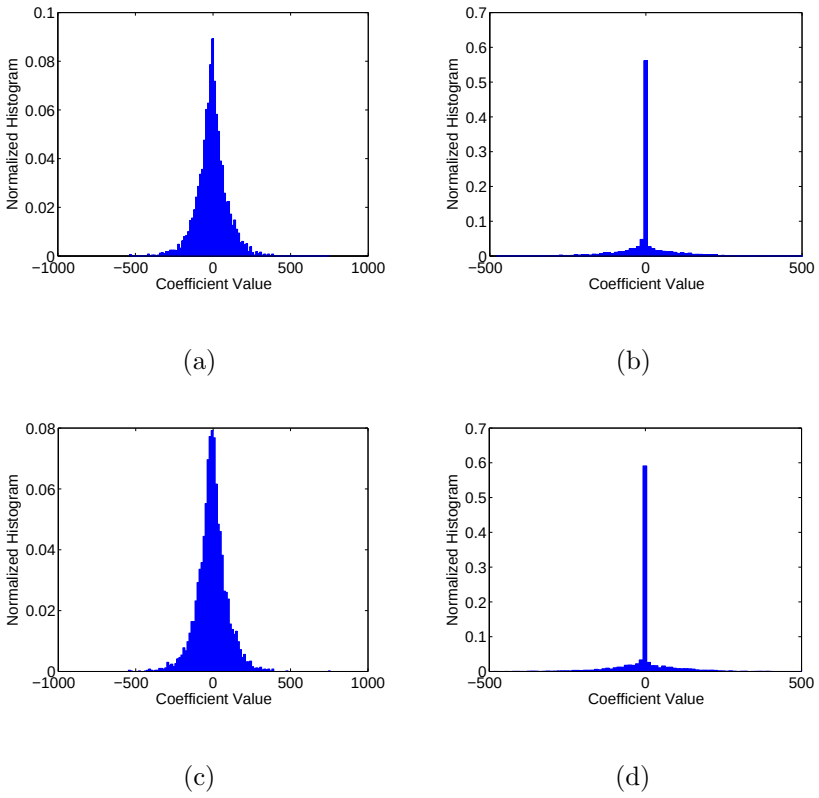
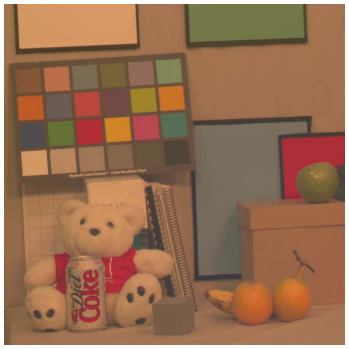
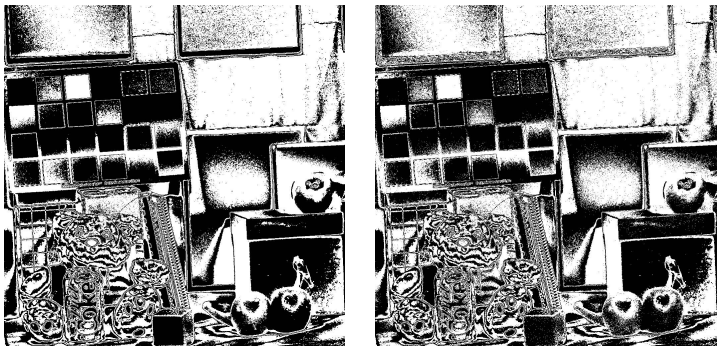


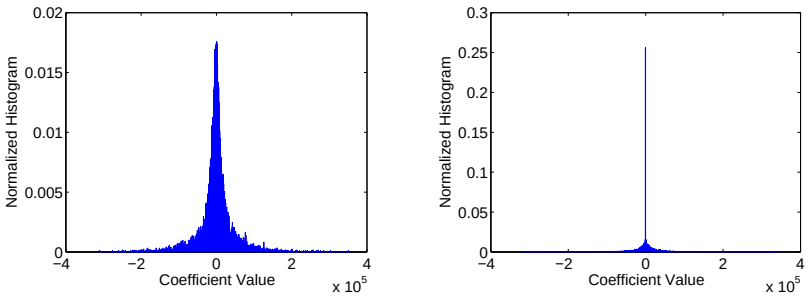
Figure 2.2: Example showing the fingerprints of compressive sensing in a nearly sparse signal with and without the presence of noise. The top row shows the histograms observed from a nearly sparse signal after (a) traditional sensing and (b) compressive sensing. The bottom row shows the histograms observed from a nearly sparse signal in the presence of noise after (c) traditional sensing and (d) compressive sensing.



(a)



(b) (c)



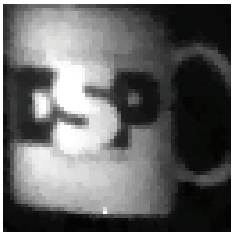
(d) (e)

Figure 2.3: (a) A hyperspectral image taken from [9] with dimension 1024×1024 pixel. (b) It's monochromatic image (obtained from raw data) corresponding to wavelength of 400nm. (c) The same monochromatic image obtained by compressive sensing and reconstructed from $1024^2 \times 50\%$ compressive measurements. (d) and (e) Histograms of DWT subband 3 coefficients from (d) the traditionally sensed image and (e) the compressively sensed image.

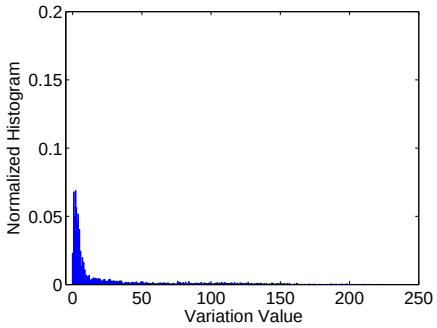
single pixel camera in [79] obtains each compressive measurement by projecting the scene onto a randomized digital micromirror array and optically calculate the linear combination. We use the 'mug' image captured by a single pixel camera in [79], as it is shown in Fig. 2.4(a), to present the fingerprints of compressive sensing. Because the reconstruction step was performed by minimizing the total variation, the domain that compressive sensing fingerprints are present in is the pixel variations, i.e., gradient magnitudes. Fig.s 2.4(b) and 2.4(c) show the histograms of pixel variations for the traditionally sensed 'mug' image, and its compressively sensed version, respectively. We can see from Fig. 2.4(c) that a peak corresponding to a large concentration of components is present at the zero bin for the compressively sensed image. These fingerprints are absent from the traditionally captured image's histogram on the left.

We note that the compressive sensing fingerprints' existence is due to the sparse representation of the signal created upon reconstruction. Because all reconstruction algorithms enforce sparsity in one way or another, these fingerprints will be present in the sparsity domain regardless of the reconstruction algorithm.

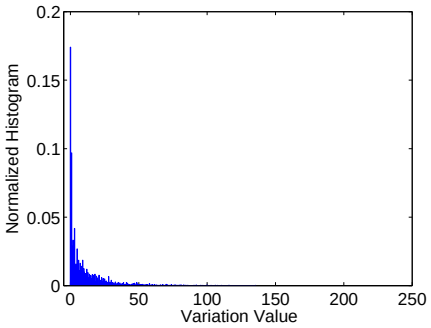
Though we focus on the basis pursuit (BP) reconstruction algorithm in this chapter, we note that there are several algorithms that can be used to reconstruct a compressively sensed signal such as orthogonal matching pursuit (OMP) [97], least absolute shrinkage and selection operator (LASSO) [95], and total variation (TV) [62]. We note that as long as a reconstruction algorithm seeks a sparse representation of the compressive measurements, similar fingerprints will be present in the reconstructed signal.



(a)



(b)



(c)

Figure 2.4: An example showing compressive sensing fingerprints in the (a) 'mug' image captured by a single pixel camera [79]. (b) The histogram of pixel variations (magnitude of the gradient) for the 'mug' image captured by a traditional digital camera. (c) The pixel variation histogram for the compressively sensed image of the same scene acquired using the single pixel camera.

2.3 Compressive Sensing Detection

Now that we have identified the fingerprints left by compressive sensing, we are able to develop a set of forensic techniques to detect its use [23]. Detecting the use of compressive sensing is equivalent to differentiating between the following hypotheses

$$\begin{aligned} H_0 &: \ \underline{x} \text{ was obtained using traditional sampling,} \\ H_1 &: \ \underline{x} \text{ was obtained using compressive sensing,} \end{aligned} \tag{2.9}$$

where $\underline{x}$ is a discretely indexed signal of unknown origin. To do this, we first need to obtain some measure of the strength of any compressive sensing fingerprints present in $\underline{x}$. Measurement of these fingerprints' strength, however, depends on the appropriate signal model for $\underline{x}$ as well as the amount of side information known by the forensic investigator. To account for this, we propose two different compressive sensing detection techniques that are appropriate in different forensic scenarios.

2.3.1 Zero Ratio Detection Scheme

In many cases, a forensic investigator knows little more than the fact that the signal in question fits one of the three signal models outlined in Section 2.2. If this is the case, the forensic investigator cannot leverage any side information such as the signal or noise distribution while measuring the strength of compressive sensing fingerprints. The investigator can, however, make use of the fact that if compressive sensing was performed, it was done under nonideal conditions.

Assume temporarily that $\underline{x}$ can be modeled as a sparse signal $\underline{s}$ sensed in

the presence of noise. We assume that the noise has a continuous distribution and a nonzero variance, i.e. its distribution is not an impulse. From Section 2.2, we know that under hypothesis H_0 each entry of $\underline{x}^n$ will correspond directly to a noise observation. As a result, the distribution of the entries in $\underline{x}^n$ will match the noise distribution. By contrast, under hypothesis H_1, an impulsive peak located at zero will occur in the distribution of the entries of $\underline{x}^n$. Because of this, we can state

$$\mathbb{P}(x_i^n = 0|H_0) \ll \mathbb{P}(x_i^n = 0|H_1). \tag{2.10}$$

Though a forensic investigator may not know the noise distribution, the investigator can use (2.10) to measure the strength of compressive sensing fingerprints by calculating the ratio of zero valued entries in $\underline{x}^n$ to its total length.

Since in practice many of the techniques used to solve (2.3) or (2.7) result in values of $\underline{x}^n$ close to but not exactly equal to zero, we measure the strength of the fingerprints as follows. Let $\Lambda_\varepsilon(\underline{x}^n)$ denote the number of elements in $\underline{x}^n$ which have an absolute value no greater than ε. We calculate the zero ratio fingerprints' strength using the equation

$$\xi_z(\underline{x}^n) = \frac{\Lambda_\varepsilon(\underline{x}^n)}{\ell(\underline{x}^n)}, \tag{2.11}$$

where $\ell(\underline{x}^n)$ is the length of the vector $\underline{x}^n$. When calculating Λ_ε, ε is chosen to be $\varepsilon = ||\,\underline{x}^n\,||_\infty/\alpha$, where α is a parameter that controls the range of values of $\underline{x}^n$ that are counted as zeros. Experimentally, we have observed that choosing $\alpha = 100$ yields desirable results. We then perform compressive sensing detection using the

following decision rule

$$\delta_z = \begin{cases} H_0 & \text{if } \xi_z(\underline{x}^n) < \tau_z, \\ H_1 & \text{if } \xi_z(\underline{x}^n) \geq \tau_z. \end{cases} \tag{2.12}$$

where τ_z is a decision threshold.

In reality, the locations of the nonzero values of $\underline{s}$ may not be known to a forensic investigator, thus making it difficult to form $\underline{x}^n$ from $\underline{x}$. In this scenario, two approaches can be taken to perform compressive sensing detection. Since $\underline{s}$ will contain a small number of nonzero entries, entries in $\underline{x}$ corresponding to these entries in $\underline{s}$ will have values significantly larger in magnitude than the rest. In the first approach, if the entries of $\underline{x}$ are sorted in descending order, a substantial drop in the values of the entries of $\underline{x}$ will be observed when transitioning between nonzero entries of $\underline{s}$ and $\underline{x}^n$. Using this information, a threshold can be chosen to separate out $\underline{x}^n$ for use in detection. If a suitable threshold cannot be chosen to separate out $\underline{x}^n$, a second approach can be used. In this approach, $\underline{x}$ can be used instead of $\underline{x}^n$ in the detection algorithm. Since $\underline{s}$ will have few nonzero entries, the statistics of $\underline{x}^n$ will dominate and there will be little effect on the detection results.

Additionally, if $\underline{x}$ can be modeled as a nearly sparse signal or a nearly sparse signal in the presence of noise, the preceding detection technique can still be used, albeit with slight modification. From Section 2.2, we know that for nearly sparse signals or nearly sparse signals in the presence of noise, the reconstruction step in compressive sensing will result in the presence of a large number of zero or near zero valued entries in $\underline{x}$. As a result, we can state

$$\mathbb{P}(x_i = 0|H_0) \ll \mathbb{P}(x_i = 0|H_1). \tag{2.13}$$

for nearly sparse signals and nearly sparse signals in the presence of noise. If we substitute $\underline{x}$ for $\underline{x}^n$ in equations (2.11), compressive sensing can be detected in nearly sparse signals using the decision rule δ_z presented in (2.12).

2.3.2 Distribution-based Detection Scheme

In some scenarios, the forensic investigator will have knowledge about the distribution $\mathcal{F}$ of the noise present during sensing, like the quantization noise [19], or about the distribution $\mathcal{G}$ of the coefficients in a nearly sparse signal. This knowledge can be used as side information to perform improved compressive sensing detection. To develop a detection scheme that makes use of this distribution information, let us examine the case of nearly sparse signals.

Let us assume that a forensic examiner knows that the coefficients of a nearly sparse signal are distributed according to some parametric distribution $\mathcal{G}(\theta)$, where the true value of the parameter θ is unknown. Additionally, assume that the forensic investigator knows an estimator $\hat{\theta}$ for the parameter θ on the basis of i.i.d. realizations of $\mathcal{G}(\theta)$. Under hypothesis H_0, each entry of $\underline{x}$ will be a direct observation of the nearly sparse signal, therefore the entries of $\underline{x}$ will be distributed according to $\mathcal{G}(\theta)$. If $\hat{\theta}$ is calculated using the entries of $\underline{x}$, an appropriately chosen measure of the distance between $\mathcal{G}(\hat{\theta})$ and the normalized histogram of $\underline{x}$ should be small. We know from Section 2.2, however, that under hypothesis H_1 the entries of $\underline{x}$ will no longer be distributed according to $\mathcal{G}(\theta)$. This will cause $\hat{\theta}$ to be an inaccurate estimate of θ if it is calculated from $\underline{x}$ under hypothesis H_1. Now, given an appropriately

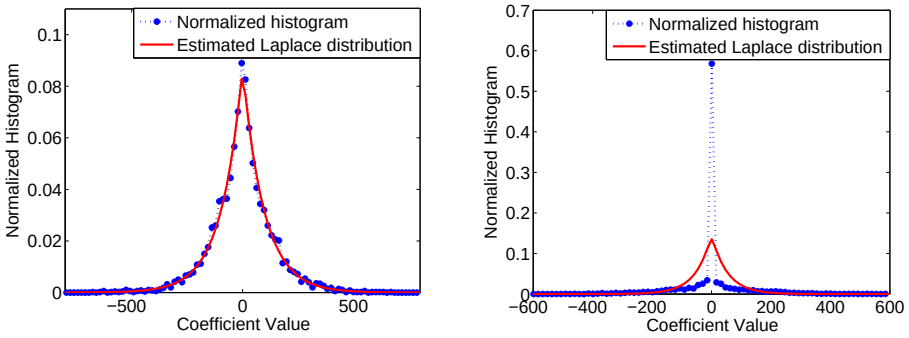


Figure 2.5: Fitting the histogram of the observed signal to the estimated signal distribution. The Laplace distribution was used to generate each sample of the nearly sparse signal. The left figure shows the fitting result when this signal was obtained by traditional sensing, while the right one shows the result for a when the signal was compressively sensed.

chosen distance metric, the distance between $\mathcal{G}(\hat{\theta})$ and the normalized histogram of $\underline{x}$ will be large. This can be seen in Fig. 2.5. As a result, we can measure the strength of compressive sensing fingerprints in $\underline{x}$ by measuring the distance between the normalized histogram of $\underline{x}$ and $\mathcal{G}(\hat{\theta})$.

A problem arises when measuring the distance between these two quantities: $h_k(\underline{x})$ is an estimate of the probability that the value of x_i falls within the k^{th} histogram bin, while $\mathcal{G}(\hat{\theta}, t)$ is the probability that x_i takes the value t. As a result, these two quantities cannot be compared directly by any distance measurement. To resolve this disparity, we integrate $\mathcal{G}(\hat{\theta}, t)$ over each histogram bin to obtain $g(\hat{\theta})$ where

$$g_k(\hat{\theta}) = \int_{b(k-1/2)}^{b(k+1/2)} \mathcal{G}(\hat{\theta}, t)dt \tag{2.14}$$

and b is the width of each histogram bin.

Let $\xi_d(h_k, g_k)$ denote some distance measure between h_k and g_k, such as mean square distance (MSD) or Kullback-Leibler divergence (KL divergence), then, we perform compressive sensing detection using the following decision rule

$$\delta_d = \begin{cases} H_0 & \text{if } \xi_d(h_k, g_k) < \tau_d \\ H_1 & \text{if } \xi_d(h_k, g_k) \geq \tau_d. \end{cases} \tag{2.15}$$

where τ_d is a decision threshold. The choice of the distance measure $\xi_d(h_k, g_k)$ is made based on the performance of this compressive sensing detector in different applications. For example, when detecting compressively sensed images, using mean square error as the distance measure yield the best detection performance. We will discuss this case in the next section.

Besides the conventional distance measures, such as MSD and KL divergence, we also propose their modified versions as the candidates of $\xi_d(h_k, g_k)$. These modified distance measures take into account the particular manner in which compressive sensing changes the distribution of the entries in $\underline{x}$. Take the KL divergence measure as an example. Since compressive sensing dramatically increases the kurtosis of the distribution of the entries in $\underline{x}$, the most forensically significant differences between h and g should occur around $k = 0$. As a result, we modify the KL divergence to measure the strength of compressive sensing fingerprints as follows

$$\xi_d(h_k, g_k) = \sum_k w_k \ln \frac{h_k}{g_k}, \tag{2.16}$$

where w_k is a normalized set of weights used to emphasize differences in the forensically significant region around $k = 0$. Since we wish to weight the regions around $k = 0$ more heavily, we construct the weighting function using a Laplace distribu-

tion. Other distributions obeying power law decay may also be good candidates. Given that the weights are discrete, we integrate the Laplace distribution over each histogram bin to obtain the weighting function as follows,

$$w_k = \begin{cases} 1 - e^{-\nu b/2}\cosh(\nu k) & \text{if } k = 0, \\ e^{-\nu|k|}\sinh(\nu b/2) & \text{otherwise,} \end{cases} \tag{2.17}$$

where the parameter ν is chosen to be

$$\nu = \frac{\beta n}{\sum_{i=1}^{n} |x_i|}, \tag{2.18}$$

and where β is a user specified parameter that adjusts the size of the forensically significant region. Experimentally, we have found that $\beta = 100$ yields desirable results. Similar modifications can be applied on other conventional distance measures.

If the signal being examined can be modeled as a sparse signal in the presence of noise and the forensic investigator has a parametric model $\mathcal{F}(\theta)$ of the noise distribution, the detection technique presented above can be used, only with slight modifications. Since the noise distribution rather than the signal distribution is known, $\mathcal{F}$ should be substituted for $\mathcal{G}$ in (2.14). Additionally, $\hat{\theta}$ should be calculated using $\underline{x}^n$ and the histogram of $\underline{x}^n$ should be substituted for $h(\underline{x})$ in (2.16). If the signal is more appropriately modeled as a nearly sparse signal in the presence of noise, the distribution of $\underline{x}$ is given by the convolution of $\mathcal{G}$ and $\mathcal{F}$. If the noise distribution is unknown or if $\mathcal{G} * \mathcal{F}$ is difficult or intractable, the noise distribution can be ignored when performing compressive sensing detection as long as the noise power is sufficiently low.

We note that, although only the original signal's distribution is explicitly used

in this distribution-based detection scheme, our model for compressively sensed signals has also been implicitly applied when designing the detector. Specifically, both detection schemes are designed based on the assumption that the distribution of a compressively sensed signal has much more kurtosis than that of a traditionally sensed signal. While this is enough for identifying compressively sensed signals from traditionally sensed signals, more explicit models for the distribution of compressively sensed signals can be proposed for particular applications where more complicated detection scenarios exist. We will discuss this in detail for images in the next section.

2.4 Detecting Compressive Sensing in Digital Images

While the compressive sensing detection techniques proposed in Section 2.3 can be used on a wide variety of signals, in some scenarios it is desirable to create a compressive sensing detection technique specifically tailored to a particular class of signals. This is the case for digital images.

An image's compression history can reveal important information about how an image was captured and stored. It can also reveal important information about the device used to capture an image [88]. As a result, a variety of techniques have been developed to determine if an image was previously compressed. Fingerprints left by compressive sensing, however, can be mistaken for traditional image compression fingerprints by existing forensic techniques such as those proposed by Lin et al. [58] and Luo et al. [61]. As a result, when we are given a compressively sensed

and reconstructed image, it may be easily misidentified as a traditionally sensed and compressed image. In this section, we propose a forensic technique specifically designed to both detect evidence of compressive sensing in digital images and to differentiate compressive sensing fingerprints from those left by traditional forms of image compression.

2.4.1 Compressive Sensing Fingerprints in Digital Images

Since the pixel values of an image do not form a sparse signal, digital images may not initially seem well suited for compressive sensing. It is well known, however, that within each subband, the set of discrete wavelet transform (DWT) coefficients of a natural image are sparse. As a result, compressive sensing reconstruction is often performed on images in the wavelet domain.

From our discussion of compressive sensing fingerprints in Section 2.2, we would naturally expect an impulsive peak to occur at zero in the DWT coefficient distribution of a compressively sensed image. While this is true after the compressively sensed DWT coefficients are reconstructed, the inverse DWT of the image must be performed and the resulting pixel values must be projected back into the set $\{0, \ldots, 255\}$ of allowable pixel values. This will introduce a small but nontrivial amount of noise into the DWT coefficients when DWT is applied to the image again to extract the coefficients. As a result, the peak in the image's DWT coefficient distribution at zero will no longer correspond to an impulse. Though the peak will be slightly smoothed by this noise source, the DWT coefficient distribution of a

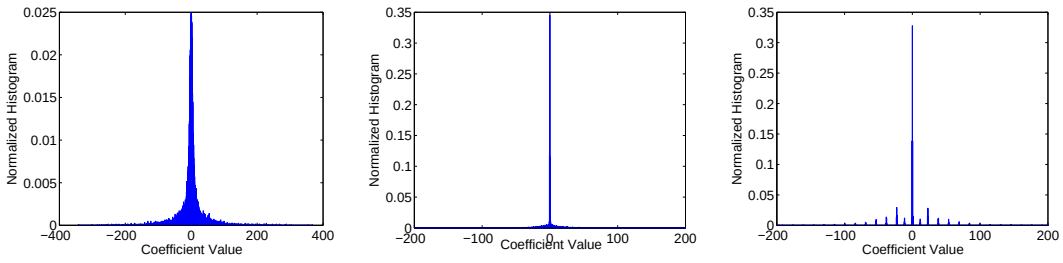


Figure 2.6: Histograms of DWT coefficients taken from uncompressed Lena (left), the same image after JPEG 2000 compression (right), and the reconstructed compressively sensed Lena (center).

compressively sensed image will still exhibit a large degree of kurtosis, as can be seen in Fig. 2.6. We use this characteristic feature of a compressively sensed image's DWT coefficient distribution as the fingerprints.

Wavelet-based image compression techniques such as JPEG 2000 and SPIHT also introduce fingerprints in an image's DWT coefficient distribution. During compression, these techniques use a bit-plane encoder to store the most significant digits of each DWT coefficient in a subband. This has the same effect as quantizing each DWT coefficient. As a result, the DWT coefficients in an image compressed using a wavelet-based technique will tightly cluster around certain values, forming a series of peaks in the DWT coefficient distribution that can be seen in the rightmost plot in Fig. 2.6. These peaks are the fingerprints of wavelet based image compression. Since the most prominent peak occurs at zero, compressive sensing fingerprints and wavelet-based compression fingerprints can easily be confused by existing detectors.

To demonstrate that compression history detection techniques can mistake compressive sensing fingerprints for JPEG 2000 compression fingerprints, we per-

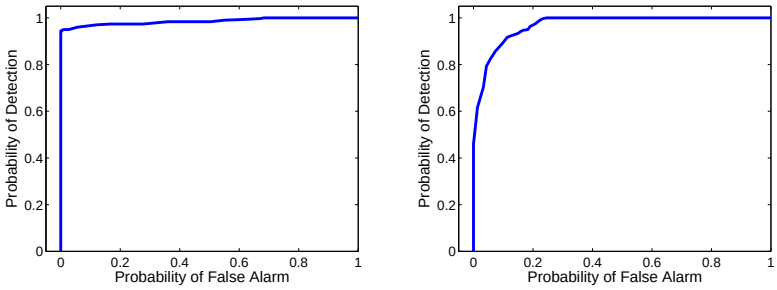


Figure 2.7: ROC curves obtained by using the image compression detection technique in [58] to identify JPEG 2000 compression in a set of unaltered and JPEG 2000 compressed images (left) and a set of unaltered and compressively sensed images (right). In the right figure "false alarms" correspond only to unaltered images misclassified as JPEG 2000 compressed. Since there is no JPEG 2000 compressed image in the seconde test set, the results in the right figure demonstrate that compressive sensing can be easily misidentified as JPEG 2000 compression.

formed an experiment using the compression history detection technique proposed in [58]. When performing this experiment, we used the Uncompressed Colour Image Database (UCID) [82] to create a testing database of 300 unaltered images, 300 JPEG 2000 compressed images, and 300 compressively sensed images. First, we evaluated the baseline performance of the wavelet-based compression detection technique from [58] by using it to distinguish between the set of unaltered and JPEG 2000 compressed images. An ROC curve showing the results of this experiment is displayed in the left figure of Fig. 2.7, which shows that this technique can reliably detect wavelet-based compression. Next, we used this technique to identify evidence of JPEG 2000 compression in the set of compressively sensed and unaltered images. Since none of the images in this second experiment were compressed

using JPEG 2000, we would expect the detector to find no evidence of JPEG 2000 compression. An ROC curve showing the results of this experiment is displayed in the right figure of Fig. 2.7. "false alarms" correspond only to unaltered images misclassified as JPEG 2000 compressed, and "detections" correspond to compressively sensed images been identified as JPEG 2000 compressed images. These results show that compressively sensed images can be easily misidentified as images that have undergone JPEG 2000 compression by existing forensic techniques. This reinforces the need for a technique to distinguish between compressive sensing and traditional wavelet-based compression.

Moreover, while the proposed universal detection schemes in Section 2.3 can be used on images to distinguish compressively sensed images from traditionally sensed images, their performance may be affected when traditionally sensed but wavelet-based compressed images are involved in the acquisition detection analysis. To demonstrate this, we used the universal detector proposed in section 2.3.2 to differentiate between compressively sensed images and both uncompressed traditionally sensed images as well as traditionally sensed images that have been compressed using JPEG 2000. The results of this experiment are shown in Fig. 2.8. The left figure demonstrates that our proposed general compressive sensing detection scheme can be successfully used on image signals. While the right figure shows the degradation of this scheme's performance when traditionally sensed but JPEG 2000 compressed images are involved in the analysis. Therefore, in order to determine the acquisition process of an image signal and identify compressive sensing, we need more specific models for compressively sensed images to distinguish them from traditionally

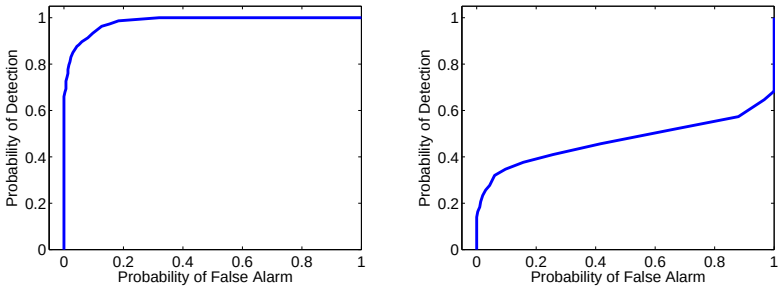


Figure 2.8: ROC curves obtained by using the proposed scheme in Section 2.3.2 to identify compressively sensed images from traditionally sensed images (left) and to identify compressively sensed images from traditionally sensed but JPEG 2000 compressed images (right).

sensed but wavelet-based compressed images.

2.4.2 DWT Coefficient Distribution Models

Because both compressive sensing fingerprints and wavelet-based compression fingerprints present themselves in an image's DWT coefficient distribution, we must adopt a set of models for an image's DWT coefficient distribution in order to develop our forensic technique. Let X be a random variable representing the value of a DWT coefficient in a particular subband of an image. For uncompressed images, we model the distribution of X using the Laplace distribution [87]

$$f_X(x) = \frac{\lambda_0}{2} e^{-\lambda_0 |x|}. \tag{2.19}$$

Since traditional DWT-based image compression is equivalent to nonuniform quantization [87], we then model the DWT coefficient distribution of an image that has

undergone traditional wavelet-based compression as

$$\mathbf{P}[X = q] = \int_{q-\triangle_q}^{q+\triangle_q} \frac{\lambda_0}{2} e^{-\lambda_0|x|} dx, \tag{2.20}$$

where $q \in \mathbb{Z}$ and $\triangle_q$ is half of the width of the quantization interval that maps DWT coefficients to q.

When examining compressively sensed images, we must account for the noise introduced into the image's DWT coefficients described in Section 2.4.1. Since this noise will slightly smooth out the impulsive spike that we would expect to occur in the distribution of X at zero, we instead model the DWT coefficients of a compressively sensed image using a Laplace mixture distribution [24]

$$f_X(x) = \omega_1 \frac{\lambda_1}{2} e^{-\lambda_1|x|} + \omega_2 \frac{\lambda_2}{2} e^{-\lambda_2|x|} \tag{2.21}$$

where $\omega_1 + \omega_2 = 1$ and $0 < \lambda_1 < 1 < \lambda_2$. Fig. 2.9 shows an example of a compressively sensed image's DWT coefficient histogram fit to both a Laplace and a Laplace mixture distribution. We can see from this figure that an appropriately chosen Laplace mixture distribution very accurately models the compressively sensed image's DWT coefficient distribution.

2.4.3 Compressive Sensing Detection

Because the fingerprints left by traditional wavelet-based compression techniques can be confused with the compressive sensing fingerprints, we propose performing compressive sensing detection on images in two steps [24]. In the first step, we separate unaltered traditionally sensed images from those that are either traditionally compressed or compressively sensed. In the second step, we differentiate

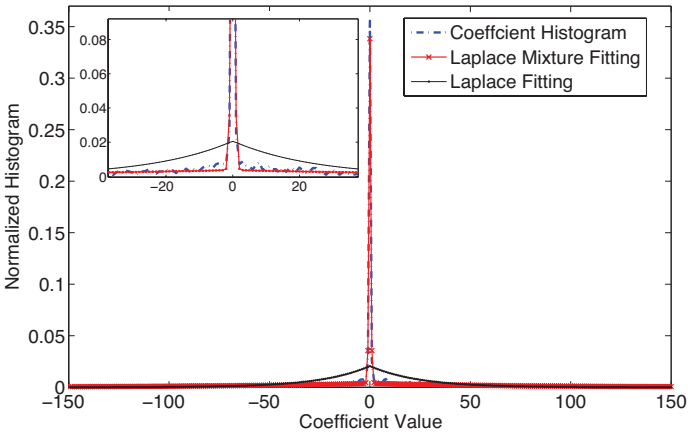


Figure 2.9: Fit the coefficient histogram of compressively sensed Lena with both Laplace model and Laplace mixture model. Coefficients are taken from the third subband after 6-level DWT decomposition with wavelet basis 'bior4.4'.

between compressively sensed images and those that have traditionally undergone wavelet-based compression.

Step 1 - Identify Uncompressed Traditionally Sensed Images

The goal of the first step of our compressive sensing detection scheme is to remove uncompressed traditionally sensed images from further examination. This step is equivalent to differentiating between the following two hypotheses

$$\begin{aligned} &H_0\text{: The image is uncompressed and traditionally sensed,} \\ &H_1\text{: The image is traditionally compressed or compressively sensed.} \end{aligned} \tag{2.22}$$

where hypothesis H_1 is a composite hypothesis. To accomplish this, we exploit the fact that the DWT coefficient distributions of both compressively sensed images and traditionally compressed images will significantly differ from the Laplace distribution.

We begin by assuming that hypothesis H_0 is correct. Under this assumption,

the parameter λ_0 in (2.19) can be estimated for a particular subband of an image's DWT coefficients using the maximum likelihood estimator

$$\hat{\lambda}_0 = \frac{N}{\sum_{i=1}^{N} |x_i|}, \tag{2.23}$$

where each x_i represents a DWT coefficient in the subband being examined and N is the number of DWT coefficients in the subband. Once the estimate $\hat{\lambda}_0$ is obtained, we use $\hat{\lambda}_0$ and (2.19) to calculate the expected histogram g_k^{unalt} according to (2.14). We then measure the mean squared distance (MSD) between the observed histogram of DWT coefficients h_k and g_k^{unalt} according to the formula

$$MSD_1 = \frac{1}{B} \sum_k \left(h_k - g_k^{unalt}\right)^2, \tag{2.24}$$

where B is the total number of histogram bins.

We note that this step is an application of our distribution-based detection scheme proposed in section 2.3.2. MSD is chosen instead of KL divergence to avoid the "divide by zero" problem when calculating the KL divergence.

If the MSD between h_k and g_k^{unalt} is sufficiently large, we conclude that an image's DWT coefficient histogram cannot be modeled using (2.19), therefore the image either has undergone wavelet-based compression or has been compressively sensed. As a result, we differentiate between the hypotheses in (2.22) using the decision rule

$$\delta_1 = \begin{cases} H_0: \text{ If } MSD_1 < \tau_1 \\ H_1: \text{ If } MSD_1 \geq \tau_1, \end{cases} \tag{2.25}$$

where τ_1 is the decision threshold. If δ_1 returns a decision of H_1 for an image, then we proceed to step 2 of our detection process.

Step 2 - Detect Compressive Sensing

Once we have decided that an image has been either traditionally compressed or compressively sensed, we must differentiate between these two possibilities. In the second step of our detector, we frame this problem as deciding between the hypotheses

$$\begin{aligned} H_0&: \text{The image has undergone wavelet-based compression,} \\ H_1&: \text{The image was compressively sensed.} \end{aligned} \tag{2.26}$$

We know that under hypothesis H_1, an image's DWT coefficient distribution will be given by (2.21). As a result, we can identify compressively sensed images by determining how well the distribution of an image's DWT coefficients within a subband fits a Laplace mixture distribution.

To do this, we first estimate the parameters in the parameter set $\theta = \{\omega_1, \omega_2, \lambda_1, \lambda_2\}$ using the expectation maximization (EM) algorithm [67]. Let Z_i be latent random variables that denote which component of the Laplace mixture distribution each DWT coefficient x_i originates. As a result, we can write the following equations:

$$f_{X_i}(x_i|Z_i = 1) = \frac{\lambda_1}{2} e^{-\lambda_1|x_i|}, \tag{2.27}$$

$$f_{X_i}(x_i|Z_i = 2) = \frac{\lambda_2}{2} e^{-\lambda_2|x_i|}, \tag{2.28}$$

$$\mathbf{P}[Z_i = 1] = \omega_1 \text{ and } \mathbf{P}[Z_i = 2] = \omega_2. \tag{2.29}$$

At the t^{th} iteration of the EM algorithm, the updated estimates of the parameters

in the parameter set are given by the equations

$$w_j^{(t+1)} = \frac{1}{n}\sum_{i=1}^{N} T_{j,i}^{(t)} \quad j = 1, 2 \tag{2.30}$$

$$\lambda_j^{(t+1)} = \frac{\sum_{i=1}^{N} T_{j,i}^{(t)}}{\sum_{i=1}^{N} T_{j,i}^{(t)}|x_i|} \quad j = 1, 2 \tag{2.31}$$

where

$$T_{j,i}^{(t)} = \frac{w_j^{(t)}\lambda_j^{(t)} e^{-\lambda_j^{(t)}|x_i|}}{\omega_1^{(t)}\lambda_1^{(t)} e^{-\lambda_1^{(t)}|x_i|} + \omega_2^{(t)}\lambda_2^{(t)} e^{-\lambda_2^{(t)}|x_i|}}. \tag{2.32}$$

The EM algorithm's iterations are terminated after either the maximized log-likelihood ratio

$$\max_{\theta} Q(\theta|\theta^{(t)}) = \sum_{i=1}^{N}\sum_{j=1}^{2} T_{j,i}^{(t)}\left[\ln\left(\omega_j^{(t+1)}\lambda_j^{(t+1)}/2\right) - \lambda_j^{(t+1)}|x_i|\right].$$

converges or a fixed number of iterations have been reached.

After the values of $\omega_1, \omega_2, \lambda_1$, and λ_2 have been estimated, we compute the expected DWT coefficient histogram g_k^{cs} under hypothesis H_1 using (2.14). Next, we calculate the MSD between the g_k^{cs} and the observed histogram of DWT coefficients h_k

$$MSD_2 = \frac{1}{B}\sum_{k}\left(h_k - g_k^{cs}\right)^2, \tag{2.33}$$

where B is the total number of histogram bins. Finally, we perform compressive sensing detection according to the decision rule

$$\delta_2 = \begin{cases} H_0: \text{ If } MSD_2 > \tau_2 \\ H_1: \text{ If } MSD_2 \leq \tau_2, \end{cases} \tag{2.34}$$

where τ_2 is a decision threshold.

2.5 Measurement Number Estimation

Once a signal has been identified as compressively sensed, a forensic investigator may wish to ascertain additional information about how the signal was captured. One significant piece of information is the number of compressive measurements that were used to acquire the signal. In this section, we propose a technique to estimate the number of compressive measurements obtained when sensing a signal.

When a compressively sensed signal is reconstructed by solving (2.3), the sparsest solution $\underline{x}$ such that $\Phi\,\underline{x} = \underline{y}$ is chosen. Since the values of $\underline{x}$ can be thought of as weights for the column vectors of Φ, and $\underline{y}$ is obtained also by weighted sum of these vectors with non-sparse weighting values, it seems natural that the sparsity of the reconstructed signal will be closely related to dimension of the column vectors of Φ, which is approximated to be the rank of Φ, i.e., the number of compressive measurements. In fact, we are able to prove that the relationship between the number of compressive measurements and the number of zeros in the reconstructed signal is given by the relationship stated below in Theorem 1.

Theorem 1. *Let $\underline{y}$ be a vector of m compressive measurements obtained by compressively sensing a signal that fits one of the three signal models proposed in Section 2.1.2. Assume that the noise, if applicable, is continuously distributed. Additionally, let the m by n sensing matrix Φ have orthonormal row vectors selected uniformly at random from an orthonormal vector set in $\mathbb{R}^n$. If the reconstructed signal $\underline{x}$ is obtained by solving the l_1 minimization problem*

$$\min_{\tilde{\underline{x}}} ||\tilde{\underline{x}}||_{l_1} \quad s.t. \;\; \Phi\tilde{\underline{x}} = \underline{y}, \tag{2.35}$$

then with probability close to one, $\underline{x}$ will have m non-zero coefficients. As a result, the number of compressive measurements is given by

$$m = n - \Lambda_0(\underline{x}), \tag{2.36}$$

where $\Lambda_0(\underline{x})$ denotes the number of zero valued entries in $\underline{x}$.

Proof. We prove this theorem by deriving a lower and upper bound on $n - \Lambda_0(\underline{x})$ respectively, then showing that the only value of $n-\Lambda_0(\underline{x})$ that satisfies both bounds is m.

To derive the lower bound, we begin by defining vector space V as the linear span of the column vectors $\underline{\phi}_1, \underline{\phi}_2, \ldots, \underline{\phi}_n$ of the sensing matrix Φ. Since Φ has orthogonal row vectors, it is full rank. Thus, $\dim(V) = \dim\{\underline{\phi}_1, \underline{\phi}_2, \ldots, \underline{\phi}_n\} = m$. Next, we define the dimension of an m length vector $\underline{v}$ on space V as the size of the smallest subset of $\{\underline{\phi}_1, \underline{\phi}_2, \ldots, \underline{\phi}_n\}$ whose linear span contains $\underline{v}$.

The compressive measurements $\underline{y}$ can be expressed as $\underline{y} = \sum_{i=1}^{n} \underline{\phi}_i s_i$, where $\underline{s}$ is the signal being acquired by compressive sensing. If $\underline{s}$ fits any of the signal models in Section 2.1.2, then the dimension of $\underline{y}$ is equal to the dimension of V with probability close to one. Specifically, in the case of signals corrupted by environmental noise and nearly sparse signals, either the noise or the nature of the signal itself will cause each entry of $\underline{s}$ nonzero. Otherwise, if the signal is corrupted by measurement noise, then the independent white noise added to the compressive measurements will cause $\underline{y}$ to lie in the span of any subset of V of size $m - 1$ or less with probability nearly zero.

Because the reconstructed signal $\underline{x}$ is just another decomposition of $\underline{y}$ on space

V, the number of non-zero entries in $\underline{x}$ can not be less than the dimension of $\underline{y}$ on this space. Thus,

$$n - \Lambda_0(\underline{x}) \geq \dim(\underline{y}) = \dim(V) = m. \tag{2.37}$$

To derive the upper bound, we reformulate (2.35) as the following equivalent problem [17]

$$\min_{\tilde{\underline{z}}} \underline{1}^T \tilde{\underline{z}}, \quad s.t. \;\; A\tilde{\underline{z}} = \underline{y}, \;\; \tilde{\underline{z}} \geq 0. \tag{2.38}$$

where $\underline{1}$ denotes a column vector of length $2n$ of all ones and $A = (\Phi, -\Phi)$ is of size $m \times 2n$. If the solution to (2.38) is partitioned into two vectors of equal length such that $\underline{z} = (\underline{u}^T, \underline{v}^T)^T$, then the solution to (2.35) can be expressed as $\underline{x} = \underline{u} - \underline{v}$.

By examining this intermediate problem, the following lemma and corollary can be proved by using Karush-Kuhn-Tucker conditions [51].

Lemma 1. *Let $\underline{z}'$ denote the sparsest solution of problem* (2.38)*, i.e., the one with smallest number of non-zero coefficients. Then*

$$n - \Lambda_0(\underline{z}') \leq m. \tag{2.39}$$

Corollary 1. *For any solution $\underline{z}$ of* (2.38)*, the corresponding solution $\underline{x}$ for* (2.35) *will have the same number of non-zero coefficients with $\underline{z}$.*

Given these two results, we conclude our proof by recalling that the solution to (2.35) is unique (see *Theorem 1.1* in [12]), so that the sparsest solution $\underline{x}'$ to (2.35) is the only solution, i.e., $\underline{x} = \underline{x}'$. Therefore, $n - \Lambda_0(\underline{x}) = n - \Lambda_0(\underline{z}') \leq m$. Combining this result with (2.37), we conclude that $n - \Lambda_0(\underline{x}) = m$, thus Theorem 1 is proved. □

In practice, a number of iterative techniques are often used to solve (2.35). Since these techniques are typically terminated after the difference between two iterations is sufficiently small or a fixed number of iterations has been reached, the solution yielded by these techniques will often differ slightly from the optimal solution. As a result, several values of $\underline{x}$ that would ideally be zero will instead take small nonzero values. To compensate for this effect, we instead count the number of entries $\Lambda_\zeta(\underline{x})$ that fall within a ball of radius ζ around zero. Our measurement number estimator for the observed signal $\underline{x}$ is defined as follows:

$$\hat{m} = n - \Lambda_\zeta(\underline{x}), \tag{2.40}$$

where $\zeta = ||\breve{\underline{x}}||_\infty/\rho$. If the signal $\underline{x}$ is modeled as a sparse signal in noise, $\breve{\underline{x}}$ is taken as the noise component, otherwise $\breve{\underline{x}} = \underline{x}$. The choice of ρ depends on how accurate the reconstruction is. For example, in the ideal where the iteration in simulation can go to infinity, then $\rho \to \infty$ and $\zeta \to 0$. In our simulations, we have experimentall observed that $\rho = 100$ yields desirable performance.

2.6 Simulations and Results

To verify the effectiveness of our proposed forensic techniques, we have evaluated their performance through a series of experiments. In this section, we present the results of these experiments and show that our proposed techniques can reliably detect the use of compressive sensing. We first evaluate the ability of our forensic techniques to identify compressive sensing in sparse signals in the presence of noise, nearly sparse signals, and nearly sparse signals in the presence of noise. We then

evaluate the performance of our compressive sensing detection technique for images and our technique to estimate the number of compressive measurements used to acquire a signal.

2.6.1 Sparse Signals in the Presence of Noise

To evaluate the ability of both the zero ratio detector and the distribution-based detector to identify compressive sensing in sparse signals in the presence of noise, we first created a database of testing signals. This database consisted of 200 compressively sensed sparse signals in the presence of environmental noise, 200 compressively sensed sparse signals in the presence of measurement noise, and 200 sparse signals in the presence of additive noise which were not compressively sensed. Each signal was created by first randomly generating a sparse signal of length $N = 1000$ with 20 nonzero entries. For each nonzero entry, its location was chosen uniformly at random and its value was drawn from a Gaussian distribution with a mean of 10 and unit variance. We then corrupted each signal with additive Gaussian noise distributed $\mathcal{N}(0, 0.1)$. For signals which were not compressively sensed, we added the noise directly to the sparse signal to obtain the observed signal. For compressively sensed signals corrupted by environmental noise, we added the noise to the sparse signal, then performed M compressive measurements. For signals corrupted by measurement noise, we first obtained M compressive measurements of the sparse signal, then added the Gaussian noise to each compressive measurement. Each compressively sensed signal was reconstructed using the basis pursuit de-noising al-

gorithm [17]. We obtain the noise component of the observed signal by excluding the 20 entries that have the largest magnitudes, since these likely correspond to the nonzero components of the sparse signal. We then used both detection techniques to determine if each signal was compressively sensed.

In our first set of experiments, we evaluated the performance of both detection techniques as the ratio of the number of compressive measurements to the total signal length was varied from $M/N = 0.1$ to 0.9 in increments of 0.1. For distribution-based detector, the modified KL divergence was chosen as the distance measure for it performs better than other distance measures do. When performing these experiments, we varied the decision thresholds of each detector over a range of values. For each threshold value, we determined the associated probabilities of detection P_d and false alarm P_f by calculating the percentage of compressively sensed signals that were correctly identified and the percentage of signals that were incorrectly identified as compressively sensed respectively. We then used these probabilities to construct a set of ROC curves showing the performance of each detector. Selected ROC curves showing the performance of both detectors for $M/N = 0.1$, 0.4, and 0.9 are shown in Fig.s 2.10(a) through (c).

From the full set of ROC curves, we found that both detectors achieved perfect detection, i.e. $P_d = 100\%$ with $P_f = 0\%$, for $M/N \leq 0.8$. When M/N reaches 0.9, both detectors can still identify compressive sensing with $P_d = 99\%$ at a $P_f \leq 5\%$. Since in most real world scenarios compressive sensing will be applied with M/N less than 0.5, these results show that both techniques perform strongly under realistic conditions. Furthermore, we can see from Fig. 2.10(c) that the distribution-

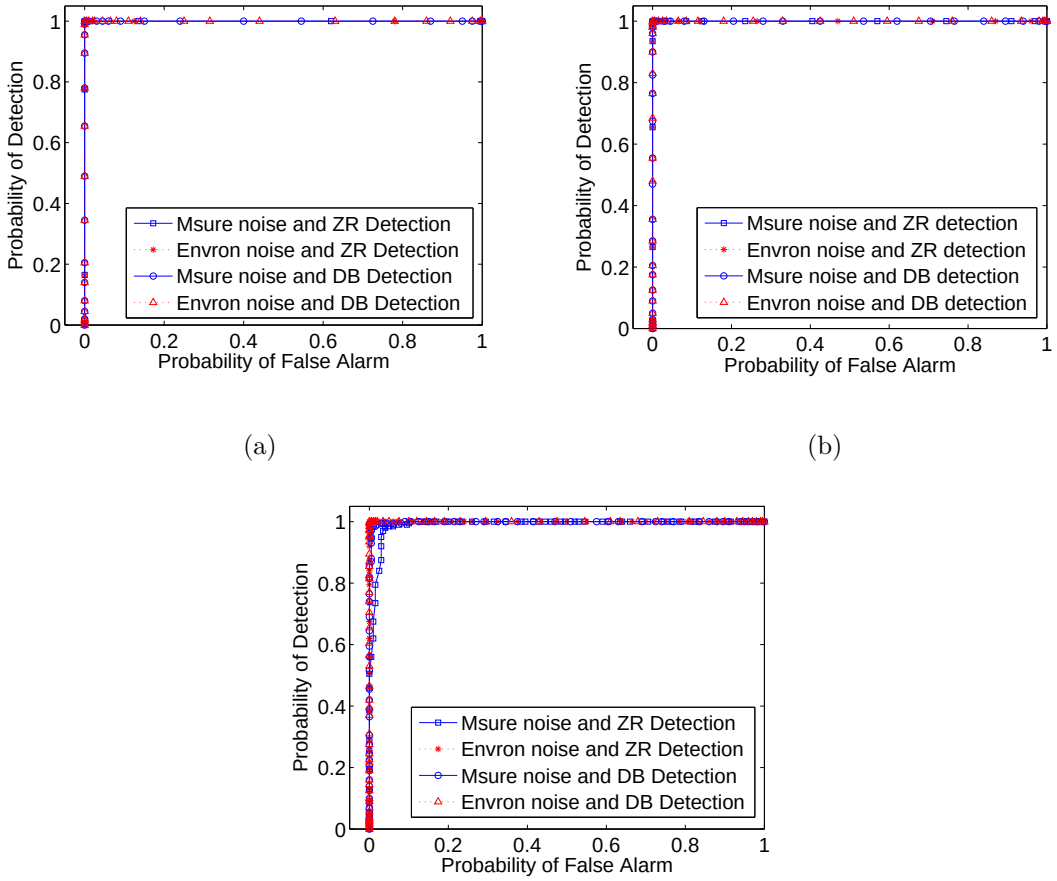


Figure 2.10: ROC curves of zero ratio detector and distribution-based detector on signals modeled as sparse signals in the presence of noise for (a) M/N=0.1, (b) M/N=0.4 and (c) M/N=0.9. 'Msure' is short for measurement and 'Envron' is short for environment. 'ZR' denotes the zero ratio detector and 'DB' denotes the distribution-based detector.

based detector outperforms the zero ratio detector because the forensic investigator is able to make use of additional information about the noise's distribution. We also note that the performance of our detectors decrease as M increases because with more compressive measurements, the noise can be accurately reconstructed. Since compressive sensing fingerprints manifest themselves as changes in the noise distribution, this impedes compressive sensing detection. Nevertheless, our results show that compressive sensing detection can be performed with a high degree of accuracy under realistic values of M/N.

Next, we evaluated the robustness of both detectors to different signal and noise powers, as well as different noise distributions. To evaluate the performance with different signal and noise powers, we fixed the number of compressive measurements so that $M/N = 0.5$. This was done because $M/N = 0.5$ is typically an upper bound in real world applications [63], therefore it provides a lower bound on the performance of both detectors in realistic scenarios. We then repeated the previous experiments using the same noise power with signal powers of 10, 100, and 1000, and while using the same signal power with noise powers of 0.1, 1, and 10. For each of these experiments, both detectors achieved $P_d = 100\%$ at a false alarm rate of $P_f = 0\%$. These results show that both detectors can perform strongly under a variety of signal and noise powers. Next, we kept $M/N = 0.5$ and performed compressive sensing detection when each signal corrupted by noise from the exponential, Laplace, Gaussian, uniform and Rayleigh distributions. Again, under each scenario both detectors were able to acheive $P_d = 100\%$ at a false alarm rate of $P_f = 0\%$. Taken together with our previous results, these results show that both our zero ratio

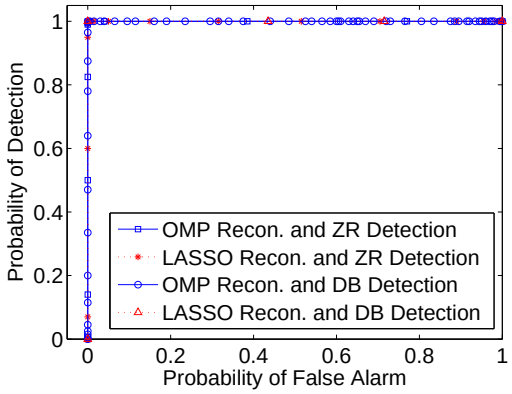


Figure 2.11: ROC curves of zero ratio detector and distribution-based detector on signals modeled as sparse signals in the presence of noise when different reconstruction algorithms were used.

detector and distribution-based detector can be used to identify compressive sensing in sparse signals corrupted by noise under a wide range of conditions.

In addition, since several different algorithms are available to reconstruct a compressively sensed signal, we performed a set of experiments to demonstrate the robustness of our compressive sensing detection technique to different reconstruction algorithms. In these experiments, we used used both orthogonal matching pursuit (OMP) [97] and the LASSO error variation minimization reconstruction algorithm [95] to reconstruct the compressively sensed signals. We then repeated our first set experiments, this time setting $M/N = 0.5$. ROC curves obtained from the results of these experiments are shown in Fig. 2.11. These results demonstrate that both of our detectors can identify compressive sensing regardless of the reconstruction algorithm.

2.6.2 Nearly Sparse Signals and Nearly Sparse Signals in the Presence of Noise

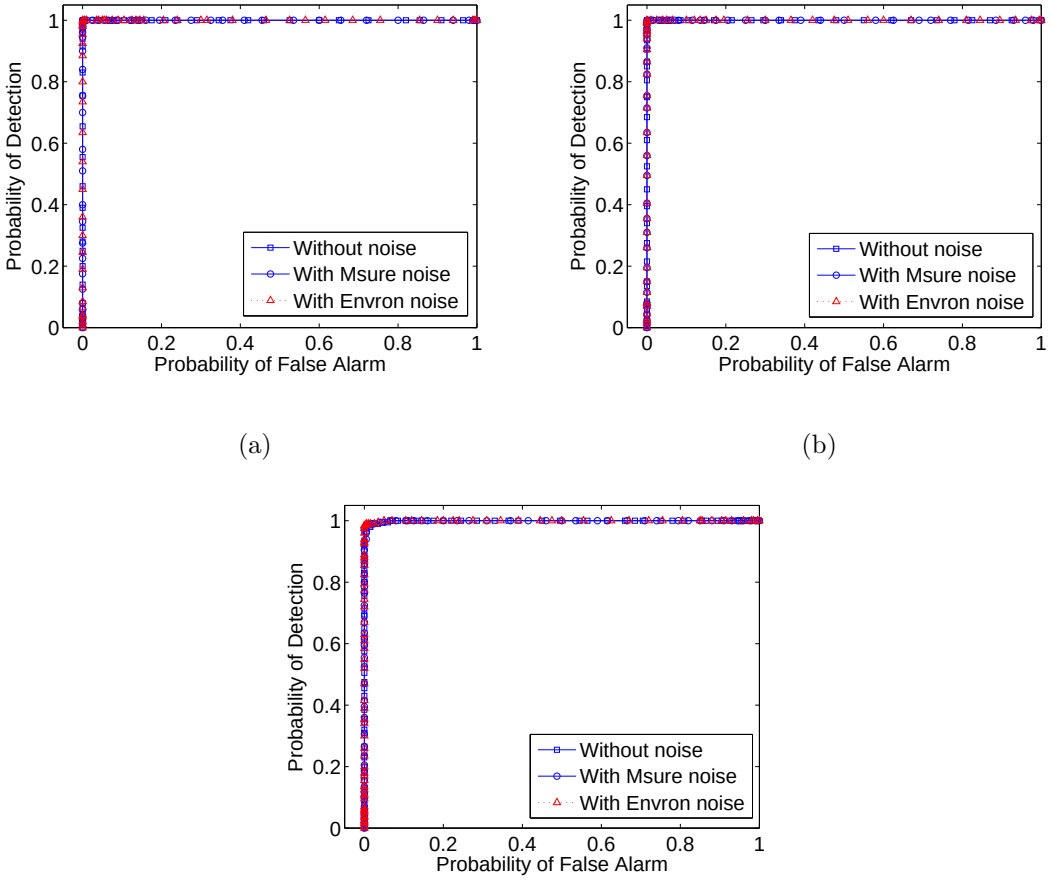
For nearly sparse signals and nearly sparse signals in the presence of noise, we evaluated our distribution-based detector's ability to identify compressive sensing. To do this we created a testing database of 1000 signals consisting of 200 of each of the following types of signals; compressively sensed nearly sparse signals, nearly sparse signals which were not compressively sensed, compressively sensed nearly sparse signals corrupted by environmental noise, compressively sensed nearly sparse signals corrupted by measurement noise, and nearly sparse signals corrupted by additive noise which were not compressively sensed.

Each signal was generated by first creating a nearly sparse signal of length $N = 1000$ whose entries were drawn from a Laplace distribution with variance 10^4. The Laplace distribution was chosen because it is commonly used to model the coefficients of several nearly sparse signals [46, 54]. For compressively sensed nearly sparse signals, we performed M compressive measurements of the signal, then reconstructed it. For compressively sensed nearly sparse signals in the presence of noise, we applied zero mean additive Gaussian noise with variance 10 to either the signal or the M compressive measurements, then performed reconstruction using the basis pursuit de-noising algorithm. To create nearly sparse signals in noise which were not compressively sensed, we added zero mean Gaussian noise with variance 10 to the nearly sparse signal. We then used our distribution-based detector to determine if each signal had been compressively sensed.

In our first set of experiments on these signals, we varied the ratio of the number of compressive measurements to the signal length from $M/N = 0.1$ to 0.9 in steps of 0.1 as was done in Section 2.6.1. We evaluated our distribution-based detector's performance by varying its decision threshold over a range of values, calculating the corresponding P_d and P_f for each threshold value, then creating a set of ROC curves. Selected ROC curves for $M/N = 0.1$, 0.4, and 0.9 are shown in Fig. 2.12.

From the full set of ROC curves we found that when $M/N \leq 0.8$, our distribution-based detector could achieve a probability of detection of $P_d = 100\%$ with $P_f = 0\%$ for both nearly sparse signals and nearly sparse signals in the presence of either type of noise. When M/N was increased to 0.9, our detector was able to achieve a performance of $P_d = 99\%$ with $P_f \leq 3\%$ for all cases. These results show that our distribution-based detector can accuratley identify compressively sensed nearly sparse signals and nearly sparse signals in noise for realistic values of M/N.

Next, we evaluated our distribution-based detector's robustness when performing compressive sensing detection on nearly sparse signals and nearly sparse signals in noise. To do this, we performed a series of experiments in which we fixed M/N at 0.5 as was done in Section 2.6.1, then varied the signal variance as well as the noise power and distribution when appropriate. For nearly sparse signals, we allowed the signal variance to take values of 10^{-4}, 1 and 10^4. In each case, the detector achieved $P_d = 100\%$ with $P_f = 0\%$, i.e. perfect detection. For nearly sparse signals in the presence of noise, we repeated experiments using signal powers of 10^3, 10^4 and 10^5 and with noise powers of 0.1, 1 and 10. Additionally, we performed experiments



(a) (b)

(c)

Figure 2.12: ROC curves of distribution-based detection on nearly sparse signals and nearly sparse signals in the presence of noise for (a) M/N=0.1, (b) M/N=0.4 and (c) M/N=0.9. 'Msure' is short for measurement and 'Envron' is short for environment.

in which we fixed the signal power at 10 and varied the noise distribution between the Gaussian, Rayleigh, Laplace, exponential and uniform distributions. In each of these experiments, our detector was able to achieve $P_d = 100\%$ with $P_f = 0\%$. These results show that our detector can be used to reliably identfy compressive sensing in both nearly sparse signals and nearly sparse signals in the presence of noise under a wide variety of conditions.

2.6.3 Images

To evaluate the performance of our compressive sensing detection technique for images, we first created a testing database of images. For each experiment, we used 300 unaltered images, 300 JPEG 2000 compressed images, and 300 compressively sensed images from the UCID database [82]. Each image in this database has size of 512×256 pixels. During JPEG 2000 compression and compressive sensing reconstruction, the 'bior4.4' DWT basis was used to perform the discrete wavelet transform of each image. To fairly evaluate our detector, during each set of experiments the compression quality factor for the JPEG 2000 images and the number of compressive measurements for the compressively sensed images were chosen so that both sets of images had the same average PSNR. For example, the average PSNRs for $M/N = 0.67$ and $M/N = 0.25$ are 36dB and 26dB, respectively.

In our first experiment, when performing compressive sensing we chose the compression ratio to be $N/M = 4$. After creating an appropriately compressed set of JPEG 2000 images, we classified each image in the testing database using

our two-step image compressive sensing detection technique. When doing this, we obtained classification results using DWT subbands 2 through 6 for both detection steps. We used these results to create the set of ROC curves for each step of our detection scheme shown in Fig. 2.13.

The leftmost plot in Fig. 2.13 shows ROC curves for the first step of our detection process in which unaltered images are separated from both JPEG 2000 compressed and compressively sensed images. From these results, we can see that performing detection on subbands 3, 4, or 5 yields the best performance. For each of these subbands, our detector achieves a P_d of 100% at a P_f of 4% or less. The rightmost plot in Fig. 2.13 shows ROC curves for the second step of our detector. From these curves we can see that when using subbands 2 or 3 to perform detection, our detector achieves a P_d of approximately 90% at $P_f = 10\%$. Taken together, these results show that the detection scheme proposed in Section 2.4.3 can be used to reliably discriminate between unaltered, compressively sensed, and JPEG 2000 compressed images. For both steps of the detection process, we note that the performance decreases sharply when subband 6 or higher is used to perform detection. This is because the kurtosis of the distribution of DWT coefficients typically increases as the subband increases. This, together with the fact that the effective quantization interval used in JPEG 2000 is typically larger for higher DWT subbands, will result in the DWT coefficient distributions of unaltered, compressively sensed, and JPEG 2000 compressed images appearing very similar.

Next, we repeated the previous experiment while varying the number of compressive measurements so that the compression ratio of the compressively sensed

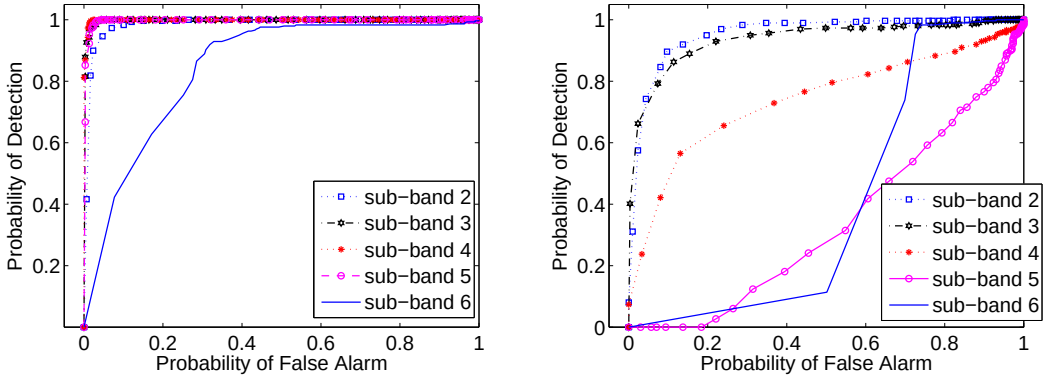


Figure 2.13: ROC curves of the first (left) and second (right) step detections on each DWT sub-band coefficients. $M/N = 0.25$ is used in compressive sensing.

images ranged between $N/M = 1.5$ and 4. In this set of experiments, we used subband 3 to perform both steps of our detection process. We used the results of this set of experiments to create the ROC curves shown in Fig. 2.14. We can see from the leftmost plot in Fig. 2.14 that the first step of our detector can achieve $P_d > 90\%$ with $P_f < 5\%$ when $N/M \geq 2$. Since in most realistic scenarios $N/M > 2$, these ROC curves show that the first step of our detector performs strongly. The rightmost plot in Fig. 2.14 shows that the second step of our detector can acheive a P_d of approximately 90% or higher at $P_f = 10\%$ for each value of N/M. These results show that our dector can be used to reliably discriminate between unaltered, compressively sensed, and JPEG 2000 compressed images in a variety of scenarios.

2.6.4 Estimator of the Number of Compressive Measurements

We performed a final set of experiments to evaluate the performance of our technique to estimate the number of compressive measurements used to capture a

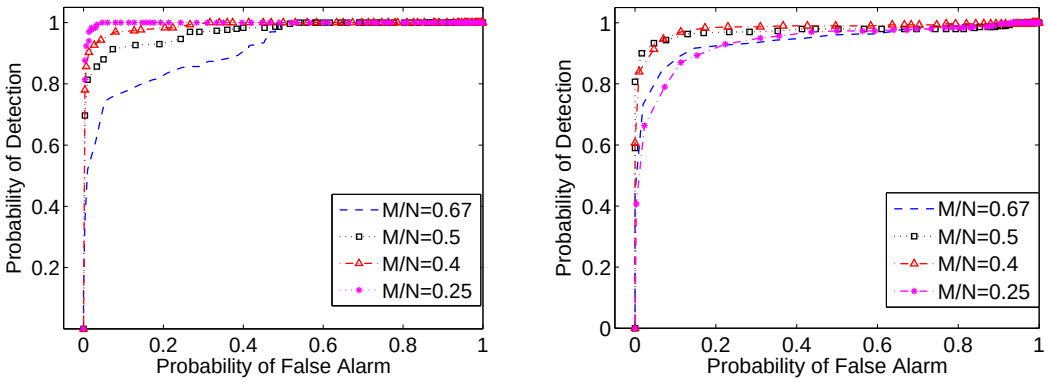
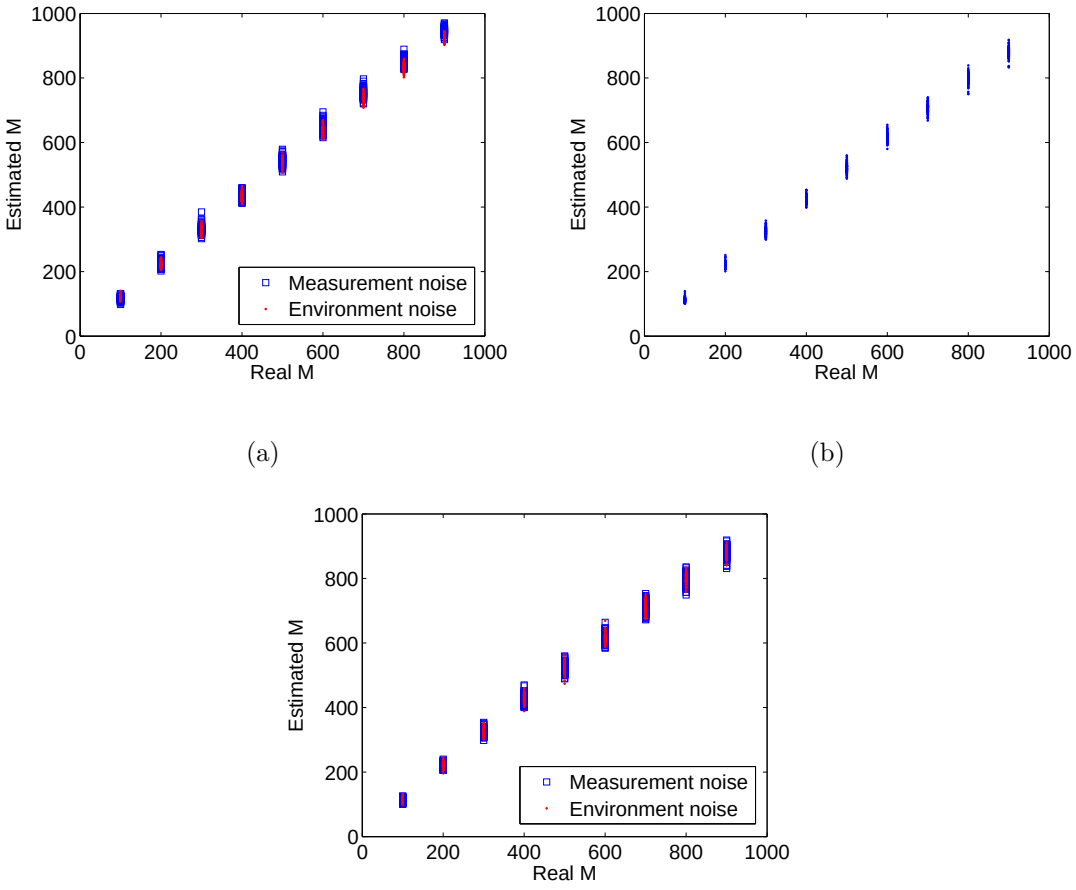


Figure 2.14: ROC curves of the first (left) and second (right) step detections on coefficients of DWT sub-band 3 under different compression ratios of compressive sensing.

signal. In these experiments, we created a set of sparse and nearly sparse signals of length $N = 1000$ as was done in Sections 2.6.2 and 2.6.1, then corrupted them using both environmental and measurement noise to create a database of 100 of each of the following signals; sparse signals in the presence of environmental noise, sparse signals in the presence of measurement noise, nearly sparse signals, nearly sparse signals in the presence of environmental noise, and nearly sparse signals in the presence of measurement noise. When creating signals corrupted by noise, we used Gaussian noise whose variance corresponded to a signal to noise ratio (SNR) of 10^3. This was done because the performance of our forensic technique decreases as the SNR decreases, thus our results can be interpreted as a conservative evaluation of our estimator's performance.

Once we created our testing database, we compressively sensed each signal while varying the number of compressive measurements from $M = 100$ to 900. We



(a)

(b)

(c)

Figure 2.15: Estimated $\hat{M}$ versus the real M for (a) sparse signals in the presence of noise, (b) nearly sparse signals, (c) nearly sparse signals in the presence of noise.

then used our forensic technique to obtain an estimate $\hat{M}$ of the number of compressive measurements used to acquire each signal. The results of this experiment are displayed in Fig. 2.15 which shows a series of plots comparing the estimated number of compressive measurements to the true number. We can see from this figure that for each signal model, our estimate closely matched the true number of measurements. Furthermore, we can see that our estimate lies within ± 25 measurements of the true number of measurements.

Additionally, we also testified the effectiveness of our proposed estimator of the number of compressive measurements on images. In these experiments, we tested our estimator on the database of compressively sensed images created in section 2.6.3. We have found that higher frequency subbands tend to have higher estimation accuracies due to their sufficient numbers of coefficients. Thus, we used subband 6 to estimate the number of compressive measurements in this subband, and then obtain the estimated ratio of M/N. The relative square error of the estimated M/N ratio was calculated as $\mathbb{E}\Big[\frac{\left(\frac{\hat{M}}{N}-\frac{M}{N}\right)^2}{(\frac{M}{N})^2}\Big]$. Table 2.1 lists these relative estimation errors for some typical choices of M/N ratios. The results show that the relative square error of our estimator on images is no greater than 5.2% for typical choices of compression ratios in compressive sensing.

Table 2.1: Relative error of estimating compressive measurements for images.

True M/N	0.25	0.4	0.5
Relative Error	3.2%	3.5%	5.2%

2.7 Summary

In this chapter, we have proposed a set of techniques to identify the use of compressive sensing in a wide variety of signals. To do this, we first identified the fingerprints left in a compressively sensed signal. We then developed two general techniques to identify compressively sensed signals; one that operates by analyzing the ratio of zero valued entries in a signal, and another that operates by identifying

changes to a signal's coefficient distribution caused by compressive sensing. Since evidence of compressive sensing in images can be confused with fingerprints left by JPEG 2000 compression, we designed a compressive sensing detection technique specifically tailored to digital images. Additionally, we proposed a technique to estimate the number of compressive measurements used to acquire a compressively sensed signal.

Our experimental results have shown that both our zero ratio and distribution-based detection schemes are able to reliably detect compressive sensing in a wide variety of realistic scenarios. Similarly, we have shown that our technique to identify compressive sensing in images can reliably distinguish compressively sensed images from both uncompressed and JPEG 2000 compressed images. Additionally, we have provided both a theoretical proof and experimental results verifying the effectiveness of our technique to estimate the number of compressive measurements used to acquire a signal.

Chapter 3

Fundamental Limits in Operation Forensics

Due to the ease of tampering a multimedia file, forensics has gained much attention in the recent decade for providing technical tools to verify the authenticity of multimedia content [88]. Enabled by techniques in existing forensic literature, forensic investigators can not only identify the acquisition environment of multimedia content [24, 38, 58, 59, 78, 94], but also detect the processing history that the content has gone through after acquisition [32, 50, 74, 77, 85, 86]. For the purpose of improving the detection performance and identifying more sophisticated manipulations, forensic researchers have always been working on discovering new fingerprints and designing new schemes [53, 66, 69, 84].

However, as the effort of developing more powerful forensic techniques goes on, evidence has shown difficulties when dealing with complicated manipulation scenarios [66]. One would then wonder if there exists a fundamental limit on forensic capability that can never be exceeded? In other words, what is the limit of investigators' capability? How many manipulation operations that investigators can detect at most? Given this information, we would be able to tell whether the existing technique has achieved the limit. If not, how far can it go? Furthermore, by quantifying the forensic capability, we may also obtain information about how to achieve the capability limit. In addition, given that forgers may manipulate multimedia content

to the extent beyond the limit of forensics, special care would be needed for such cases.

There are few works exploring the fundamental limit of forensic capabilities. To the best of our knowledge, the most related work on fundamental limit analysis of forensics was done by Swaminathan *et al.* [92, 93]. They explored the fundamental limit in component forensics by establishing two theoretical frameworks: an estimation framework and a pattern classification framework. Three types of forensic scenarios were defined in each framework regarding how much information investigators have about the components of a camera. Then, fundamental relationships of their forensic performance were derived using the above two theoretical frameworks. Moreover, in the estimation framework, Fisher information was used to obtain the optimal input for semi non-intrusive component forensics. However, these theoretical frameworks were designed for camera identification forensics, and thus they may not be suitable for answering fundamental questions in operation forensics, which focuses on detecting manipulation operations.

In this chapter, we explore the fundamental limit of operation forensics by building an information theoretical framework. We consider the forensic scenario of detecting the processing history of given multimedia content. We aim to answer the question of how many operations that investigators can detect, at most? To answer this question, we define *forensicability* as the forensic capability of detecting operations. Unlike the measure of distinguishability proposed in [25], which was based on a simple hypothesis model, our definition is applicable for more general scenarios where multiple operations may happen and many hypotheses can be

considered. Given that investigators often use features to estimate process history, in our information theoretical framework, forensicability indicates the maximum forensic information that extracted features can contain about detecting operations. Furthermore, it determines the fundamental limit of forensic detection performance of any scheme based on those features. Then, by introducing a statistical concept of *expected perfect detection*, we are able to use forensicability to determine the maximum number of operations investigators can detect. In addition, the fundamental measure of forensicability provides insights and theoretical support for predicting forgers' behavior and designing optimal forensic schemes.

The remaining of this chapter is organized as follows. Section 3.1 introduces our information theoretical framework for operation forensics, where forensicability is defined and analyzed for general scenarios. Then, to demonstrate our framework, we apply it to the forensic problem of multiple JPEG compression detection in Section 3.2. In this section, specific models for DCT coefficient histogram features are proposed to derive the expression of forensicability in this example. Then, Section 5.5 performs all experiments corresponding to the theoretical analysis in Section 3.2. Among these experimental results, we obtain the maximum number of JPEG compressions one can detect using DCT coefficient histograms. In addition, the best strategies for investigators and forgers are also analyzed in this section. Lastly, Section 4.5 concludes our work.

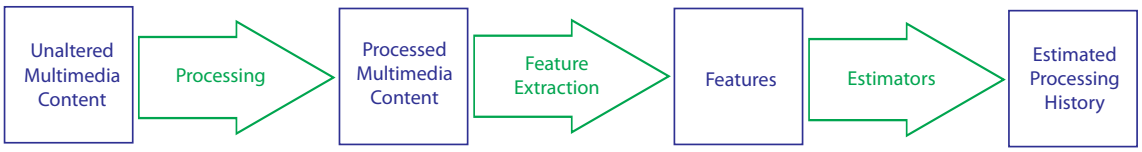


Figure 3.1: Typical process that a multimedia signal may go through when considering forensics.

3.1 Information Theoretical Framework

In this section, we introduce our information theoretical framework for general operation forensic systems. Under this model, we define the capability of investigators as forensicability, which determines the lower bound of estimation error probability and helps us answer the question of when we cannot detect any more operations.

3.1.1 Channel between Multimedia States and Features

Let us consider the process of a typical forensic analysis shown in Fig. 3.1. Unaltered multimedia content may go through some processing before investigators obtain it. In order to identify the processing history that the obtained multimedia content went through, investigators extract features from the content. Based on the extracted features, specific estimators are proposed to finally estimate the processing history.

During this process, it is often assumed that there are a finite number of hypotheses on processing histories that the multimedia content may go through. Investigators determine which hypothesis actually happened based on the analysis

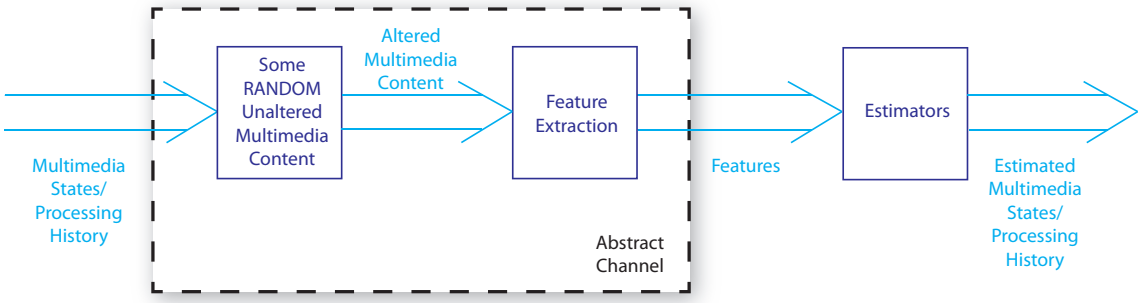


Figure 3.2: Abstract channel model in our information theoretical framework.

of extracted features. For example, to detect if the multimedia content was edited by a certain operation, like contrast enhancement [86], resizing [77] or compression [31], simple hypothesis test was used to distinguish the unaltered multimedia content and the content edited by the certain operation. In another example of detecting the number of compressions, the hypotheses would include single compression, double compression, triple compression and so on. In this work, processing history hypotheses considered in a certain forensic analysis are denoted as *multimedia states*. Then, investigators' goal is to distinguish multimedia states based on extracted features.

Given the discussion above, we reformulate the forensic system in a different way such that the relationship between multimedia states and features can be emphasized. As it is shown in Fig. 3.2, in this new formulation, the multimedia state is the input to the system. When a certain multimedia state is applied on unaltered multimedia content, features can be extracted from the processed multimedia content. Then, estimators will be applied on these features to estimate the input multimedia state.

By exploring fundamental limits in operation forensics, we want to answer

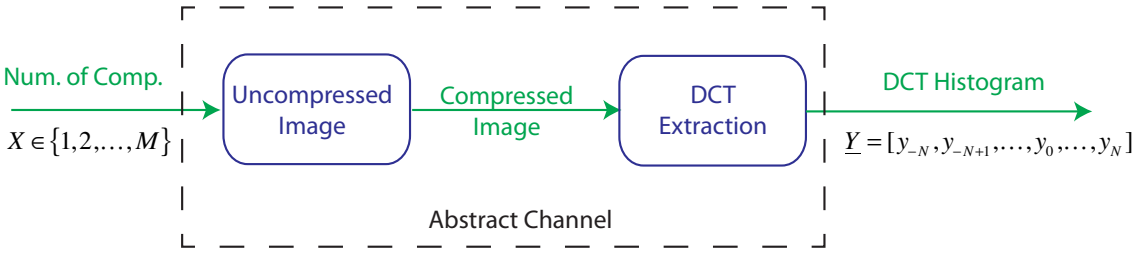


Figure 3.3: Channel model for the example of multiple compression detection forensics.

"what is the maximum information about multimedia states that investigators can obtain from the extracted features?" In other words, we are concerning the fundamental relationships between multimedia states and features, regardless of specific detectors or estimators that investigators may use to make final decisions. This motivates us to abstract all processes between multimedia states and features as a channel. Within this channel, the unaltered multimedia content can be any particular content, and it is modeled as a random variable. As a result, the relationship between multimedia states and features becomes stochastic instead of deterministic.

To demonstrate our abstract channel and further explain the relationship between multimedia states and features, let us consider an example of detecting the number of JPEG compressions using the DCT coefficients feature. As it is shown in Fig. 3.3, the multimedia state is the number of JPEG compressions from 1 to M. The feature is DCT coefficient histogram represented in a vector. Fig. 3.4 illustrates the mapping between multimedia states and features in this example. Specifically, with the same number of compressions applied, different images result in different DCT coefficient histograms, which we call them a histogram set. When

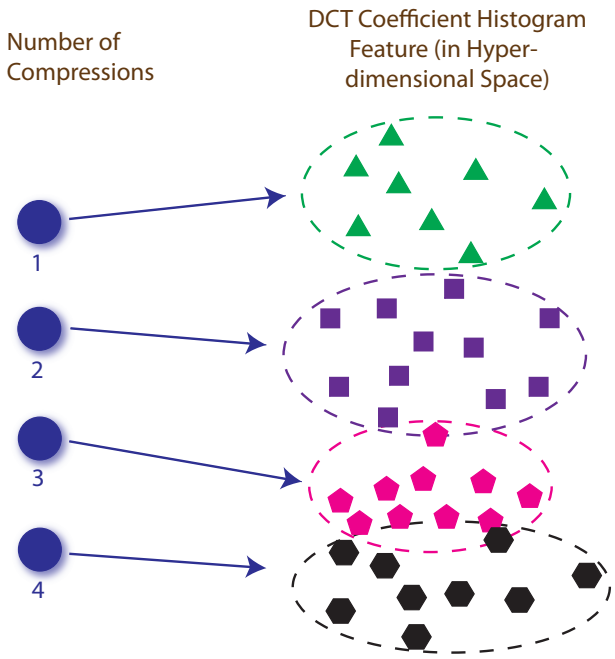


Figure 3.4: An illustration of the mapping between multimedia states and features in the example of multiple compression detection.

we detect double compressions, we are distinguishing single compression, $X = 1$, and double compression, $X = 2$. Given the distinctive fingerprints for single compression and double compression, the DCT coefficient histogram sets resulted from these two inputs can be well separated after some post-processing [74]. Thus, for $M = 2$, classification schemes can be used to distinguish the input according to the output. However, as the number of compressions considered in the system increases, more overlapping between different histogram sets may occur, which will affect the accuracy of the detection. Finally, at a certain point, we cannot distinguish all inputs and we say that we have reached our limit of detecting multiple compressions. Detailed modeling and analysis will be discussed in Section 3.2.

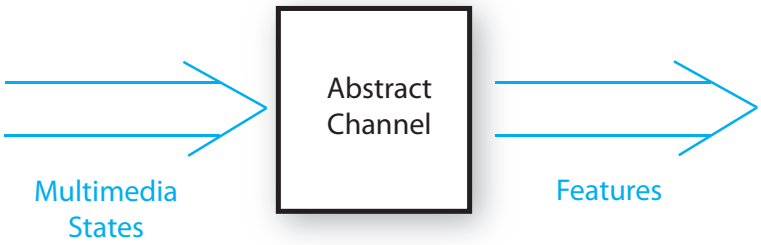


Figure 3.5: Abstract channel between multimedia states and features in the information theoretical framework for operation forensics.

3.1.2 Forensicability

Given the channel model built up between multimedia states and features, we are ready to define forensicability for operation forensics. Let us consider the general abstract channel proposed in our information theoretical framework, as it is simplified in Fig. 3.5. Let $X \in \{1, 2, \ldots, M\}$ denote the input of the channel, i.e., the multimedia state considered in a forensic analysis. Let $\underline{Y}$ denote the output of the channel, which is a vector containing features that examined by investigators. After obtaining feature $\underline{Y}$, investigators design estimators based on their statistics to estimate X. We define *forensicability* in this forensic system as the maximum information that features contain about multimedia states, regardless of any specific estimators used afterward. It is well known that, in a channel model, mutual information implies the reduction in uncertainty of input due to the knowledge of output. Thus, given the similarity between these two concepts, we define forensicability as follows.

Definition 1. *In operation forensics, where features are used to identify multimedia*

states, forensicability *of using feature* $\underline{Y}$ *towards identifying multimedia state* X *is defined as the mutual information between* X *and* $\underline{Y}$*, i.e.,* $I(X;\underline{Y})$.

Forensicability of an operation forensic system implies the maximum forensic information that features contain about multimedia states. More importantly, it determines the best performance investigators can obtain by examining these features through all possible estimators. We demonstrate this significance in the following theorem.

Theorem 2. *Consider any estimator of the multimedia state* $\hat{X}$ *such that* $X \to \underline{Y} \to \hat{X}$ *is a Markov Chain, i.e., the value of* $\hat{X}$ *depends only on* $\underline{Y}$ *and not on* X*. Let* $P_e = \mathbb{P}(X \neq \hat{X})$ *denote the error probability. If the estimator is better than a random decision where* $\hat{X}$ *is uniformly and randomly drawn from the set of* X*, i.e.,* $P_e \leq \frac{M-1}{M}$*, then we have*

$$P_e \geq P_e^0, \tag{3.1}$$

where P_e^0 *is the lower bound of error probability. It is unique and satisfies the following equation*

$$H(P_e^0) + P_e^0 \log_2(M-1) = H(X) - F(X;\underline{Y}). \tag{3.2}$$

Proof. From the corollary of Fano's inequality in [28], we have

$$H(P_e) + P_e \log_2\big(|\mathcal{X}| - 1\big) \geq H(X|\underline{Y}), \tag{3.3}$$

where $|\mathcal{X}|$ is the cardinality of the input X and thus $|\mathcal{X}| = M$. In order to later examine the equality conditions, we briefly review the derivation of (3.3) in [28] as

follows. First, let $E = \mathbf{1}(\hat{X} \neq X)$ denote an error random variable, then $H(E, X|\hat{X})$ can be expanded in two ways,

$$\begin{aligned} H(E, X|\hat{X}) &= H(X|\hat{X}) + H(E|X, \hat{X}) \\ &= H(E|\hat{X}) + H(X|E, \hat{X}). \end{aligned} \tag{3.4}$$

While $H(E|X, \hat{X}) = 0$ and $H(E|\hat{X}) \leq H(E) = H(P_e)$, the upper bound of $H(X|E, \hat{X})$ is obtained as

$$\begin{aligned} H(X|E, \hat{X}) &= \mathbb{P}(E = 0)H(X|\hat{X}, E = 0) \\ &\quad + \mathbb{P}(E = 1)H(X|\hat{X}, E = 1) \\ &\leq (1 - P_e)0 + P_e \log_2 \big(|\mathcal{X}| - 1\big). \end{aligned} \tag{3.5}$$

Thus, combining the above results, we have

$$H(P_e) + P_e \log_2 \big(|\mathcal{X}| - 1\big) \geq H(X|\hat{X}) \geq H(X|\underline{Y}). \tag{3.6}$$

Given (3.3), we examine the derivative of the left hand side of this inequality with respect to P_e,

$$\frac{\partial\big(H(P_e) + P_e \log_2(M-1)\big)}{\partial P_e} = \log_2 \left(\frac{1 - P_e}{P_e}(M-1) \right) \geq 0. \tag{3.7}$$

The last step holds because $P_e \leq \frac{M-1}{M}$. Therefore, the left hand side of (3.3) is an increasing function of P_e for $P_e \leq \frac{M-1}{M}$. Then, the minimum of P_e, which is denoted by P_e^0, can be obtained by solving the equality of (3.3). Hence, we have, $P_e \geq P_e^0$, where P_e^0 is the unique solution of the following equation,

$$H(P_e^0) + P_e^0 \log_2(M-1) = H(X|\underline{Y}) = H(X) - I(X; \underline{Y}). \tag{3.8}$$

□

The lower bound P_e^0 can be achieved if and only if all of the following conditions are satisfied.

1. $H(E|\hat{X}) = H(E)$, i.e., E and $\hat{X}$ are independent. Furthermore, it can be easily proved that the independence between E and $\hat{X}$ implies that the error probability for each given estimated result is the same, i.e., $\mathbb{P}(X \neq i|\hat{X} = i) = \mathbb{P}(X \neq j|\hat{X} = j), \forall 1 \leq i, j \leq M$. For the specific setting of this work, it indicates that multimedia states are equally hard to be correctly identified.

2. $H(X|\hat{X}, E = 1) = \log_2(M - 1)$, which implies that no information can be inferred from a known missed detection towards finding the correct one. For the specific setting of this work, this condition means that, given a wrong estimated multimedia state, probabilities of the true multimedia state being any other multimedia states are the same.

3. $I(X; \hat{X}) = I(X; \underline{Y})$, i.e., $X \to \hat{X} \to \underline{Y}$ is also a Markov chain. This implies that, the estimated input contains all information that the real input has about the channel output. For the specific setting of this work, it means that the distribution of features given an estimated multimedia state will not change if the real multimedia state is also known.

In addition, with the assumption of uniform prior for X, which is commonly used in forensic analysis, the error probability lower bound will be only dependent on forensicability:

$$H(P_e^0) + P_e^0 \log_2(M - 1) = \log_2 M - I(X; \underline{Y}). \tag{3.9}$$

Note that, while uniform priors are adopted in this work, cases with non-uniform priors can be similarly handled by using the initial equation (3.2) instead of (3.9).

3.1.3 Expected Perfect Detection

While the lower bound of error probability gives fundamental limit on estimators' performance, we also want to answer the question of "when cannot we detect any more operations?" For example, in the multiple compression detection problem discussed earlier, we may want to know how many compressions we can detect at most. To answer these questions, we need a criterion to make decisions on whether we can or cannot detect more. One possible way is to check the equality of $I(X;\underline{Y}) \leq H(X)$. If equality holds, then there exists some estimator which can distinguish all considered multimedia states with zero error probability. Otherwise, it implies that not all multimedia states can be distinguished with zero error probability by any estimator.

However, for most cases, the equality of $I(X;\underline{Y}) = H(X)$ may never hold. This is because that multimedia state X is always a discrete variable, while the feature vector $\underline{Y}$ given a certain multimedia state X is usually modeled as continuous variables, such as multivariate Gaussian random variables. Thus, as long as the supports of conditional distributions of $\underline{Y}$ given different X have any overlap, it is impossible to perfectly estimate X from $\underline{Y}$, i.e., $I(X;\underline{Y}) < H(X)$. For these cases, the question becomes "how small should the error probability be so that we can still consider it as a perfect detection?"

Such a question leads us to examine the relationship between theoretical and experimental results. Given a rare incident, i.e., the probability that this incident happens tends to zero, it is very likely that we will not observe it in real experiments. Therefore, if the theoretical error probability is small enough, then we may not see the occurrence of error within a limited number of observations. Inspired by this idea, we reformulate the process of experimental testing as follows.

Given an image that may belong to any multimedia state considered in the analysis, there is probability P_e that the image will be misidentified. When we experimentally evaluate the performance of a detector on a database, we go through the following steps. First, an image is picked from a database containing images of all possible multimedia states. Then the detection scheme is applied on this image to obtain an estimated multimedia state. Lastly, by comparing the estimated multimedia state with the ground truth, we know whether the detection was correct or not. Given that nothing is known until the last step, each image is treated equally during estimation. By iterating these steps for every image in the database, the experimental error probability can be calculated as the total number of misclassifications divided by the size of the database. This process can be considered as a sequential process, where each time an image is randomly picked and its multimedia state is estimated by a detector, whose theoretical detection error probability is P_e. Then, by definition of P_e, for each individual detection, the tested image has probability P_e of being misidentified and probability $1 - P_e$ of being correctly detected. From this formulation, we can see an analogy between the process of experimental testing and a Bernoulli process.

Motivated by the discussion above, we model each sample in the testing database as an independent and identical Bernoulli random variable with probability P_e of missed detection. It is well known in probability theory that, the expected time of the first occurrence of missed detection happens at $1/P_e$. In other words, if the experimental database only has $S < 1/P_e$ samples, then the missed detection may not occur in expected sense, where the expectation is taken among all databases with the same size S. Thus, we propose the definition of *expected perfect detection* as follows.

Definition 2. *Given an experimental database of size S, the* expected perfect detection *happens if and only if the theoretical error probability satisfies $P_e < 1/S$.*

Based on this definition, a simple corollary below can give us the criterion to determine when we cannot detect any longer.

Corollary 2. *For an experimental database of size S, if the lower bound of error probability obtained from* (3.2) *satisfies $P_e^0 > 1/S$, then no expected perfect detection can be obtained for any estimators.*

We note that all above analysis is based on the law of large number. Experimentally, we find that the size of the database needs to be at the order of thousands for the expected perfect detection argument being hold. Fortunately, most experimental databases used in forensic analysis satisfy this condition.

3.2 Information Theoretical Modeling for JPEG Compression Forensics

To demonstrate the effectiveness of our proposed framework for operation forensics, we use the multiple JPEG compressions detection forensics as an example [20].

3.2.1 Background on JPEG Compression Forensics

An image's JPEG compression history is forensically important because it helps investigators to identify the image's acquisition process and detect possible manipulations [73, 76]. Specifically, by estimating the quantization table of a singly compressed image, one can identify the model of the camera that captured the image [73]. Furthermore, when a forger manipulates a JPEG image and re-saves it in the same format, double JPEG compression fingerprints may left in the image [35,44, 64,74,76]. The more times the JPEG image is manipulated, the more times of JPEG compressions it may go through. Thus, detecting the number of JPEG compressions that an image has gone through can help investigators to understand how much the image has been tampered. However, as the number of JPEG compressions increases, the multiple compression fingerprints become less distinguishable [53, 66]. So a natural question would be "how many JPEG compressions can we detect, at most?"

Before applying our information theoretical model to answer this question, let us first review the typical process of a JPEG compression. When JPEG compressing

an image, block-wise DCT transform is first applied on the pixel domain to obtain coefficients in DCT domain. Then, these coefficients are quantized and encoded by an entropy coder to get the JPEG data file. Whenever the image is edited or processed, decompression is needed, which follows the reverse procedure of compression. During decompression, the quantized DCT coefficients cannot be recovered. Thus, by examining the difference of DCT coefficients between uncompressed and compressed images, one can observe important fingerprints of JPEG compression. Furthermore, multiple JPEG compressions can also be detected by examining these coefficients.

Let D_0 denote a coefficient of a certain DCT subband of an uncompressed image. We use the Laplacian model to characterize the distribution of D_0 [54], where

$$f_{D_0}(\rho) = \frac{\lambda}{2} e^{-\lambda|\rho|}, \ \ \rho \in \mathbb{R}. \tag{3.10}$$

During JPEG compression, let a_1 be the quantization step used in this subband, and D_1 denote the DCT coefficient after compression, then

$$D_1 = \text{round}\left(\frac{D_0}{a_1}\right) \cdot a_1. \tag{3.11}$$

Thus, D_1 has a discrete distribution of

$$\begin{aligned} \mathbb{P}(D_1 = l_1 a_1) &= \int_{(l_1-1/2)a_1}^{(l_1+1/2)a_1} f_{D_0}(\rho) d\rho, \ \ l_1 \in \mathbb{Z}, \\ &= \begin{cases} 1 - e^{-\lambda a_1/2}, & \text{if } l_1 = 0, \\ e^{-\lambda|l_1 a_1|} \sinh\left(\frac{\lambda a_1}{2}\right), & \text{if } l_1 \neq 0. \end{cases} \end{aligned} \tag{3.12}$$

By examining the DCT coefficient histogram, investigators can detect whether the image is singly compressed or not. Furthermore, quantization step sizes can also be estimated if the image is detected as a singly compressed one [73].

When recompressing this singly compressed image using quantization step of $a_2, a_2 \neq a_1$, in the examined subband, let D_2 denote the DCT coefficient after two compressions, then we have

$$D_2 = \text{round}\left(\frac{D_1}{a_2}\right) \cdot a_2 = \text{round}\left(\text{round}\left(\frac{D_0}{a_1}\right) \cdot \frac{a_1}{a_2}\right) \cdot a_2, \tag{3.13}$$

and

$$\mathbb{P}(D_2 = l_2 a_2) = \sum_{(l_2-\frac{1}{2})a_2 \leq l_1 a_1 < (l_2+\frac{1}{2})a_2} \mathbb{P}(D_1 = l_1 a_1), \;\; l_2 \in \mathbb{Z}. \tag{3.14}$$

Due to the effect of double quantization, the histogram of D_2 will present periodic characteristics, either periodic peaks or periodic zeros. Then, by examining the Fourier transform of the histogram, investigators can distinguish between singly compressed images and doubly compressed images [64, 74, 76].

3.2.2 DCT Coefficients Feature Model

Given that the histogram of DCT coefficients is a commonly used feature to detect JPEG compressions, in this example, we examine the fundamental limit of using DCT coefficient histograms to detect multiple JPEG compressions. We note that, other features used to detect JPEG compressions can be analyzed by similar approaches. As it is shown in Fig. 3.3, we consider an abstract channel where the input $X \in \{1, 2, ..., M\}$ is the number of JPEG compressions and the output $\underline{Y}$ is the DCT coefficient histogram written in a vector form.

To demonstrate the relationship between X and $\underline{Y}$, we take one subband as an illustration. We use λ to denote the parameter of the Laplace distribution of the coefficient D_0 in this subband when it is not compressed (3.10). Let $\mathcal{Q}_M = (q_1, q_2, ..., q_M)$ denote the set of quantization step sizes that may be used for this subband during compressions. Since in multiple compression detection forensics, the given image is a JPEG image and investigators try to detect how many compressions have been done before this last one, we keep the last compressions the same for all hypotheses. Without loss of generality, we take q_M as the quantization step size used in the last compression for all hypotheses. Then, if there are actually m applications of JPEG compressions, the DCT coefficient should have been quantized by step sizes $\{q_{M-m+1}, q_{M-m+2}..., q_M\}$ in order. Let D_m denote the DCT coefficients if m times of JPEG compressions are applied. By following the analysis in (3.13) and substituting $\{a_1, a_2..., a_m\}$ with $\{q_{M-m+1}, q_{M-m+2}..., q_M\}$, we have,

$$D_m = \text{round}\left(...\,\text{round}\left(\,\text{round}\left(\tfrac{D_0}{q_{M-m+1}}\right) \times \tfrac{q_{M-m+1}}{q_{M-m+2}}\right)\right) \times q_M. \tag{3.15}$$

Given this equation and (3.10), we can derive the distribution of D_m, which only has nonzero values at integer multiples of q_M. Let vector $\underline{v}_m(\lambda, \mathcal{Q}_M)$ denote this theoretical distribution, with each element $v_{n,m}(\lambda, \mathcal{Q}_M)$ representing the nonzero probability mass function $\mathbb{P}(D_m = nq_M)$, then

$$\underline{v}_m(\lambda, \mathcal{Q}_M) = \big[\mathbb{P}(D_m = -Nq_M), ..., \mathbb{P}(D_m = Nq_M)\big]. \tag{3.16}$$

In reality, however, we may not observe the theoretical distribution from the DCT histogram due to the model mismatch and/or the rounding and truncation in the compression and decompression. Instead, the normalized DCT coefficient

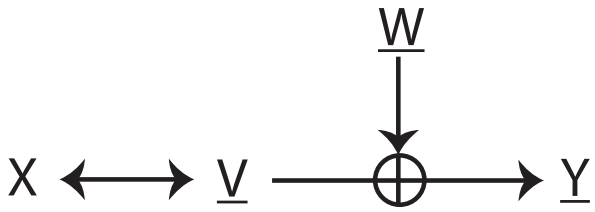


Figure 3.6: Abstract channel inner structure for the model in Fig. 3.3.

histogram that we observe may be a noisy version of the theoretical distribution. Let random variable $\underline{Y}_m(\lambda, \mathcal{Q}_M)$ denote the observed normalized histogram if m applications of JPEG compressions were applied, i.e.,

$$\underline{Y}_m(\lambda, \mathcal{Q}_M) = [B_m(-Nq_M), ...B_m(Nq_M)], \tag{3.17}$$

where $B_m(nq_M), -N \leq n \leq N$, denotes the normalized histogram bin at location nq_M when m times of compressions happened. Then, by assuming that the observation noise, denoted by $\underline{W}$, is an additive noise, we have

$$\underline{Y}_m(\lambda, \mathcal{Q}_M) = \underline{v}_m(\lambda, \mathcal{Q}_M) + \underline{W}. \tag{3.18}$$

Let random variable $\underline{V}(\lambda, \mathcal{Q}_M) \in \{\underline{v}_1(\lambda, \mathcal{Q}_M), \underline{v}_2(\lambda, \mathcal{Q}_M), \ldots, \underline{v}_M(\lambda, \mathcal{Q}_M)\}$ denote the theoretical distribution of DCT coefficients. Then, for a certain subband, given a fixed λ and $\mathcal{Q}_M$, the abstract channel in Fig. 3.3 can be depicted as the diagram in Fig. 3.6. Specifically, for each hypothesis on the number of JPEG compression X, it dictates a theoretical distribution on DCT coefficients $\underline{V}$, which can be calculated by (3.10) and (3.15). But due to the observation noise $\underline{W}$, the obtained normalized DCT coefficient histogram is $\underline{Y}$ in (3.18).

3.2.3 Forensicability for JPEG Compression Forensics

Based on our information theoretical framework, forensicability of using DCT histogram to detect multiple JPEG compressions is $I\big(X;\underline{Y}(\lambda,\mathcal{Q}_M)\big)$.

To calculate forensicability, we first assume that the observation noise on different histogram bins are independent with each other, then the covariance of $\underline{W}$ is a diagonal matrix. Furthermore, based on experimental results, which will be shown in Section 5.5, we use the multivariate Gaussian distribution to model the observation noise as follows

$$\underline{W}(\lambda,\mathcal{Q}_M)\sim\mathcal{N}\Big(\underline{d},diag\big(\beta\underline{V}^{2\alpha}(\lambda,\mathcal{Q}_M)\big)\Big), \tag{3.19}$$

where $\underline{d}$, $\beta>0$ and $\alpha>0$ are constant parameters, which will be estimated later. We note that, in our model, the variance of observation noise, $\mathrm{Var}(\underline{W})$, is proportional to the signal $\underline{V}$ that the noise is added on. This is because that the model mismatch and the rounding and truncation effect in the compression and decompression are more obvious on significant histogram bins.

In this example, we consider the case where we have no biased information on how many compressions that the image might have gone through, i.e., X has equal probability of being any value in $\{1,2,\ldots,M\}$. Then, given (3.18), (**??**) and (3.19), we can derive the forensicability of using DCT histogram to detect multiple JPEG compressions as the following expression

$$F_{\lambda,\mathcal{Q}_M}(X;\underline{Y})=\log_2 M-\frac{1}{M}\sum_{m=1}^{M}\mathbb{E}\bigg[\log_2\sum_{j=1}^{M}exp\Big(\Phi_j^m(\underline{V})\Big)\bigg], \tag{3.20}$$

where

$$\Phi_j^m(\underline{V}) = \sum_{n=-N}^{N} \left[\alpha \ln \frac{v_{n,m}}{v_{n,j}} - \frac{(Y_n - v_{n,j})^2}{2\beta v_{n,j}^{2\alpha}} + \frac{(Y_n - v_{n,m})^2}{2\beta v_{n,m}^{2\alpha}} \right]. \tag{3.21}$$

Proof. Since $\underline{d}$ is a constant, we have $I(X;\underline{Y}) = I(X;\underline{Y}-\underline{d})$. Thus, in the following derivation, we take $\underline{d} = \underline{0}$ for simplicity. Then, the conditional probability of $\underline{Y}$ given $X = m$ is

$$\begin{aligned} f_{\underline{Y}}(\underline{y}|X=m) &= \prod_{n=-N}^{N} \frac{1}{\sqrt{2\pi\beta v_{n,m}^{2\alpha}}} exp\left[- \frac{(y_n - v_{n,m})^2}{2\beta v_{n,m}^{2\alpha}} \right] \\ &= \frac{1}{\prod_{n=-N}^{N} \sqrt{2\pi\beta v_{n,m}^{2\alpha}}} exp\left[\sum_{n=-N}^{N} - \frac{(y_n - v_{n,m})^2}{2\beta v_{n,m}^{2\alpha}} \right]. \end{aligned} \tag{3.22}$$

Based on Bayes' theorem and $\mathbb{P}(X = m) = \frac{1}{M}, \forall m \leq M$, we calculate the conditional entropy of X given $\underline{Y}$ as follows,

$$\begin{aligned} & H(X|\underline{Y}) \\ &= \int_{\mathbb{R}^{2N+1}} \sum_{m=1}^{M} p_X(m|\underline{Y}=\underline{y}) \log_2 \frac{1}{p_X(m|\underline{Y}=\underline{y})} f_{\underline{Y}}(\underline{y}) d\underline{y} \\ &= \int_{\mathbb{R}^{2N+1}} \sum_{m=1}^{M} \frac{f_{\underline{Y}}(\underline{y}|X=m)}{M f_{\underline{Y}}(\underline{y})} \log_2 \frac{\sum_{j=1}^{M} f_{\underline{Y}}(\underline{y}|X=j)}{f_{\underline{Y}}(\underline{y}|X=m)} f_{\underline{Y}}(\underline{y}) d\underline{y} \\ &= \frac{1}{M} \sum_{m=1}^{M} \int_{\mathbb{R}^{2N+1}} f_{\underline{Y}}(\underline{y}|X=m) \log_2 \left[\sum_{j=1}^{M} \frac{f_{\underline{Y}}(\underline{y}|X=j)}{f_{\underline{Y}}(\underline{y}|X=m)} \right] d\underline{y}. \end{aligned} \tag{3.23}$$

By (3.22), the ratio between $f_{\underline{Y}}(\underline{y}|X=j)$ and $f_{\underline{Y}}(\underline{y}|X=m)$ can be calculated

as

$$\begin{aligned}
&\frac{f_{\underline{Y}}(\underline{y}|X=j)}{f_{\underline{Y}}(\underline{y}|X=m)} \\
&= \prod_{n=-N}^{N} \frac{v_{n,m}^{\alpha}}{v_{n,j}^{\alpha}} exp\left[\sum_{n=-N}^{N} -\frac{(y_n - v_{n,j})^2}{2\beta v_{n,j}^{2\alpha}} + \frac{(y_n - v_{n,m})^2}{2\beta v_{n,m}^{2\alpha}} \right] \\
&\triangleq \quad exp\Big[\Phi_j^m(\underline{v})\Big], \qquad (3.24)
\end{aligned}$$

where

$$\Phi_j^m(\underline{v}) = \sum_{n=-N}^{N} \left[\alpha \ln \frac{v_{n,m}}{v_{n,j}} - \frac{(y_n - v_{n,j})^2}{2\beta v_{n,j}^{2\alpha}} + \frac{(y_n - v_{n,m})^2}{2\beta v_{n,m}^{2\alpha}} \right]. \qquad (3.25)$$

Take the notation of (3.24) into (3.23), we have

$$\begin{aligned}
&H(X|\underline{Y}) \\
&= \frac{1}{M} \sum_{m=1}^{M} \int_{\mathbb{R}^{2N+1}} f_{\underline{Y}}(\underline{y}|X=m) \log_2 \left\{ \sum_{j=1}^{M} exp\Big[\Phi_j^m(\underline{v})\Big] \right\} d\underline{y} \\
&= \frac{1}{M} \sum_{m=1}^{M} \mathbb{E}\left[\log_2 \sum_{j=1}^{M} exp\Big(\Phi_j^m(\underline{V})\Big) \right]. \qquad (3.26)
\end{aligned}$$

Given that $I(X;\underline{Y}) = \log_2 M - H(X|\underline{Y})$, we have completed the derivation of (3.20) and (3.21). □

Note that the right hand side expression in (3.20) and (3.21) still depend on λ and $\mathcal{Q}_M$. We remove these dependencies from variables in the sequel to simplify the expression. It is also noticed from (3.20) and (3.21) that forensicability does not depend on the constant mean $\underline{d}$ of the observation noise. This is because that any constant deviation of the output can be directly subtracted from input without any effect on the channel performance.

Before calculating forensicability, we need to estimate parameters β and α in the variance of observation noise (3.19). Based on (3.18) and (3.19), we apply

maximum likelihood estimator to obtain the optimal β and α. Given that $\underline{d}$ has no effect on forensicability, we first derive the estimator for $\underline{d} = \underline{0}$. Let $Y_{\lambda_i,n,m}$ denote the n^{th} histogram bin of the i^{th} image (whose Laplace parameter is λ_i) after m times of compressions. Then, the optimal β and α are

$$(\hat{\beta}, \hat{\alpha}) = \arg\max_{\beta>0,\alpha>0} \log \sum_{i=1}^{K} \sum_{n=-N}^{N} \sum_{m=1}^{M} \mathbb{P}(Y_{\lambda_i,n,m} = y_{\lambda_i,n,m}). \tag{3.27}$$

According to Karush-Kuhn-Tucker conditions, we have

$$\begin{cases} \sum_{i=1}^{K} \sum_{n=-N}^{N} \sum_{m=1}^{M} (y_{\lambda_i,n,m} - v_{\lambda_i,n,m})^2 \ln v_{\lambda_i,n,m} (\frac{1}{v_{\lambda_i,n,m}})^{2\hat{\alpha}} = \hat{\beta} \sum_{i=1}^{K} \sum_{n=-N}^{N} \sum_{m=1}^{M} v_{\lambda_i,n,m}, \\ \sum_{i=1}^{K} \sum_{n=-N}^{N} \sum_{m=1}^{M} \frac{(y_{\lambda_i,n,m} - v_{\lambda_i,n,m})^2}{v_{\lambda_i,n,m}^{2\hat{\alpha}}} = \hat{\beta} K (2N+1) M. \end{cases} \tag{3.28}$$

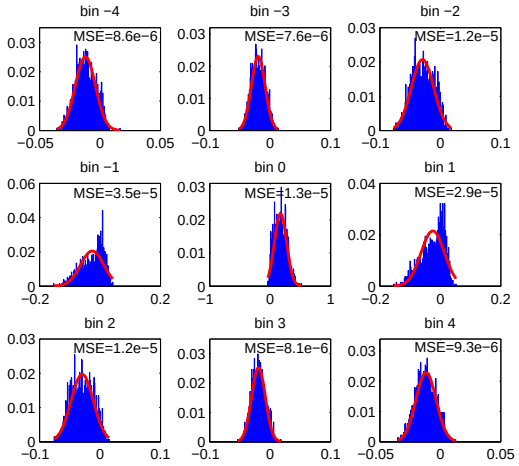
Given that the theoretical distribution $v_{\lambda_i,n,m} \in [0, 1]$, the left hand side of (3.28) is monotonically increase with $\hat{\alpha}$. Then $\hat{\alpha}$ can be approximated for any given $\hat{\beta}$. In addition, from (3.28), $\hat{\beta}$ can be derived for any fixed $\hat{\alpha}$. Thus, an iterative algorithm can be used to obtain the optimal $\hat{\beta}$ and $\hat{\alpha}$ from (3.28) and (3.28). For $\underline{d} \neq \underline{0}$ cases, similar estimators can be derived with $y_{\lambda_i,n,m}$ substituted by $y_{\lambda_i,n,m} - d_n$, where $d_n, n \in [-N, N]$, is the n^{th} element in $\underline{d}$.

Lastly, we note that, as the first work proposing and calculating forensicability in operation forensics, JPEG compression forensics has been chosen as it is a well studied problem in literature. Furthermore, the existing model of DCT coefficient histograms has helped us simplify the analysis of channel characteristics. Nevertheless, similar approaches can be applied to other forensic problems to find their fundamental limit of forensicability. For example, in contrast enhancement detection [86], the input of the channel is either unaltered, i.e., $X = 0$, or contrast

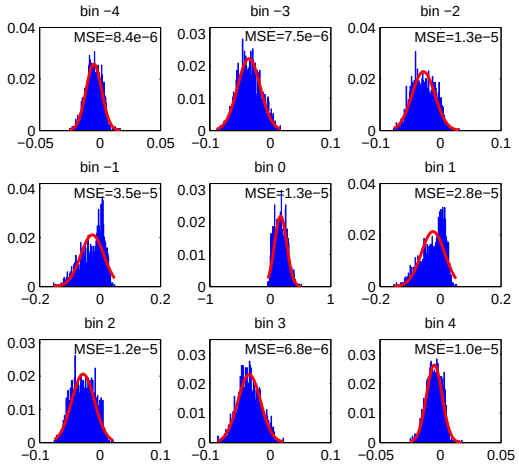
enhanced, i.e., $X = 1$. The extracted feature can be taken as the high frequency component of the image pixel histogram. Then, similar approaches can be applied to model the relationship between features and multimedia states. Forensicability can also be calculated to imply the best performance one can possibly obtain. Furthermore, by comparing the forensicability of contrast enhancement detection and those of other detections, such as resizing detection [77], one can find which manipulation is fundamentally easier to be detected. In addition, our framework may also be used to explore the fundamental limit of detecting the order of manipulation operations [84]. In this case, multimedia states would be any combinations of considered operations, and features can be built by concatenating all useful features for distinguishing the order of these operations.

3.3 Data-Driven Results and Analysis

In this section we provide experimental support for our proposed framework and calculate the forensicability for JPEG compression forensics. From analyzing forensicability, we are able to answer how many JPEG compressions, at most, that investigators can detect. Furthermore, we also examine the effect of compression quality factors and different DCT subbands on forensicability in order to provide guidance of strategies for both investigators and forgers.



(a)



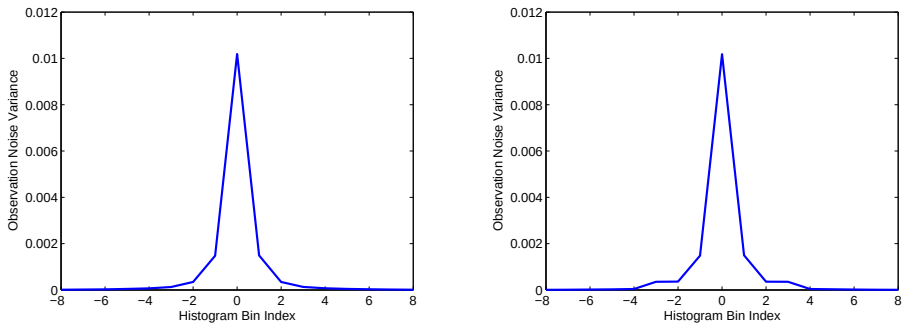
(b)

Figure 3.7: Normalized histograms of observation noise and their estimated Gaussian distributions (plotted in red lines) on different histogram bins for (a) single compressed images with quantization step size of 6 in the examined subband and (b) doubly compressed images with quantization step size of 6 then 7 in the examined subband. Bin i means that the observation noise on normalized histogram bin $B(iq_{last})$ is examined, where q_{last} denotes the last quantization step size. The mean square error of each estimation is also shown in the subfigure.

3.3.1 Verification of Observation Noise Model

To support our proposed observation noise model in (3.19), we conduct an experiment to examine the difference between observed normalized histograms and their theoretical distributions. Our test images are generated from the 1338 uncompressed images from UCID database [82]. We first create the 1338 singly compressed images by JPEG compressing the uncompressed images using quality factor of 80. We examine the $(2, 3)$ subband, where the corresponding quantization step size is 6. Double compressed images are also examined for verification, where we obtain these test images by double JPEG compressing the uncompressed 1338 images using quality factors 80 and then 75. The corresponding quantization step sizes for the examined subband are 6 and 7 respectively. The observed normalized histograms are obtained directly from these two sets of compressed images. We calculate the theoretical distributions for singly compressed images and doubly compressed images based on their uncompressed versions. Specifically, for each of the 1338 images, we first estimate the Laplace parameter λ based on the DCT coefficients of the uncompressed image. Then the theoretical distribution is calculated according to (3.15) and (3.16) for given λ and quantization step sizes. Observation noise is calculated by subtracting the theoretical distributions from the observed normalized histograms.

Fig. 3.7 plots the histograms of observation noise and their estimated Gaussian distributions for different histogram bin locations for both singly compressed images and doubly compressed images. From these results, we can see that Gaussian distributions can well approximate the distributions of the observation noise for



(a) (b)

Figure 3.8: Variance of observation noise versus histogram bin index for (a) single compressed images with quantization step size of 6 in the examined subband; and (b) doubly compressed images with quantization step size of 6 then 7 in the examined subband.

most of cases. Furthermore, the mean of the histograms does not change much between singly compressed images and doubly compressed images. This gives support on our constant mean model of the observation noise.

Fig. 3.8 plots the variance of observation noise for different histogram bin locations for both singly compressed images and doubly compressed images. Given the discussion in Section 3.2.1, the DCT coefficient distribution of singly compressed image is quantized Laplace distribution. Although different images have different Laplace parameters and their DCT coefficient distributions may be different, these distributions share a common shape of having a central peak at zero and decreasing fast as the absolute value of the variable increases. The observation noise variance of singly compressed images exhibits similar characteristics as it is plotted in Fig. 3.8(a). Furthermore, for double compressed images where the second quality factor

is lower than the first one, double compression fingerprints of periodic peaks will be presented in DCT coefficient histograms. Similar fluctuation of the observation noise variance is shown in Fig. 3.8(b). Therefore, both figures in Fig. 3.8 show that the variance of observation noise changes in the similar way as the value of theoretical distribution changes. In other words, these experimental results show that the variance of observation noise is proportional to the theoretical distribution. This validates the proposed variance model of the observation noise in (3.19). Furthermore, instead of using a linear model, an exponential proportionality principle is adopted in the variance model to make it more general.

We note that there may be more accurate but complicated models for the observation noise. We use the model in (3.19) as a tradeoff between the accuracy of modeling and the complexity of analysis.

3.3.2 Forensicability Calculation

In order to calculate forensicability, we first estimate parameters β and α from (3.28) and (3.28). We use the normalized DCT coefficient histograms of singly compressed images and their corresponding theoretical distributions obtained from last subsection to estimate. Due to the nonzero mean of observation noise, we subtract this mean from the observed normalized histograms before using them in (3.28) and (3.28). Then, we exclude insignificant histogram bins due to the severe noise effect on those small histogram bins. Specifically, we use those normalized histogram bins whose theoretical probabilities are equal or greater than 5×10^{-4}. This results in

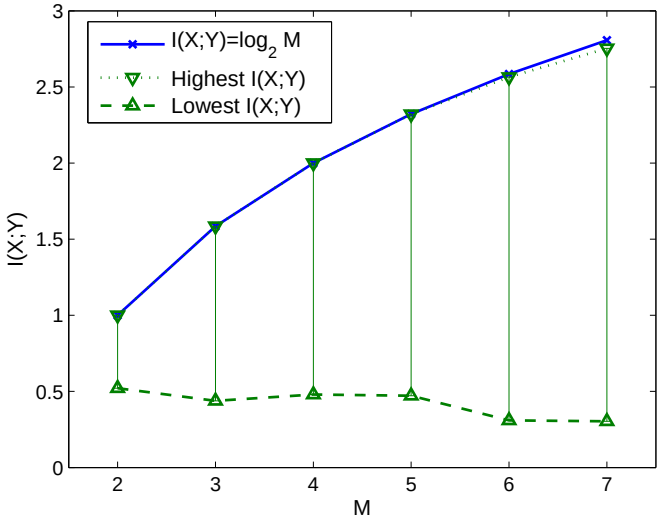


Figure 3.9: The reachable forensicabilities of different compression quality factors $\mathcal{Q}_M$ and the upper bound of forensicability for different M's.

total 36298 histogram bins used for estimation. The estimated parameters' values are

$$\hat{\beta} = 0.0494, \quad \hat{\alpha} = 0.744. \tag{3.29}$$

Given β and α, forensicability of multiple JPEG compression forensics can be obtained from (3.20) and (3.21). Since (3.20) is not a closed form and we cannot calculate the precise value, we use Monte Carlo simulation to approximate the result. This is a commonly used method in information theoretic analysis [68]. We demonstrate the results for subband $(2, 3)$, where we take a typical value of $\lambda = 0.1$. We find that the quantization step size in this subband changes from 1 to 14 when varying the JPEG compression quality factor from 50 to 100. By excluding the trivial cases where one quantization interval is an integer multiple of another, we

choose the candidate quantization step sizes as

$$\{5, 6, 7, 8, 9, 11, 13\}. \tag{3.30}$$

Then, for each M, we randomly select values from this candidate set to construct $\mathcal{Q}_M$, under the constraint that two adjacent elements are not equal.

For each different $\mathcal{Q}_M$, $I_{\lambda,\mathcal{Q}_M}(X;\underline{Y})$ is estimated by Monte Carlo averaging and plotted in Fig. 3.9. The green lines with triangle ending points show the range of all possible forensicabilities at each M for different $\mathcal{Q}_M$'s. As we can expect, quantization step sizes play an important role in determining forensicabilities. We will analyze this effect in later sections. In Fig. 3.9, we also plot the line of $I_{\lambda,\mathcal{Q}_M}(X;\underline{Y}) = \log_2 M$, which is the upper bound of forensicability for uniform priors, indicating perfect detection. Despite variations of forensicabilities for different $\mathcal{Q}_M$'s, the gap between the highest reachable forensicability and its upper bound becomes more obvious when M increases. This indicates that, as M increases, even when we encounter the scenario with the highest forensicability, i.e., the case having the best detection performance, we still cannot obtain perfect detection. Furthermore, the distance of the best performance to perfect detection will be larger with the increase of M. Therefore, when M increases, it will be much harder to detect the exact number of applied JPEG compressions, which validates our theory.

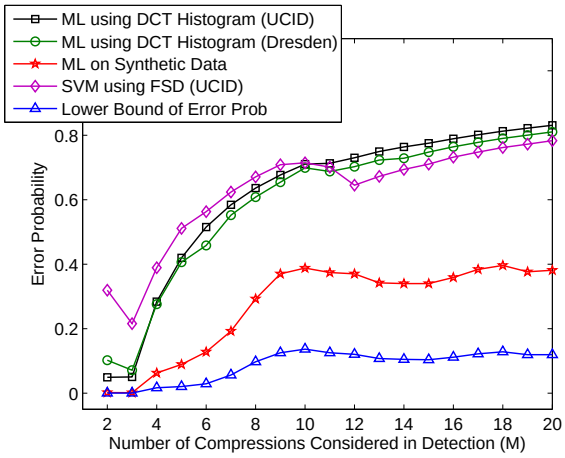
3.3.3 Estimation Error Probability Lower Bound

According to theorem 2, forensicability determines the lower bound of error probabilities. In this section, we perform several experiments to examine the effec-

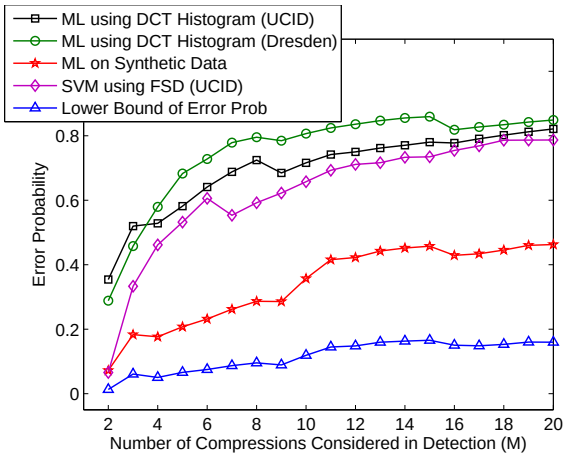
tiveness of the lower bound by comparing the theoretical lower bound of all possible error probabilities with the experimental error probability obtained from specific estimators. We perform this comparison on two examples of $\mathcal{Q}_{20}$, which are constructed by randomly selecting quantization step sizes from the candidate set in (3.30).

The first estimator we examine is the maximum likelihood estimator. The experimental error probabilities of this estimator on real images are obtained as follows. For each $M \in [2, 20]$, $\mathcal{Q}_M$ is obtained as the last M quantization step sizes in $\mathcal{Q}_{20}$. The 1338 uncompressed images from the UCID database are first used to construct a test database. Specifically, for each M, we JPEG compress each of the 1338 images M times using quality factors, whose quantization step sizes in the $(2, 3)$ subband are $\{q_{20-M+1}, \ldots, q_{20}\}$. The resulting 1338 images compose the data set of M times compressed images. Then, normalized DCT coefficient histograms in the $(2, 3)$ subband are extracted for analysis. Their theoretical distributions are also calculated based on $\mathcal{Q}_{20}$ and the estimated λ's from their uncompressed versions.

Given the assumption of uniform priors and the proposed conditional distribution of a normalized histogram given the theoretical distribution in (3.18) and (3.19), the maximum likelihood estimator is used to estimate the number of compressions for each M. Specifically, when M hypotheses of X are considered in the system, let m be the actual number of compressions that an image has gone through. Its normalized DCT coefficient histogram is denoted as $\underline{y}_m$. Then the maximum likelihood



(a)



(b)

Figure 3.10: Experimental error probabilities of several estimators comparing with the theoretical lower bound of error probabilities, where two randomly selected $\mathcal{Q}_{20}$'s are taken as examples: (a) $\mathcal{Q}_{20} = \{..., 8, 11, 13, 6, 5\}$ and (b) $\mathcal{Q}_{20} = \{..., 11, 9, 7, 8, 13\}$. Estimators used in experiments are, in order of displayed legends, maximum likelihood estimator using DCT coefficient histogram on UCID database, Dresden databases, and synthetic data; support vector machine using first significant digit of DCT coefficients on UCID database.

estimator for m is

$$\hat{m} = \underset{1 \le m^* \le M}{\arg\max} \ \mathbb{P}(\underline{Y}_{m^*} = \underline{y}_m), \tag{3.31}$$

where the distribution of $\underline{Y}_{m^*}$ is given in (3.18) and (3.19).

To examine the experimental result for different databases, we applied the above maximum likelihood estimator on another database, the Dresden Image Database [39]. This database contains 1491 unprocessed images, with each has size of 2000×3008 or larger. We can see from Fig. 3.10 that the estimator performs similarly on these two databases.

Furthermore, we examine the error probability of maximum likelihood estimator on synthetic data. The synthetic database is generated as follows. First, we take λ as the mean value of 1338 uncompressed images in UCID database. Based on $\mathcal{Q}_M$ and λ, we calculate the theoretical distribution of the DCT coefficient for each M. Then, based on the conditional probability of the observed histograms given these theoretical distributions, we generate 1000 synthetic observed histograms for each M to compose a test database. To calculate the experimental error probability for each M, maximum likelihood is used to obtain the estimation results of the number of compressions for each synthetic histogram in the test database. We can see from Fig. 3.10 that the error probabilities obtained from synthetic data are lower than those obtained from real data.

Another estimator we examine is the forensic technique in [66], where the histogram of first significant digit (FSD) of DCT coefficients is used to train a support vector machine. Because the histogram of the FSD of DCT coefficients can

be obtained directly from the histogram of the DCT coefficients, this estimator is eventually based on the feature of DCT coefficient histograms and fits our model. The estimation results of using this estimator on subband $(2, 3)$ is also plotted in Fig. 3.10. We can see that as the number of considered compressions increases, the performance of this estimator becomes comparable to the maximum likelihood estimator.

Last, for every M, the theoretical lower bound of error probabilities is calculated for each image, i.e., each estimated λ, using (3.2), then we take the mean value and plot it in Fig. 3.10.

Both examples in Fig. 3.10(a) and Fig. 3.10(b) show that the error probability of specific estimators are higher than the theoretical lower bound, which verifies the validity of our proposed lower bound. For the example in Fig. 3.10(b), most experimental results are worse than those in Fig. 3.10(a), even when detecting double compressions, i.e., $M = 2$. This matches the results in forensic literatures of detecting double compressions, which shows difficulty when the detected image has a secondary compression quality factor lower than the primary one [74]. The distance between the experimental error probability of one specific estimator and the theoretical error probability lower bound of all estimators suggests the existence of better estimators or better features.

3.3.4 Maximum Number of Detectable Compressions

Given the error probability lower bound, we can determine what is the maximum number of compressions investigators can detect by using corollary 2. First, based on Theorem 2, we use the highest reachable forensicability for each M to calculate the minimum lower bound of error probabilities for all possible compression quality factors. The calculation results are shown in Table 3.1. From this table we can see that, for double compression detection where $M = 2$, the lower bound of error probability is approximately 0 (note that it is not exactly zero, it is just smaller than the precision of Matlab processor), which matches the result of existing techniques [74]. Furthermore, the table shows that the minimum lower bound of error probability increases dramatically with M.

Table 3.1: $\min_{\mathcal{Q}_M} P_e^0$ for different M.

M	2	3	4	5	6
$\min_{\mathcal{Q}_M} P_e^0$	0	3.9×10^{-9}	5×10^{-5}	2.1×10^{-4}	0.0016

Then, to determine the point where we cannot perfectly detect any more compressions, we adopt the concept of expected perfect detection defined in definition 2 and use the conclusion in corollary 2. For example, if the forensic investigator performs experiments on a test database of size $S = 5000$, then because $\min_{\mathcal{Q}_4} P_e^0 < 1/S = 2 \times 10^{-4}$ but $\min_{\mathcal{Q}_5} P_e^0 > 2 \times 10^{-4}$, we claim that no expected perfect detection exists for $M > 4$.

Furthermore, by noticing that

$$\frac{1}{\min_{\mathcal{Q}_4} P_e^0} = 20000, \qquad \frac{1}{\min_{\mathcal{Q}_5} P_e^0} = 4762, \tag{3.32}$$

we have the following conclusion. For any database of size bigger than 4762 and smaller than 20000, expectedly, no perfect detection can be achieved for detecting more than 4 times of JPEG compressions. In other words, for typical sizes of database, investigators can only perfectly detect up to 4 times of JPEG compressions using DCT coefficient feature.

We note that, since we are analyzing the minimum lower bound of error probability, which is the best performance we may get from all estimators and all compression quality factors, these results only provides an upper limit of investigators' capability. In other words, "cannot perfectly detect 5 compressions" does not mean "can perfectly detect 4 compressions for sure". Our theorem tells what we cannot do rather than what we can do.

It is also noted that, for databases bigger than 20000, the maximum number of compressions can be detected may be less than 4. It implies that the number of detectable compressions depends on the test database size. It is reasonable because, as the database size goes bigger, there will be higher probability that we may meet an instance that is hard to detect and thus error may occur.

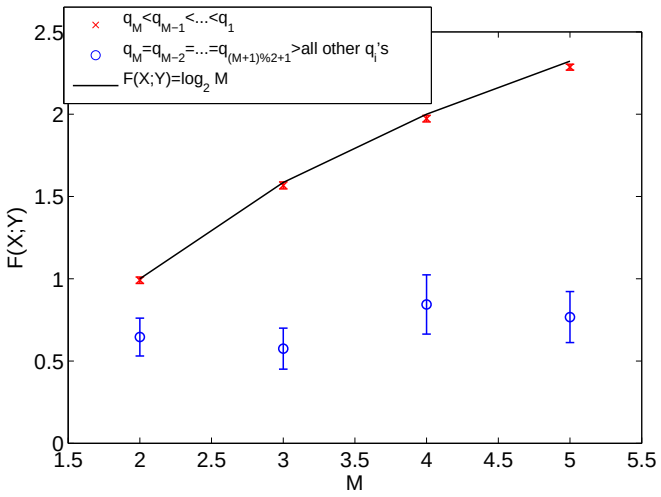


Figure 3.11: Patterns of $\mathcal{Q}_M$ yielding the highest and lowest forensicabilities.

3.3.5 Quality Factor Patterns having the Highest and Lowest Forensicabilities

As Fig. 3.9 shows, forensicability varies significantly with $\mathcal{Q}_M$. In order to characterize this effect, we examine all combinations of quantization step sizes and their forensicabilities. From there, we find the patterns of $\mathcal{Q}_M$ which will yield the highest and lowest forensicabilities, as they are shown in Fig. 3.11.

We find that, if the next compression always uses a higher quality factor than the previous one, forensicabilities will be the highest, i.e., they are easiest to be detected. Denote the set of quality factors yielding the highest forensicabilities as $\mathcal{Q}^h$, then

$$\mathcal{Q}^h = \{\mathcal{Q}_M | q_m < q_{m-1}, \forall 1 < m \leq M, M \in \mathbb{Z}^+\}. \tag{3.33}$$

To explain this phenomenon, let us examine a DCT coefficient histogram of an image that has been compressed m times using decreasing quantization step sizes $a_1 > a_2 >$

$\ldots > a_m$ in the concerned subband. Recall the discussion in Section 3.2.1, the singly quantized coefficients D_1 obeys a quantized Laplace distribution with quantization step size a_1. Then, given that the next quantization step size is smaller than the current one, when re-quantizing this histogram, every bin will remain its original value but be shifted to its nearby integer multiple of a_2. Zeros may be introduced into the histogram of D_2, but all nonzero histogram bins will be the same as those in the histogram of D_1. Similar analysis applies for the following quantizations. Therefore, the normalized DCT histogram after m times of quantizations will have all of its nonzero bins being equal to those after the first quantization.

For detecting M times of compressions with quantization step sizes $\mathcal{Q}_M$, we are distinguishing the following M hypotheses on the DCT coefficient histogram:

$$\begin{cases} H_1 : 1 \text{ time of quantization by } q_M, \\ H_2 : 2 \text{ times of quantizations by } \{q_{M-1}, q_M\} \text{ in order}, \\ H_3 : 3 \text{ times of quantizations by } \{q_{M-2}, q_{M-1}, q_M\}, \\ \vdots \\ H_M : M \text{ times of quantizations by } \{q_1, q_2, \ldots, q_M\}. \end{cases} \tag{3.34}$$

It is easy to notice that, for different hypotheses, the first quantization step sizes are different. Thus, for case of $q_1 > q_2 > \ldots > q_M$, theoretically, the nonzero bins of the normalized histogram obtained from one hypothesis are completely different from those obtained from another hypothesis. Furthermore, there may also have cases where a location of a zero histogram bin in one hypothesis has a nonzero bin in another hypothesis. This will further enlarge the disparity of DCT histograms obtained from different hypotheses. Therefore, the complete distinguishability of

theoretical distributions of DCT coefficients among different hypotheses results in the easiest detection and the highest forensicability.

The compression quality factors resulting in the lowest forensicabilities, as it is shown in Fig. 3.11, are those which use same quality factors periodically. More specifically, denote the set of quality factors yielding the lowest forensicabilities as $\mathcal{Q}^l$. We have found that

$$\mathcal{Q}^l = \{\mathcal{Q}_M | q_M = q_{M-2} = \ldots = q_{(M+1)\%2+1} > \text{all other } q_i's, M \in \mathbb{Z}^+\}, \quad (3.35)$$

where % is a remainder operator. The reason can be explained by the following theorem.

Theorem 3. *Given a quantized DCT coefficient D_{m-2} with the last quantization step size as q_{m-2}. We further quantize it two more times using quantization step sizes q_{m-1} then q_m. The obtained coefficient is denoted as D_m. If the quantization step sizes satisfy $q_m = q_{m-2} > q_{m-1}$, then the DCT coefficient remains the same after these two more compressions, i.e., $D_m \equiv D_{m-2}$.*

Proof. Take any possible value of $D_{m-2} = l_{m-2}q_{m-2}$, where $l_{m-2} \in \mathbb{Z}$, after the two quantizations, we obtain

$$D_m = round\left(round\left(\frac{l_{m-2}q_{m-2}}{q_{m-1}}\right)\frac{q_{m-1}}{q_m}\right)q_m. \quad (3.36)$$

Given that $\forall A \in \mathbb{R}, A - 1/2 < round(A) \leq A + 1/2$, we have

$$\begin{aligned}
\frac{D_m}{q_m} &> round\Big(\Big(\frac{l_{m-2}q_{m-2}}{q_{m-1}} - \frac{1}{2}\Big)\frac{q_{m-1}}{q_m}\Big) && (3.37)\\
&= round\Big(l_{m-2} - \frac{1}{2}\frac{q_{m-1}}{q_m}\Big) && (3.38)\\
&> round\Big(l_{m-2} - \frac{1}{2}\Big) && (3.39)\\
&> l_{m-2} - 1, && (3.40)
\end{aligned}$$

where (3.38) and (3.39) are obtained from the condition $q_{m-2} = q_m > q_{m-1}$. Since $\frac{D_m}{q_m}$ is an integer, we obtain $\frac{D_m}{q_m} \geq l_{m-2}$. Similarly, we can prove that $\frac{D_m}{q_m} \leq l_{m-2}$. Thus,

$$D_m = l_{m-2}q_m = D_{m-2}. \quad (3.41)$$

□

Given the above theorem, the M hypotheses in (3.34) can be reduced to only singly quantized hypothesis and double quantized hypothesis. Specifically, all odd numbered hypotheses will be identical to each other. While all even numbered hypotheses will be simplified to 2 times of quantization with different primary quantization step sizes. Furthermore, for the simplified double quantization hypotheses, the second quantization step size is larger than the first one, which is harder for estimation compared to its opposite case. Therefore, such a pattern of compression quality factors is the hardest to be detected, and thus has the lowest forensicability. Moreover, since the estimation performance will always be similar to a double compression detection regardless of how many compressions investigators really want to detect, forensicability almost remains the same as M increases.

3.3.6 Optimal Strategies for Forgers and Investigators

The fundamental measurement of forensicability can also be used to obtain the optimal strategies for both investigators and forgers. In this multiple compression detection system, investigators try to detect the number of compressions forgers have done on an image. Thus, investigators can choose examined subbands to maximize forensicability, while forgers have the right of choosing compression quality factors to minimize forensicability. Given that forensicability is a function of both subband parameter λ and compression quality factors $\mathcal{Q}_M$, we model the optimal strategies for forensic investigators and anti-forensic forgers in this multiple compression detection system as

$$\delta_F = \arg\max_{(i,j)} \mathbb{E}_{\mathcal{Q}_M}\Big[F_{\lambda_{(i,j)},\mathcal{Q}_M}(X;\underline{Y})\Big], \tag{3.42}$$

$$\delta_{AF} = \arg\min_{\mathcal{Q}_M} \mathbb{E}_{\lambda_{(i,j)}}\Big[F_{\lambda_{(i,j)},\mathcal{Q}_M}(X;\underline{Y})\Big], \tag{3.43}$$

respectively, where $(i,j), i,j \in [1,8]$, denotes the subband index.

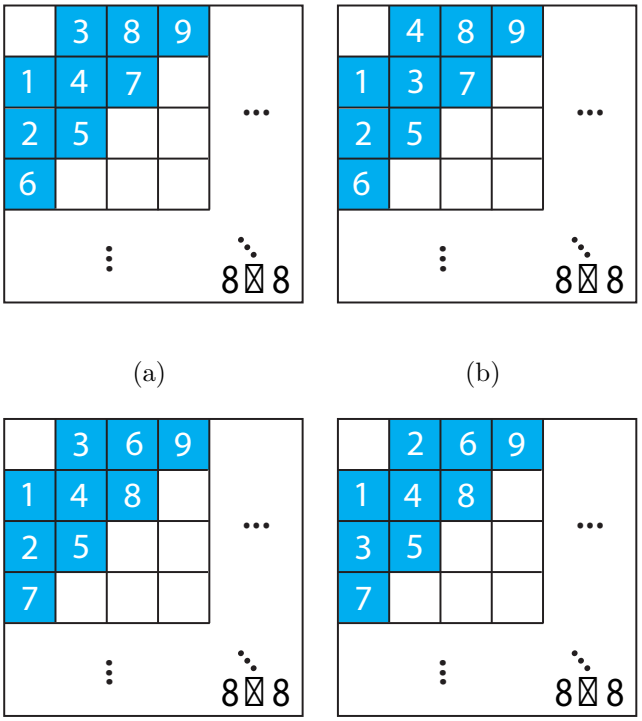
Since we have just discussed the effect of compression quality factors on forensicability, let us obtain the optimal strategy for forgers (3.43). From the discussion in last subsection, we notice that the patterns of compression quality factors yielding the highest and lowest forensicabilities do not depend on the subband parameter λ. Instead, the results are merely dependent on how the DCT coefficients are quantized. Thus, regardless of which subband or subbands investigators will choose, $\mathcal{Q}^l$ will always yield the lowest forensicability. Thus, we obtain the optimal strategy for forgers is

$$\delta_{AF} = \mathcal{Q}^l. \tag{3.44}$$

We note that, when $M = 2$, we have $\delta_{AF} = \mathcal{Q}^l = \{\mathcal{Q}_2 | q_1 < q_2\}$, which is opposite to the pattern of $\mathcal{Q}^h$. This result matches our early work on the concealability-rate-distortion tradeoff of compression anti-forensics, where we found that forgers would prefer to use a lower secondary quality factor instead of a higher one in their second compression [21].

To obtain the optimal strategy for investigators, we take $\lambda_{(i,j)}$ as the mean value of all estimated λ's from the $(i,j)^{th}$ subband coefficients of 1338 uncompressed images in the UCID database. We examine the cases of detecting 2, 3, 4 and 5 times of compressions, i.e., we take $M \in [2,5]$. For each M, $\mathcal{Q}_M$ for the $(2,3)$ subband is still constructed by randomly selecting quantization step sizes from the candidate set $\{5,6,7,8,9,11,13\}$ in (3.30). Given that the compression quality factors corresponding to these quantization step sizes are $\{82,78,75,70,67,60,55\}$, $\mathcal{Q}_M$ for other subbands can also be determined from their corresponding quantization tables. Then, for each of the 63 alternating current (AC) DCT subbands, forensicabilities are calculated for all $\mathcal{Q}_M$'s, whose number of possibilities can reach $(7 \times 6^4 =)9072$ when $M = 5$. We assume that investigators do not know the priori of the compression quality factors used by forgers. Thus, for each subband, $\mathbb{E}_{\mathcal{Q}_M}\left[F_{\lambda_{(i,j)},\mathcal{Q}_M}(X;\underline{Y})\right]$ is calculated as the mean value of forensicabilities with respect to different $\mathcal{Q}_M$'s.

By comparing the expected value of forensicabilities for all 63 subbands, we order them in descending order and take the top 9 subbands to show in Fig. 3.12. Our results show that, the top 9 subbands yielding the highest forensicabilities remain the same when detecting different numbers of compressions, though their orders are slightly different. Thus, if investigators take the best 9 subbands for



(a) (b)

(c) (d)

Figure 3.12: The best 9 DCT subbands (shown as blue cells) for detection, which yield the highest forensicabilities for (a) $M = 2$, (b) $M = 3$, (c) $M = 4$ and (d) $M = 5$. Numbers 1 through 9 represent the order of these subbands regarding their forensicabilities from the highest to the lowest.

detection, their the optimal strategy, which is denoted as $\delta_F^{(9)}$, will be

$$\delta_F^{(9)} = \{(2,1),(1,2),(3,1),(1,3),(2,2),(4,1),(1,4),(3,2),(2,3)\}. \tag{3.45}$$

It matches the set of subbands that many successful double compression forensic techniques have used in their algorithms [74]. This result gives theoretical support of why we use those subbands for detecting double compressions. It also suggests that we should continue to use these subbands to detect 3, 4 or 5 times of compressions. Furthermore, the ranks on these subbands tell us which subband contains more forensic information and which one will give us the most trustful result.

3.3.7 Forensicabilities for Image Outliers

Given that forensicability depends on the Laplace parameter λ of DCT coefficients, it may also vary for different types of images. While our results were obtained by choosing a representative λ value and thus can be considered as the most expected performance for natural images, there are some outliers that are much harder or much easier to be detected. For example, if an image is underexposed and most of its pixels are equal to zero, then it would be very hard to detect the number of compressions on this image.

To track the change of the Laplace parameter λ for different images, we examine natural images from both the UCID database (1338 images) and the Dresden image database (1491 images). Fig. 3.13 shows the histogram of λ in the $(2,3)$ subband of these 2829 images. We can see that most images have their λ values close to 0.1, which was chosen as the representative value of λ in Section 3.3.4.

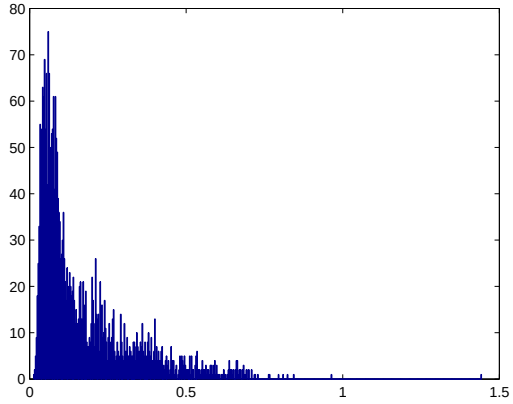


Figure 3.13: Histogram of λ in subband (2,3) of images from UCID and Dresden databases.

In order to examine forensicabilities for other images, we take two extreme cases of $\lambda = 0.02$ and $\lambda = 0.7$ to obtain the bounds of performance. Table 3.2(a) and 3.2(b) show the minimum error probability lower bound for different numbers of compressions when $\lambda = 0.02$ and $\lambda = 0.7$, respectively. By comparing these two tables with Table 3.1, we can see that the minimum lower bound of error probabilities $\min_{\mathcal{Q}_M} P_e^0$ increases with λ, and thus forensicability decreases with λ. This matches the results in the previous subsection where forensicability decreases for higher frequency subbands which have higher values of λ. This is because for large λ's, the DCT coefficient histograms have high kurtosis and low variances. Most bins in these histograms have small values that can be severely contaminated by noise. Only a few histogram bins have large enough values that can be used for estimation. Thus, little information can be extracted from these histograms. By following the analysis in Section 3.3.4 we can infer that, if we have a database of size 10000, then for image outliers whose $\lambda = 0.02$, investigators can detect up to 7

Table 3.2: $\min_{\mathcal{Q}_M} P_e^0$ for different M when (a) $\lambda = 0.02$ and (b) $\lambda = 0.7$.

(a)

M	2	3	4	5	6	7	8
$\min_{\mathcal{Q}_M} P_e^0$	0	0	1.9×10^{-9}	1.1×10^{-7}	2.2×10^{-6}	3.7×10^{-5}	5.5×10^{-4}

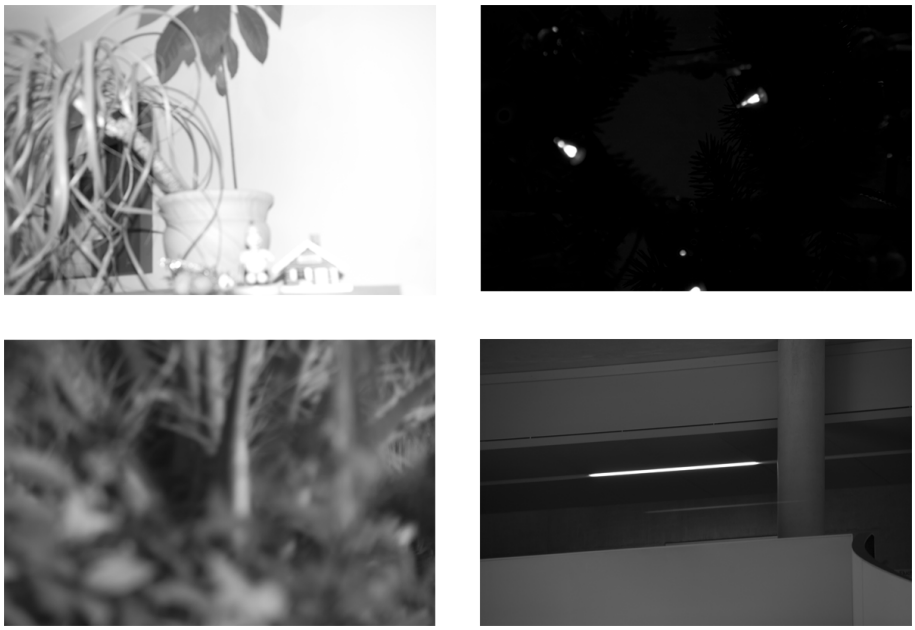
(b)

M	2	3
$\min_{\mathcal{Q}_M} P_e^0$	1.4×10^{-5}	0.0018

times of compressions. While for image outliers whose $\lambda = 0.7$, we can only detect 2 times of compressions.

Lastly, in order to see what types of images are outliers, we select some representative images from each extreme case and show them in Fig. 3.14. As it is shown in Fig. 3.14(a), the outliers having the lowest λ's and the highest forensicabilities are highly textured images whose AC components have sufficient information to be used for forensic detection. On the other hand, the outliers having the highest λ's and the lowest forensicabilities are images having a large amount of smooth or uniform areas but few textured regions. As it is shown in Fig. 3.14(b), this phenomenon may be caused by overexposure, underexposure, strong blurring, or little textured image content.

(a)



(b)

Figure 3.14: Representative image outliers in UCID and Dresden databases with (a) $\lambda \cong 0.02$ and (b) $\lambda \geq 0.7$.

3.4 Summary

In this chapter, we proposed an information theoretical framework to explore the fundamental limit of operation forensics. In this framework, we defined forensicability in operation detection forensics as the maximum information that features contain about operations. Based on this measure, we obtained the lower bound of error probabilities for any estimators using these features. Furthermore, by introducing the concept of expected perfect detection, we were able to determine the limit of how many operations we can successfully detect. To show the effectiveness of our framework, we applied it to the case of detecting multiple JPEG compressions using DCT coefficient histogram features. By appropriate modeling of the features, we calculated forensicabilities and concluded that, under typical settings of forensic analysis where the size of the testing database is less than 20000, at most 4 times of compressions were perfectly detectable. Furthermore, based on this fundamental measurement, we found the patterns of compression quality factors holding the highest and lowest forensic information. Lastly, the optimal strategies for investigators and forgers were discussed using forensicability.

Chapter 4

Fundamental Limits in Order Forensics

In recent years, many forensic techniques have been proposed to identify the use of different manipulation operations, such as compression [31, 55, 76], resizing [48, 77], contrast enhancement [86], blurring [13, 89, 96] and so on [22, 85, 88]. Most of these techniques expose specific fingerprints of the considered operations and implicitly assume that no other operations were applied [13, 31, 48, 77, 85, 86, 89, 96]. However, in reality, it is often the case that multiple operations are needed to complete a forgery. For example, if a forger wants to replace a person's face in an image using another person's face from another image, he or she may need to apply the following operations. First, the forger may need to apply resizing and contrast enhancement operations on the new face to make it match the size and color of the old face in the target image. Then, to avoid visible boundaries of the new face to the background of the target image, blurring may be applied to smooth the transition. At last, this forged image may be compressed for storage or transmission.

There have been some forensic techniques designed to identify the existence of a single operation in a certain operation chain [6, 14, 36, 55, 76]. Double compression detectors were developed to detect the existence of the first compression in a processing chain of two consecutive compressions [55, 76]. In [6], an improved double compression detector was proposed for the processing chain of two compressions

with resizing in between. Specifically, two hypotheses were considered: whether the image was single JPEG compressed, or it was double JPEG compressed with resizing applied in the middle. Authors in [36] considered a similar scenario where linear contrast enhancement was interleaved with the two compressions. In addition, the contrast enhancement detector proposed in [14] can effectively detect this operation when it was applied on previously JPEG compressed images.

While these techniques considered multiple operations, their goal is to identify the existence of a specific operation in a certain processing chain. Nothing can be inferred about the order of operations from these techniques. However, when multiple different operations may be applied on the multimedia content, detecting the order of these operations is equally important with identifying the existence of each operation. By detecting the order of operations, we can obtain the complete processing history of multimedia content. Furthermore, given that different operations may be applied by different forgers, detecting the order may also help us identify who manipulated the multimedia content and when it was manipulated. For example, if investigators receive an image that was downloaded from the internet and may be maliciously blurred by either the uploader or the downloader. Suppose that when an image is uploaded, resizing is needed to make the image fit the website standard. In this scenario, detecting the order of blurring and resizing can tell us who manipulated the image and when it was manipulated.

Few works have been done on detecting the order of operations. In [84], a forensic technique has been developed to detect the order of resizing and contrast enhancement. Nevertheless, the order of operations is not always detectable due

to the interplay between operations. One reason would be that when multiple operations are applied on the multimedia content, later applied operations may affect, or even destroy, the fingerprints of earlier applied operations. For example, if JPEG compression or Gaussian noise is applied after contrast enhancement, the fingerprints of contrast enhancement would be too weak to be detected [14].

Therefore, a natural question would be "when can and cannot we detect the order of operations?" To answer this question, we formulate the order detection problems into multiple hypotheses estimation problems. For such problems, we propose an information theoretical framework and mutual information based criteria to determine whether or not we can distinguish all the considered hypotheses. Furthermore, for those indistinguishable cases, this criterion can tell us which hypotheses are confused with each other and why they are confused. In addition, we also give a rigorous definition of the existence of conditional fingerprints. To verify the effectiveness of the proposed framework and criteria, we apply them on two known forensic problems to show that the obtained results match those published in existing works. Then, the proposed framework and criteria are applied on the problem of detecting of the order of resizing and blurring to obtain when we can or cannot detect their orders.

The remaining of this chapter is organized as follows. Section 4.1 explains the fact that the order of operations is not always detectable and presents our system model for determining when they can or cannot be detected. The mutual information based criteria are proposed in section 4.2. Section 4.3 presents our proposed estimation scheme for detecting the order of resizing and blurring. To

demonstrate the effectiveness of our proposed framework and criteria, section 5.5 provides simulation results for both existing forensic problems and the problem examined in section 4.3. Lastly, section 4.5 concludes our work.

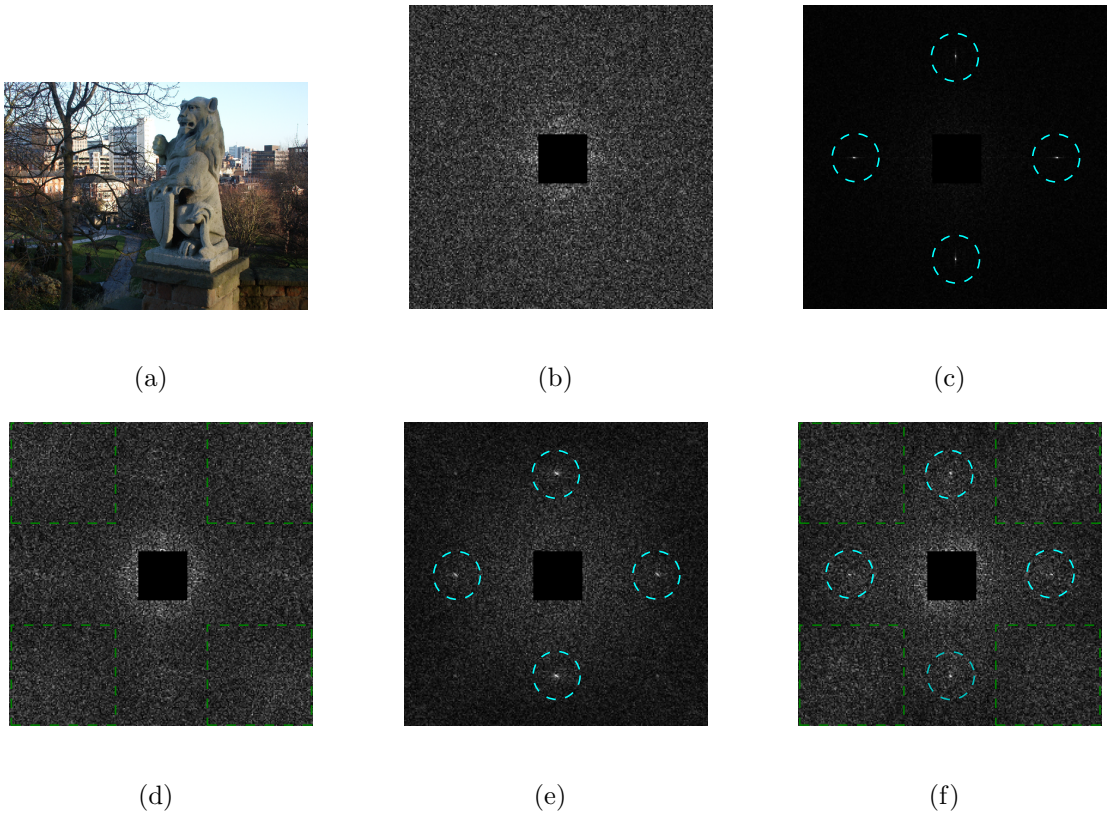
4.1 System Model

In this section, we first give an example to illustrate that the order of operations is not always detectable. Then, based on the analysis on the example, we propose our information theoretical framework for generalized multiple hypotheses estimation problems.

4.1.1 Order of Operations May Not be Detectable

When multiple operations are applied on multimedia content, the effect of later applied operations on earlier applied ones may lead to the undetectability of the order of operations. For example, let us consider a processing chain which may contain two operations: resizing and blurring. To detect the order of resizing and blurring, we assume that a given image may fall into one of the following hypotheses:

$$\begin{aligned} &H_0 : \text{The image is unaltered,} \\ &H_1 : \text{The image is resized only,} \\ &H_2 : \text{The image is blurred only,} \\ &H_3 : \text{The image is blurred then resized,} \\ &H_4 : \text{The image is resized then blurred.} \end{aligned} \tag{4.1}$$



(a) (b) (c)

(d) (e) (f)

Figure 4.1: Fingerprints for detecting the order of resizing and blurring. (a) and (b) are the original image and the DFT of its p-map, respectively. (c) - (f) show the DFT of the p-map of (c) the resized image, (d) the blurred image, (e) the blurred then resized image, and (f) the resized then blurred image. Resizing factor is 1.5 (upscaling). Gaussian blur is used with variance 1. Regions of interests are highlighted by dotted squares and circles.

The order of resizing and blurring can be distinguished if we can distinguish all the above five hypotheses.

Fig.s 4.1(b) - 4.1(f) show the different fingerprints of each hypothesis in the discrete Fourier transform (DFT) of an image's p-map. P-map is a probability matrix with each element representing the probability of the corresponding image pixel correlated with its neighbor pixels [77]. This matrix is widely used in detecting the resizing operation [48]. This is because that the linear interpolation process in resizing will lead to periodic characteristics of the p-map. Thus, when we take the DFT of the p-map, we would observe four distinct peaks in the corresponding spectrum, as they are shown in Fig. 4.1(c).

We assume that the blur operation is applied by using a linear filter on an image. Let $I_{i,j}$ and $I'_{i,j}$ denote a pixel located at (i, j) of the unaltered image I and the blurred image I', respectively. Then, each blurred pixel is a linear combination of its original neighbor pixels, for example,

$$I'_{i,j} = \sum_{|\triangle i|<k} \sum_{|\triangle j|<k} \alpha_{\triangle i,\triangle j} I_{i+\triangle i,j+\triangle j}, \tag{4.2}$$

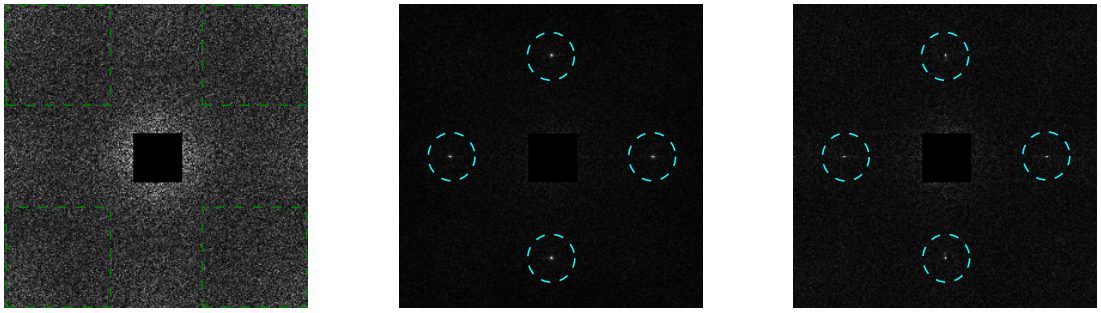
where $\alpha_{\triangle i,\triangle j}$ denote the coefficient of the linear filter. Consider a neighbor pixel of $I'_{i,j}$, such as $I'_{i+1,j}$, from (4.2) we know that this pixel is a linear combination of pixels $\left\{I_{i+\triangle i,j+\triangle j} \middle| |\triangle i - 1| < k, |\triangle j| < k\right\}$. Thus, $I'_{i,j}$ and $I'_{i+1,j}$ are both dependent on pixels $\{I_{i+\triangle i,j+\triangle j} \big| -k+1 < \triangle i < k, |\triangle j| < k\}$. This analysis shows that, though blurring does not give direct correlations between neighboring pixels, the neighbor pixels of a blurred image may still be correlated due to the overlapped dependency on the pixels of the original image. This alteration on pixel correlations cause by

blurring may result in certain fingerprints in the p-map of the blurred image.

To see how pixel correlations are altered by blurring, we examined the p-map and its DFT of a blurred image. We have found that, in the DFT of the p-map, a blurred image has an increase of energy in high frequency component while the energy in frequency domain of an unaltered image is monotonically decreasing as the frequency increases. We can see these fingerprints by comparing Fig. 4.1(d) with Fig. 4.1(b). These fingerprints can be used to detect blurring, as we will discuss in section 4.3.

Furthermore, even when the image is previously resized, these fingerprints of blurring may still exists, as it is shown in Fig. 4.1(f). However, if resizing is applied after blurring, the fingerprints of blurring will be hardly detectable, as it is shown in Fig. 4.1(e). Nevertheless, either resizing then blurring or blurring then resizing, the DFT of the p-map is more noisy than that of the only resized case. We can see this by comparing Fig. 4.1(e) and Fig. 4.1(f) with Fig. 4.1(c). Then, the peak signal to noise ratio (PSNR) at the four peaks corresponding to the resizing fingerprints may be used to distinguish the hypotheses containing both resizing and blurring and the hypothesis of pure resizing. Specific detection schemes will be discussed in section 4.3.

Based the fingerprints of each hypothesis presented in Fig.s 4.1(b) - 4.1(f), we can design algorithms to distinguish all hypotheses in (4.1) and thus detect the order of resizing and blurring. However, for some cases, these fingerprints are very weak and hardly detectable. Fig. 4.2 shows a confusing example where the same image in Fig. 4.1(a) was examined but the blurring effect is weaker that in Fig. 4.1.



(a) (b) (c)

Figure 4.2: A confusing example that we may not be able to detect the order. Plotted are DFTs of the p-map of (a) the blurred image, (b) the blurred then resized image, and (c) the resized then blurred image when resizing factor is 1.5 and the variance of Gaussian blur is 0.7. Regions of interests are highlighted by dotted squares and circles.

We can see that, though we may still be able to observe the fingerprints of blurring, we can hardly tell the difference between the blurred then reszied image and the resized then blurred image. Therefore, in this case, we may not be able to detect the order of resizing and blurring.

4.1.2 Information Theoretical Model for Multiple Hypotheses Estimation Problems

Given that the order of operations is not always detectable, a natural question would be "when can we and cannot we detect the order of operations?" To answer this question, we first consider a generalized multiple hypotheses estimation problem as follows.

Consider a forensic problem where we have assumptions on the possible hy-

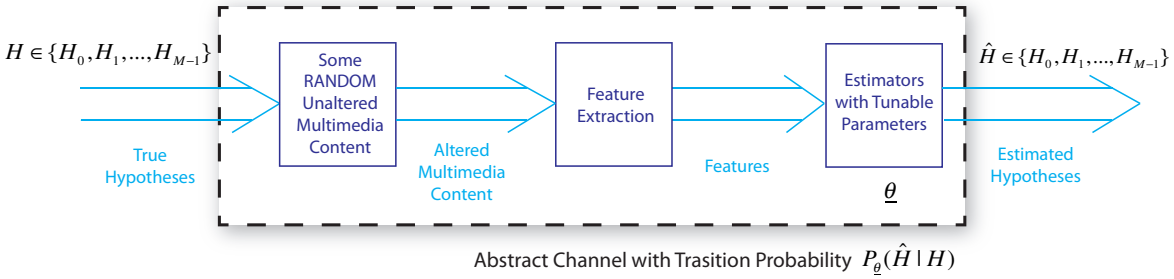


Figure 4.3: A typical process of estimating the hypotheses.

potheses that the given multimedia content may belong to. Here are some examples of hypotheses assumptions made in existing works:

- In [48] of detecting resizing, two hypotheses are considered: H_0, the image is unaltered; H_1, the image is resized.
- In [76] of detecting double JPEG compression, also two hypotheses are considered: H_0, the image is single JPEG compressed; H_1, the image is double JPEG compressed.
- In [84], five hypotheses are considered, similarly to those in (4.1) with the blurring operation substituted by the operation of contrast enhancement.

In order to distinguish these considered hypotheses, investigators go through the following typical steps [20]. First, possible fingerprints that can be used to distinguish each hypothesis are found. Then, based on these fingerprints, features are extracted from an examining image. At last, a set of estimators with tunable thresholds or parameters will be used to make the final decision of the estimated hypothesis based on the extracted features. Fig. 4.3 shows this process.

Given certain features, estimators with different parameters will lead to different estimation performance. For example, when detecting the resizing operation, we can use the fingerprints of four peaks in the DFT of an image's p-map [48]. In order to measure these fingerprints, the cumulative periodogram of the Fourier domain p-map is calculated. Then, its maximum gradient is used to make the final estimation by comparing it to a certain threshold [48]. This tunable threshold is the parameter of the estimator which determines the detection rate and false alarm rate of the estimator. The overall performance of the estimator can be measured by plotting a receiver operating characteristic (ROC) curve, which contains all reachable pairs of detection rates and false alarm rates.

Let us consider another forensic problem where more than two hypotheses are involved. In [84] of detecting the order of resizing and contrast enhancement, five hypotheses are considered in the analysis. A tree structured estimator was proposed to obtain the estimated hypothesis. In each tree node, an intermediate decision is made by comparing a certain statistic extracted from an image with a tunable threshold. The final estimation performance is determined by all the thresholds used in the detection scheme tree. Because there are more than two hypotheses considered in the problem, multiple ROC curvse would be needed to show the estimation performance [84]. This representation of multiple ROC curves could be problematic when we want to characterize the overall estimation performance, or to compare the estimation performance of different estimators.

In order to give a simple yet effective characterization, we use a transition probability matrix between the true hypotheses and the estimated hypotheses to

represent the performance of the estimator with certain parameters. This representation is applicable for general multiple hypotheses estimation problems. Furthermore, it can be used to compare the estimation performance of different estimators, which will be discussed in the next section.

The transition probability matrix is defined as follows. Let $\mathcal{H} = \{H_0, H_1, ..., H_{M-1}\}$ denote the set of considered hypotheses in a multiple hypotheses estimation problem. Then the true hypothesis and the estimated hypothesis, denoted as H and $\hat{H}$ respectively, belong to this set. Based on certain features, a set of estimators with different parameters $\underline{\theta}$, denoted as $d_{\underline{\theta}}$, are used to obtain the estimated hypotheses. For each choice of $\underline{\theta}$, the performance of the specific estimator is presented by a transition probability matrix $\mathbf{T}(\underline{\theta})$ with each element denoting the conditional probability of an estimated hypothesis given a true hypothesis, i.e.,

$$\mathbf{T}_{i,j}(\underline{\theta}) = \mathbb{P}_{\underline{\theta}}(\hat{H} = H_j | H = H_i), \quad 0 \leq i, j < M. \tag{4.3}$$

With this definition, we propose a feature dependent channel to characterize the relationship between true hypotheses and estimated hypotheses. The channel characteristics, i.e., the transition probabilities between input and output (4.3), is specified by the parameters of the set of estimators, as it is shown in Fig. 4.3.

4.2 Information Theoretical Criteria

As we have formulated the order detection problem into a multiple hypotheses estimation problem, our goal is to tell when we can and cannot distinguish all considered hypotheses. Given that for certain features, estimators with different

parameters yield different estimation performance, a natural thought would be to see if the best estimator is able to distinguish all hypotheses. Then, the question becomes "which estimator is the best?" In this section, we first propose a mutual information criterion to determine the best estimator. Then, based on this criterion, our information theoretical criterion of when we can and cannot distinguish all hypotheses are proposed.

4.2.1 Mutual Information Criterion to Obtain the Best Estimator

For simple hypothesis problems where only two hypotheses are considered in the problem, the best estimator can be obtained by comparing ROC curves. However, when more than two hypotheses are considered, because each estimator has multiple ROC curves [84], comparison among ROC curves becomes complicated. Specifically, when tuning the parameters, some ROC curves may become better while others may be worse. It is hard to say which parameters yield the best overall estimation performance. Therefore, we need a criterion to quantify the overall performance of estimators.

In the previous section, we have used a transition probability matrix to characterize the performance of an estimator. The relationship between true hypotheses and estimated hypotheses has been modeled as an abstract channel with transition probabilities $\mathbf{T}(\underline{\theta})$. Then, for the best estimator, we would expect that the estimated hypotheses contain the maximum information about the true hypotheses. Since mutual information is a measure of the information that the output of a chan-

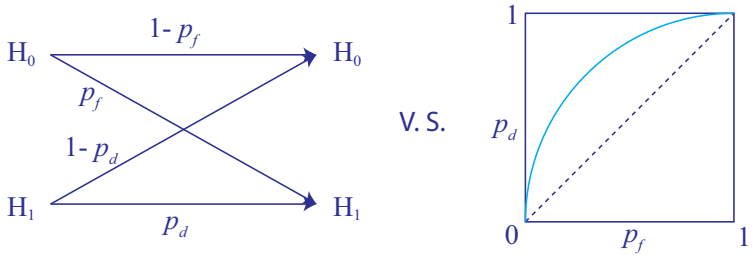


Figure 4.4: Compare a simple hypothesis channel and a ROC curve.

nel contains about the input, we define the best estimator based on this measure.

Definition 3. *In a problem of estimating hypothesis $H \in \mathcal{H}$, estimators $d_{\underline{\theta}_1}$ and $d_{\underline{\theta}_2}$ are based on the same features. Let $\hat{H}$ denote the estimated hypothesis. $\mathbf{T}(\underline{\theta}_1)$ and $\mathbf{T}(\underline{\theta}_2)$ are transition probability matrices of estimator $d_{\underline{\theta}_1}$ and $d_{\underline{\theta}_2}$, respectively. Let p_H denote the priors of H. Then,* detector $d_{\underline{\theta}_1}$ is better than $d_{\underline{\theta}_2}$, w.r.t. the mutual information criterion, *when*

$$I_{p_H,\mathbf{T}(\underline{\theta}_1)}(H;\hat{H}) > I_{p_H,\mathbf{T}(\underline{\theta}_2)}(H;\hat{H}), \tag{4.4}$$

for the cases where we know the priors of H; or

$$\max_{p_H} I_{p_H,\mathbf{T}(\underline{\theta}_1)}(H;\hat{H}) > \max_{p_H} I_{p_H,\mathbf{T}(\underline{\theta}_2)}(H;\hat{H}), \tag{4.5}$$

for the cases where we do not know the priors. $I(H;\hat{H})$ denotes the mutual information between H and $\hat{H}$.

This criterion enables us to evaluate the best estimator for general multiple hypotheses estimation problems, especially when more than two hypotheses are considered. Furthermore, the properties of this measurement also match those of the traditionally used ROC curves for simple hypothesis estimation problems.

In order to show the effectiveness of our proposed mutual information criterion in simple hypothesis cases, we make a comparison between our information theoretical characterization of the estimation performance and the traditional ROC curve, as it is shown in Fig. 4.4. In these cases, $\mathcal{H} = \{H_0, H_1\}$. Let p_d and p_f denote the detection rate and false alarm rate as follows,

$$p_d = \mathbb{P}(\hat{H} = H_1 | H = H_1), \tag{4.6}$$

$$p_f = \mathbb{P}(\hat{H} = H_1 | H = H_0). \tag{4.7}$$

Given that uniform priors are usually implied when plotting ROC curves [88], we take $\mathbb{P}(H = H_0) = \mathbb{P}(H = H_1) = 1/2$.

The performance of an estimator with a specific parameter $\underline{\theta}$ can be represented by either the value of mutual information $I_{p_H, \mathbf{T}(\underline{\theta})}(H; \hat{H})$ or a point (p_f, p_d) in the ROC curve. Because the mutual information only depends on p_d and p_f under the assumption of uniform priors, for comparison, we use a function $S(p_f, p_d)$ to denote the mutual information. Then, we have the following properties of the mutual information criterion.

Lemma 2. *For a simple hypothesis channel with uniform priors, since each parameter $\underline{\theta}$ of the estimator dictates a pair of p_d and p_f, we use a simplified function $S(p_f, p_d)$ to indicate the mutual information $I_{p_H, \mathbf{T}(\underline{\theta})}(H; \hat{H})$. Then, this function has the following properties.*

1. $S(p_f, p_d) = S(p_d, p_f)$

2. $p_{d1} > p_{d2} > p_f \Rightarrow S(p_f, p_{d1}) > S(p_f, p_{d2})$

3. $\arg\min_{p_f} S(p_f, p_d) = p_d$*;* $\arg\min_{p_d} S(p_f, p_d) = p_f$

Proof. Given the uniform priors and likelihood probabilities, we can obtain the probabilities of estimated hypotheses as

$$\mathbb{P}(\hat{H} = H_0) = 1 - \tfrac{1}{2}(p_d + p_f), \tag{4.8}$$

$$\mathbb{P}(\hat{H} = H_1) = \tfrac{1}{2}(p_d + p_f). \tag{4.9}$$

Then, the mutual information for the simple hypothesis channel is

$$\begin{aligned} S(p_f, p_d) &= I_{p_H, \mathbf{T}(\underline{\theta})}(H; \hat{H}) \\ &= h\big(\tfrac{1}{2}(p_d + p_f)\big) - \tfrac{1}{2}h(p_f) - \tfrac{1}{2}h(p_d), \end{aligned} \tag{4.10}$$

where $h(p)$ denotes the binary entropy function of $(p, 1-p)$. From (4.10), we can see that $S(p_f, p_d)$ is a symmetric, i.e., $S(p_f, p_d) = S(p_d, p_f)$ Then, the first property is proved.

To prove the second property, we take the partial derivative of $S(p_f, p_d)$ w. r. t. p_d,

$$\begin{aligned} \frac{\partial S(p_f, p_d)}{\partial p_d} &= -\tfrac{1}{2}\ln\big(\tfrac{1}{2}(p_d + p_f)\big) + \tfrac{1}{2}\ln\big(1 - \tfrac{1}{2}(p_d + p_f)\big) \\ &\quad + \tfrac{1}{2}\ln p_d - \tfrac{1}{2}\ln(1 - p_d) \\ &= \tfrac{1}{2}\ln\left(1 + \frac{p_d - p_f}{(p_d + p_f)(1 - p_d)}\right) \end{aligned} \tag{4.11}$$

Then, when $p_d \geq p_f$, the above derivative is greater than zero. Thus, for $p_d \geq p_f$, $S(p_f, p_d)$ is an increasing function of p_d. The second property is also proved.

Furthermore, we can also see from (4.11) that $p_d = p_f$ is the minimal of $S(p_f, p_d)$, i.e., $\arg\min_{p_d} S(p_f, p_d) = p_f$. Similarly, we can also prove that $\arg\min_{p_f} S(p_f, p_d) = p_d$. Then, the last property is proved and it concludes our proof. □

Since $S(p_f, p_d)$ measures the estimation performance of an estimator with detection rate p_d and false alarm rate p_f, it can also be interpreted as the measure of estimation performance at a point (p_f, p_d) of a ROC curve. Then, the properties in lemma 2 can be interpreted in the following way.

1. Estimation performance of each point in a ROC is symmetric along the random guess line $p_d = p_f$.

2. For points above the random guess line, given a certain false alarm rate, an estimator with a higher detection rate is a better estimator.

3. The worst performance is the random guess line.

We can easily see that the above properties match those in ROC curves.

Therefore, our proposed mutual information criterion is consistent with a ROC curve for simple hypothesis test cases. Furthermore, it gives a criterion for cases where more than two hypotheses are considered. Our mutual information criterion is a general measurement of the estimation performance for multiple hypotheses estimation problems.

4.2.2 Information Theoretical Criteria for Multiple Hypotheses Estimation Problems

Given the measurement of estimators' performance, we can determine the distinguishability of considered hypotheses by checking if the best performed estimator can distinguish all hypotheses. Specifically, if priors are uniform, we examine the

likelihood probabilities of the best estimator and check for each true hypothesis, if the detection probability is greater than any misdetection probabilities. If nonuniform priors are assumed or we do not know the priors, we examine the posteriori probabilities of the best estimator.

Definition 4. *For a multiple hypotheses estimation problem, where considered hypotheses are* $\mathcal{H} = \{H_0, H_1, ..., H_{M-1}\}$. *Let* H *and* $\hat{H}$ *denote the true hypothesis and the estimated hypothesis, respectively. Assume that priors are positive. Then,* under the mutual information criterion, all hypotheses can be distinguished by estimators $d_{\underline{\theta}}, \underline{\theta} \in \mathbb{R}^k$, *if and only if the following conditions are satisfied.*

- *If priors are uniform, the conditions are*

$$H_i = \arg\max_{t \in \mathcal{H}} \mathbb{P}_{\underline{\theta}^*}(\hat{H} = t | H = H_i),\ \forall i = 0, 1, ..., M-1; \tag{4.12}$$

- *If priors are nonuniform or unknown, the conditions are*

$$H_i = \arg\max_{t \in \mathcal{H}} \mathbb{P}_{\underline{\theta}^*}(H = t | \hat{H} = H_i),\ \forall i = 0, 1, ..., M-1, \tag{4.13}$$

  and

$$\mathbb{P}_{\underline{\theta}^*}(\hat{H} = H_i) > \epsilon,\ \ \forall i = 0, 1, ..., M-1, \tag{4.14}$$

where ϵ *is a small constant and* $\underline{\theta}^*$ *are parameters of the best estimator w.r.t. the mutual information criterion. That is, if we know the priors*

$$\underline{\theta}^* = \arg\max_{\underline{\theta}} I_{p_H, \mathbf{T}(\underline{\theta})}(H; \hat{H}). \tag{4.15}$$

If we do not know the priors,

$$(\underline{\theta}^*, p_H^*) = \arg\max_{\underline{\theta}, p_H} I_{p_H, \mathbf{T}(\underline{\theta})}(H; \hat{H}). \tag{4.16}$$

This criterion can be used to determine when we can or cannot distinguish all hypotheses. Furthermore, by examining the conditions in (4.12) and (4.13), we are able to tell which hypotheses are confused with each other when we cannot distinguish all hypotheses.

Definition 5. *For the problem in definition 4,* two hypotheses, H_i and H_j, $i \neq j$, are confused with each other *when we obtain the following results.*

- *If priors are uniform,*

$$\begin{aligned} \mathbb{P}_{\underline{\theta}^*}(\hat{H} = H_i | H = H_j) &\geq \mathbb{P}_{\underline{\theta}^*}(\hat{H} = H_j | H = H_j), \\ \text{or, } \mathbb{P}_{\underline{\theta}^*}(\hat{H} = H_j | H = H_i) &\geq \mathbb{P}_{\underline{\theta}^*}(\hat{H} = H_i | H = H_i). \end{aligned}$$

- *If priors are nonuniform or unknown,*

$$\begin{aligned} \mathbb{P}_{\underline{\theta}^*}(H = H_i | \hat{H} = H_j) &\geq \mathbb{P}_{\underline{\theta}^*}(H = H_j | \hat{H} = H_j), \\ \text{or, } \mathbb{P}_{\underline{\theta}^*}(H = H_j | \hat{H} = H_i) &\geq \mathbb{P}_{\underline{\theta}^*}(H = H_i | \hat{H} = H_i). \end{aligned}$$

The reason of why hypotheses may be confused with each other is related to the strength of fingerprints or conditional fingerprints [84]. As our examples in (4.1) and at the beginning of section 4.1.2 show, each hypothesis represents an operation chain. This operation chain can be an empty chain which denotes the hypothesis of unaltered multimedia content. It can also be a single operation chain or a multiple operations chain. We first define fingerprints and conditional fingerprints of operation chains as follows.

Definition 6. *Consider an operation chain and its corresponding hypothesis, denoted as* $\underline{S}_i$ *and* H_i *respectively. Let* $\underline{S}_\emptyset$ *and* $H_\emptyset$ *denote the empty operation chain, and the hypothesis of unaltered multimedia content. If* $\underline{S}_i \neq \underline{S}_\emptyset$*, then* the fingerprints of $\underline{S}_i$ *are a set of features that can be used to distinguish* $\{H_i, H_\emptyset\}$*. Next, we consider another operation chain, denoted as* $\underline{S}_j$*. If* $\underline{S}_i$ *is a sub-chain of* $\underline{S}_j$*, let* $\underline{S}_{j\backslash i}$ *denote the operation chain of* $\underline{S}_j$ *excluding* $\underline{S}_i$*.* $H_{j\backslash i}$ *is denoted as the corresponding hypothesis of* $\underline{S}_{j\backslash i}$*. Then,* the conditional fingerprints of $\underline{S}_i$ given $\underline{S}_j$ *are a set of features that can be used to distinguish the following hypotheses:*

$$\{H_{j\backslash i}, H_i, H_j\}.$$

Remarks: To better understand the difference between fingerprints and conditional fingerprints, we give the following example. Let $\underline{S}_i$ and $\underline{S}_j$ denote the operation chain of only contrast enhancement and contrast enhancement then resizing, respectively. Then, $\underline{S}_{j\backslash i}$ represents the operation chain of only resizing. When detecting contrast enhancement, the fingerprints we commonly used are the high frequency components of the DFT of the pixel histogram [86]. However, these cannot be the conditional fingerprints of contrast enhancement given contrast enhancement then resizing [84]. This is because that resized images and contrast enhanced then resized images, i.e., $\{H_{j\backslash i}, H_i\}$, cannot be distinguished by examining the fingerprints of contrast enhancement. In [84], the conditional fingerprints of contrast enhancement given contrast enhancement then resizing are two features. One is the maximum gradient of the periodogram of the Fourier transformed p-map, which is the fingerprint of resizing. The other feature is the distance of normalized histograms

between the full image and the down-sampled image [84]. By using these two features, we can distinguish resized images, contrast enhanced images, and contrast enhanced then resized images, i.e., $\{H_{j\backslash i}, H_i, H_j\}$.

Based on fingerprints and conditional fingerprints, forensic techniques can be designed to detect operations and their orders [84, 88]. Similarly, in a multiple hypotheses estimation problem, the existence of required fingerprints and conditional fingerprints enables us to distinguish all hypotheses. Based on definition 5, rigorous definitions of the existence of fingerprints and conditional fingerprints can be obtained as follows.

Definition 7. *Consider a multiple hypotheses estimation problem where* $\mathcal{H} = \{H_0, H_1, ..., H_{M-1}\}$ *Let* $H_\emptyset$ *denote the empty chain hypothesis. For a hypothesis* $H_i \in \mathcal{H}$, $H_i \neq H_\emptyset$, *let* $\underline{S}_i$ *denote the processing chain represented by this hypothesis. Then,* the fingerprints of $\underline{S}_i$ exist *if*

H_i *is* not *confused with* $H_\emptyset$ *by definition 5.*

Now, consider another hypothesis H_j, $j \neq i$ *and* $H_j \neq H_\emptyset$, *the processing chain it represents is* $\underline{S}_j$. The fingerprints of $\underline{S}_i$ and $\underline{S}_j$ are different *if*

H_i *is* not *confused with* H_j *by definition 5.*

Furthermore, if $\underline{S}_i$ *is a sub-chain of* $\underline{S}_j$, *let* $H_{j\backslash i}$ *denote the hypothesis representing* $S_{j\backslash i}$, *then* the conditional fingerprints of $\underline{S}_i$ given $\underline{S}_j$ exist *if*

any two of $\{H_{j\backslash i}, H_i, H_j\}$ *are* not *confused by definition 5.*

Having all concepts defined for general hypotheses estimation problems, let us examine the special cases of detecting the order of operations to see when we can

and cannot detect the order. For example, if two operations A and B are involved in a forensic problem, we consider the following hypotheses for a given multimedia content.

$$\begin{aligned} H_0 &: \text{It is unaltered,} \\ H_1 &: \text{It is altered by } A \text{ only,} \\ H_2 &: \text{It is altered by } B \text{ only,} \\ H_3 &: \text{It is altered by } B \text{ then } A, \\ H_4 &: \text{It is altered by } A \text{ then } B. \end{aligned} \tag{4.17}$$

Then, the order of A and B can be detected if and only if we can distinguish all hypotheses in (4.17) by definition 4. This requires that any two hypotheses cannot be confused with each other by definition 5. That is, the following conditions on fingerprints and conditional fingerprints should hold by definition 7.

- Fingerprints of A, B, $A \to B$, and $B \to A$ exist.
- Conditional Fingerprints of A given $A \to B$ exist.
- Conditional Fingerprints of B given $B \to A$ exist.
- Fingerprints of $A \to B$ and $B \to A$ are different.

4.3 Detecting the Order of Resizing and Blurring

To demonstrate the effectiveness of our framework and criteria, we examine a case study of detecting the order of resizing and blurring. In this section, we formulate this problem using our information theoretical framework and propose

an estimation algorithm to detect the order of resizing and blurring. Then, the detectability of their order is obtained by applying our mutual information based criteria on experimental data in section 4.4.3.

To detect the order of resizing and blurring, we are distinguishing the five hypotheses in (4.1). Thus, this is a multiple hypotheses estimation problem and can be analyzed using our information theoretical framework.

As we have shown in Fig. 4.1, these five hypotheses can be distinguished by their unique fingerprints in the DFT of an image's p-map. Specifically, if an image is unaltered, the DFT of its p-map can be considered a noisy signal whose energy gradually gets lower as the frequency goes higher, as it is shown in Fig. 4.1(b). If the image is only resized, the DFT of its p-map reveals distinct peaks in the corresponding spectrum, which is determined by resizing factors [77], as it is shown in Fig. 4.1(c). The lowest frequency region is removed to emphasize the fingerprints.

If an image is only blurred, due to its alteration on neighbor pixels correlations, we have found that it also shows fingerprints in the DFT of its p-map. The fingerprints of blurring are presented as an increase of the noisy energy in high frequency regions, as it is shown in Fig. 4.1(d).

When an image is altered by both resizing and blurring, regardless of the order, fingerprints of resizing are left in the image. We can still see four peaks in the DFT of the p-map, as they are shown in Fig.s 4.1(e) and 4.1(f). However, the PSNR is lower than that of the case where only resizing is applied, because blurring increases the noise energy of the DFT of the p-map.

To furthermore distinguish the hypotheses of resizing then blurring and blur-

ring then resizing, we can examine the fingerprints of blurring. Because only when blurring is applied after resizing, we can observe the fingerprints of blurring.

Based on these fingerprints, we take two features to distinguish these hypotheses. One feature is to detect the existence of four peaks and measure the strength of these peaks. The other feature is to capture the increase of noise energy in high frequency regions.

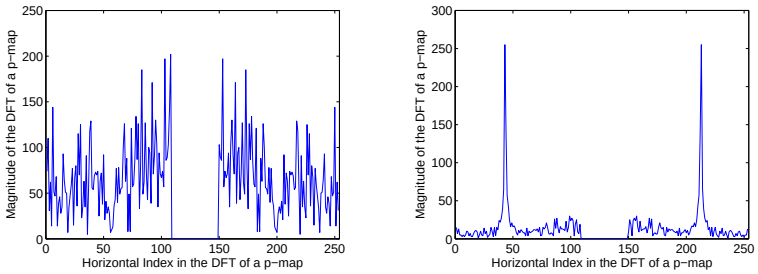
Feature 1: PSNR

As we have discussed above, the first feature can be taken as the PSNR of the DFT of an image's p-map. To calculate this measure, we first extract the central horizontal lines of the DFT of the p-maps from Fig.s 4.1(b)-4.1(e) and plot the magnitudes on these lines in Fig.s 4.5(a)-4.5(e), respectively. By appropriately choosing thresholds, this measure can be used to categorize the five hypotheses into three classes: unaltered or only blurred images; blurred then resized or resized then blurred images; only resized images; .

Specifically, let us take the resized then blurred case as an example, i.e., the signal in Fig. 4.5(d). Given the symmetry of the signal, we first consider the left half of the signal. Let y_l and x denote the magnitude and the index, respectively. Since the noise mean increases with the index, we first use the following linear regression model to make the noise mean uniform so that the peak is more prominent.

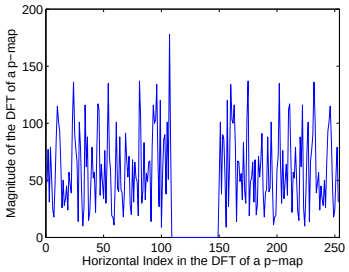
$$y_l = a_1 x + b_1 + n. \tag{4.18}$$

An example of the linear regression process is shown in the upper left figure of Fig.

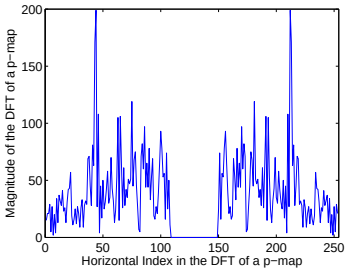


(a)

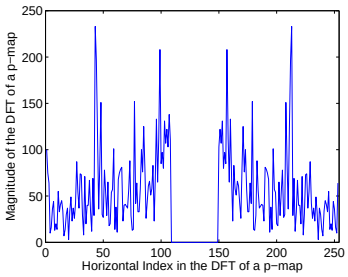
(b)



(c)



(d)



(e)

Figure 4.5: The central horizontal line of the DFT of the p-map of (a) an unaltered image, (b) a resized image, (c) a blurred image, (d) a blurred then resized image, and (e) a resized then blurred image.

4.6. After estimating the parameters as $\hat{a}_1$ and $\hat{b}_1$, we obtain the difference signal

$$d_l = y_l - \hat{a}_1 x - \hat{b}_1, \tag{4.19}$$

as it is shown in the bottom left figure of Fig. 4.6.

Then, the peak is detected from d_l by finding the coordinates of its maximum value (x_p, y_p). From the bottom left figure in Fig. 4.6 we can see that, the noise variance changes a lot as it is farther from the peak. Thus, instead of calculating the mean of the absolute value of noise in the whole range, we only consider the regions close to the peak:

$$PSNR_l = \frac{y_p}{\text{mean}_{0<|x-x_p|<\varepsilon}\Big(\big|d_l(x)\big|\Big)}. \tag{4.20}$$

Similar process is then applied to the right half of the signal in Fig. 4.5(d) to obtain $PSNR_r$. Then, the PSNR measurement for the central horizontal line of the DFT of a p-map is

$$PSNR_h = \max(PSNR_l, PSNR_r). \tag{4.21}$$

Given that the peaks also present in the central vertical line of the DFT of the p-map. We calculate the above PSNR measurement, denoted as $PSNR_v$, for the central vertical line signal as well. Then, the first PSNR feature used to distinguish hypotheses in (4.1) is

$$PSNR = \max(PSNR_h, PSNR_v). \tag{4.22}$$

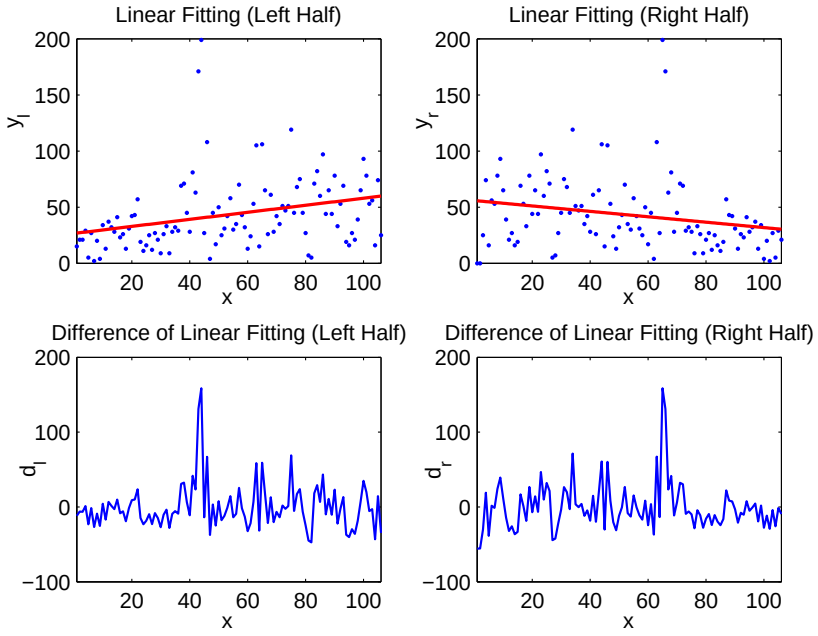


Figure 4.6: The process of how to calculate the PSNR from the central horizontal line of the DFT of a p-map. Take Fig. 4.5(d) as an example.

We can make the following estimation based on this feature,

$$\hat{H} = \begin{cases} H_0 \text{ or } H_2, & \text{if } PSNR < \tau_1, \\ H_3 \text{ or } H_4, & \text{if } \tau_1 \leq PSNR < \tau_2, \\ H_1, & \text{if } PSNR \geq \tau_2. \end{cases} \tag{4.23}$$

where τ_1 and τ_2 are tunable parameters.

Feature 2: noise energy pattern

To further distinguish H_2 from H_0 and H_4 from H_3, we examine the fingerprints of blurring. As shown in Fig. 4.1(d) and Fig. 4.1(f), when blurring is applied as the last operation, we would observe an increase of noise energy at high frequencies of the DFT of the p-map. In order to capture this change of noise energy, we calculate a noise energy pattern signal near the boundaries of the DFT of a p-map.

Specifically, let $Z = \{Z_{m,n}\}$ denote the magnitudes of the DFT of a p-map. The origin is located at the upper left corner of the matrix with size a by a. The noise energy signal, which is denoted as a matrix E, is first calculated as a summation of neighboring magnitudes in Z, i.e.,

$$E = Z \otimes \mathbf{1}_w, \tag{4.24}$$

where $\mathbf{1}_w$ is an all one matrix of size w by w, and $\otimes$ is a convolution operator. Then, we take a one dimensional signal y_e near the boundaries of E as the noise energy pattern signal:

$$y_e(x) = \frac{\left(E_{v,a/2+x} + E_{a/2+x,a-v} + E_{a-v,a/2-x} + E_{a/2-x,v}\right)}{4}, \tag{4.25}$$

where $v - a/2 \leq x < a/2 - v$ and $v = \lceil w/2 \rceil + 1$.

In Fig.s 4.7(a)-4.7(e), the dotted blue lines are noise energy pattern signals for the DFT of the p-maps in Fig.s 4.1(b)-4.1(f), respectively. We can see that the fingerprints of blurring result in an increase of the noise energy with $|x|$ for higher values of $|x|$, as shown in Fig.s 4.7(c) and 4.7(e). To measure these fingerprints, we use a second order polynomial model to fit the signal as

$$y_e = a_2 x^2 + b_2 x + c_2, \tag{4.26}$$

and see if the estimated function is convex or concave. The solid red lines in Fig. 4.7 are the estimated curves. If the estimated $\hat{a}_2$ is positive, then the noise energy pattern signal is estimated as a convex function. This indicates that the noise energy tends to increase with $|x|$ for higher $|x|$'s. Thus blurring fingerprints are detected. Otherwise, if the estimated function is concave, blurring fingerprints are

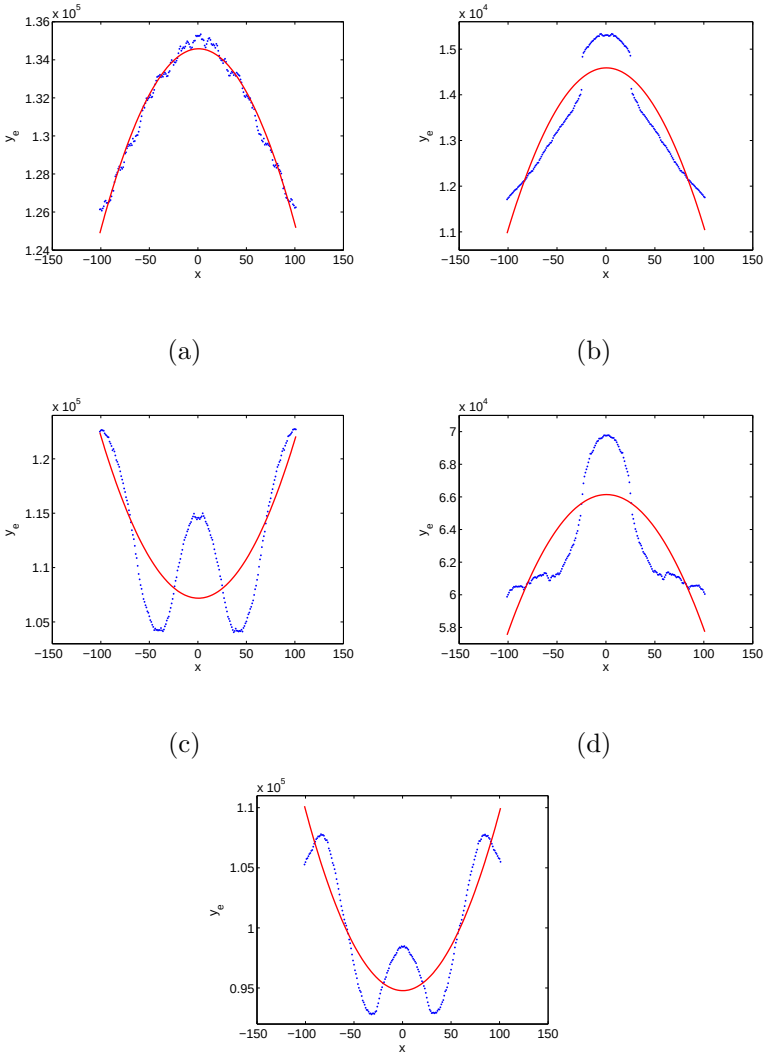


Figure 4.7: The noise energy pattern signal (dotted blue lines) extracted from the DFT of the p-map and their polynomial fitting curves (solid red lines) for (a) an unaltered image, (b) a resized image, (c) a blurred image, (d) a blurred then resized image, and (e) a resized then blurred image.

not detected. The estimation we can make by this feature is

$$\hat{H} = \begin{cases} H_0 \text{ or } H_1 \text{ or } H_3, & \text{if } \hat{a}_2 < 0, \\ H_2 \text{ or } H_4, & \text{if } \hat{a}_2 > 0. \end{cases} \tag{4.27}$$

Combining the estimation results from (4.23) and (4.27), our decision rule of the proposed estimator for detecting the order of resizing and blurring is

$$\hat{H} = \begin{cases} H_0, & \text{if } PSNR < \tau_1 \text{ and } \hat{a}_2 < 0, \\ H_1, & \text{if } PSNR \geq \tau_2, \\ H_2, & \text{if } PSNR < \tau_1 \text{ and } \hat{a}_2 > 0, \\ H_3, & \text{if } \tau_1 \leq PSNR < \tau_2 \text{ and } \hat{a}_2 < 0, \\ H_4, & \text{if } \tau_1 \leq PSNR < \tau_2 \text{ and } \hat{a}_2 > 0. \end{cases} \tag{4.28}$$

In our proposed algorithm, the estimation has two tunable parameters, and thus $\underline{\theta} = (\tau_1, \tau_2)$. Given the estimator and its parameters, we will apply our mutual information based criteria to simulation results to answer the question of "when can we and cannot we detect the order of resizing and blurring" in section 4.4.3.

4.4 Simulation Results

In this section, we conduct several simulations to demonstrate the effectiveness of our information theoretical framework and mutual information based criteria. We first examine two existing forensic problems, one simple hypothesis problem and one order detection problem, to verify the correctness of our framework and criteria. Then, the detection of the order of resizing and blurring is examined to show when we can and cannot detect the order of these two operations.

4.4.1 Detect Double JPEG Compression

Since our framework and criteria can be used for general multiple hypotheses estimation problems, we start with a well know simple hypothesis estimation problem in forensics, double JPEG compression detection [19,35,44,55,74,76]. We want to prove that the results obtained from our method match those from published literature.

To detect double JPEG compression, two hypotheses are considered in the analysis:

$$\begin{aligned} H_0 &: \text{The image is single JPEG compressed,} \\ H_1 &: \text{The image is double JPEG compressed.} \end{aligned} \tag{4.29}$$

There are many features that can be used to distinguish these hypotheses [19, 55, 74, 76]. All of them can yield over 90% detection rates for most JPEG compression quality factors. While our framework can be applied to any features and corresponding estimators, we use the first digit feature of DCT coefficients as an example to see if the results obtained from our framework match those in the existing work [55].

The estimator in [55] was proposed based on the double JPEG compression fingerprints in the first digit of DCT coefficients. Specifically, If an image is single JPEG compressed, the first digit of its DCT coefficients obeys a general Benford's law:

$$p(d) = N \log_{10}\left(1 + \frac{1}{s + d^q}\right), \;\; d \in \{1, 2, ..., 9\}, \tag{4.30}$$

where s and q are model parameters and N is a normalization factor. The first digit

d of a non-zero integer x is computed as

$$d = \left\lfloor \frac{x}{10^{\lfloor \log_{10} x \rfloor}} \right\rfloor, \tag{4.31}$$

where $\lfloor \cdot \rfloor$ is the floor rounding operation. If the image is double JPEG compressed, however, this law will not hold for the first digit of its DCT coefficients.

Given these fingerprints, an estimator for distinguishing hypotheses in (4.29) can be designed as follows. First, we obtain the normalized histogram of the first digit of DCT coefficients. Then, we use these statistics to estimate the general Benford's law and calculate the sum of squared errors (SSE) between the estimated distribution and the normalized histogram. The final decision is made by comparing the mean SSE of the 20 lowest frequency subbands with a tunable threshold θ as follows [55],

$$\hat{H} = \begin{cases} H_0, & \text{if mean SSE} < \theta, \\ H_1, & \text{if mean SSE} \geq \theta. \end{cases} \tag{4.32}$$

In order to determine whether we can detect double JPEG compression, we first generate a testing database using the 1338 unaltered images from the UCID database [82]. Specifically, these images are first JPEG compressed by quality factors from 50 to 95 with step size of 5 to obtain the single JPEG compressed image database. Then, each of the image in this database is re-compressed by the same set of quality factors to compose the double JPEG compressed image database. Let Q_1 and Q_2 denote the quality factors used in the first and second JPEG compression, respectively. Then, for each pair of Q_1 and Q_2, the testing database contains 1338 single compressed images using Q_2 and 1338 double compressed images using Q_1 then Q_2.

We first assume uniform priors for the two hypotheses as most literatures do [55, 74]. Then, using (4.4) in definition 3, we can obtain the best parameter θ^* that yield the highest mutual information between the estimated hypotheses and the true hypotheses. By checking the conditions (4.12) in definition 4, we determine whether we can distinguish these two hypotheses for a given pair of Q_1 and Q_2. The results of all combinations of Q_1 and Q_2 are shown in Fig. 4.8(a).

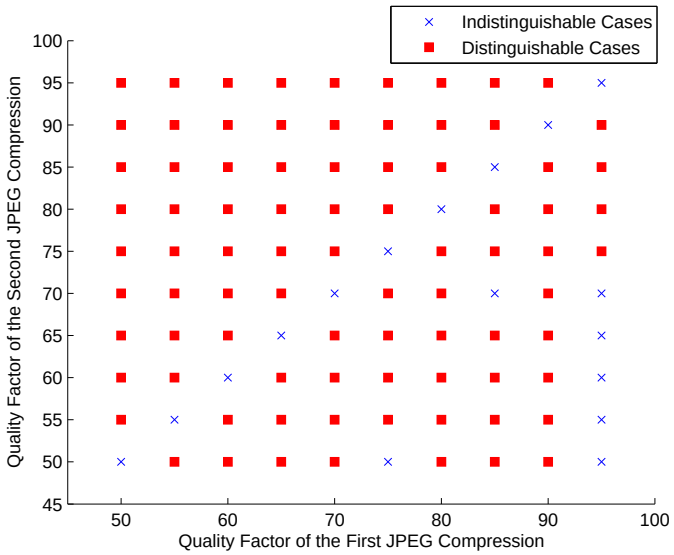
We can see that for most cases, double JPEG compression can be detected by using the proposed model. This matches the results in [55].

For indistinguishable cases, confused hypotheses are H_0 and H_1 by definition 5. This means that the conditional fingerprints of JPEG compression given the operation chain of double JPEG compression do not exist in these cases by definition 7. Specifically, this is because 1) $Q_1 = Q_2$, though there are other features that can be used to deal with this situation [44]; 2) the secondary quantization step size is a multiple integer of the first quantization step size for most of the extracted DCT subbands.

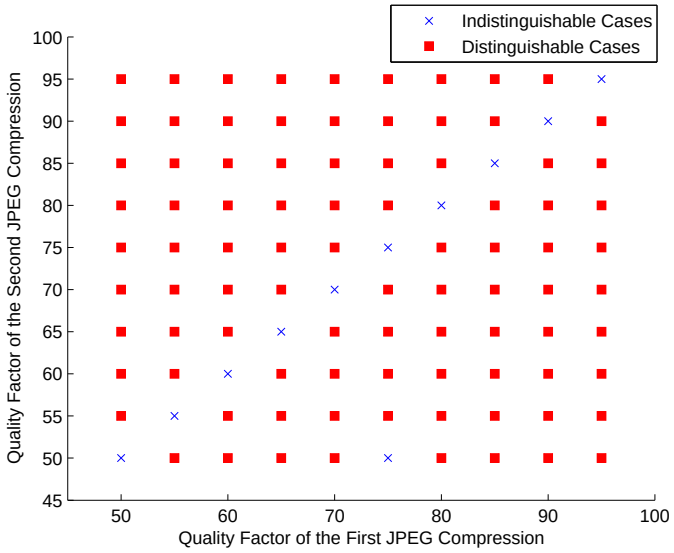
Note that we may be able to distinguish more cases if using a support vector machine (SVM) as the estimator and tune the position of the hyperplane as the parameters [55].

Then, we consider a general case where we do not know the priors of the two hypotheses. The best estimator would be determined by (4.5) in definition 3 and we should use the criterion (4.13) in definition 4 to determine whether we can distinguish the hypotheses. Fig. 4.8(b) shows the results under this assumption.

Since there are fewer constraints on the priors in the case of unknown priors,



(a)



(b)

Figure 4.8: Distinguishability test results of detecting double JPEG compression by applying our information theoretical framework and criteria. (a) priors are known and uniform. (b) priors are unknown.

the best estimator can yield higher mutual information in this case than that of the uniform priors case. Then, the best estimation performance we can get from unknown priors would be better than that from the uniform priors. Therefore, we have more distinguishable cases in Fig. 4.8(b) than those in Fig. 4.8(a).

4.4.2 Detect the Order of Resizing and Contrast Enhancement

The next case study we examine is the order detection of resizing and contrast enhancement, which contains more than two hypotheses [84]. In this forensic problem, five hypotheses are considered and needed to be distinguished:

$$\begin{aligned} &H_0 : \text{The image is unaltered,} \\ &H_1 : \text{The image is resized only,} \\ &H_2 : \text{The image is contrast enhanced only,} \\ &H_3 : \text{The image is contrast enhanced then resized,} \\ &H_4 : \text{The image is resized then contrast enhanced.} \end{aligned} \tag{4.33}$$

The fingerprints of H_3 and H_4 were found in [84] as follows. If an image is first resized then contrast enhanced, both fingerprints of resizing and contrast enhancement can be revealed from the image. However, if an image is first contrast enhanced then resized, only the fingerprints of resizing can be revealed. Nevertheless, we can still detect the previously applied contrast enhancement by examining a down-sampled image of the resized image. This is because that, if the resizing factor can be represented as a rational number $s = a/b$ such that $a, b \in \mathbb{N}$ and are mutually prime, then every a pixel in the resized image will occur at the same spatial location as a pixel in the original image. Therefore, the resizing operation

can be reverse engineered by down-sampling the image with factor $1/a$. If contrast enhancement is previously applied, then its fingerprints can be revealed from this down-sampled image.

Given these fingerprints, a tree structured estimation scheme was proposed in [84]. First, resizing fingerprints are examined [48]. In this step, the feature extracted is the maximum derivative of the cumulative periodogram calculated from the DFT of the p-map. We denote this feature as f_{rs}. If f_{rs} is greater than a threshold, denoted as α, it means that resizing has been applied on this image. Thus, we can estimate the hypothesis as one of $\{H_1, H_3, H_4\}$. Otherwise, the image belongs to either H_0 or H_2.

If resizing fingerprints have been detected from the image, then we can use the conditional fingerprints of contrast enhancement given contrast enhancement then resizing to detect the previously applied contrast enhancement [84]. The feature extracted is the distance of normalized pixel histograms between the full image and the down-sampled image with factor $1/a$. To obtain a, the resizing factor needs to be estimated [75], which involves the use of a training database and SVM. Let f_{cers} denote the feature extracted in this step. If f_{cers} is greater than a threshold λ, then previously applied contrast enhancement is detected in the resized image. Thus, the estimated hypothesis is H_3. Otherwise, the image belongs to either H_1 or H_4.

To distinguish H_2 from H_0 or H_4 from H_1, the fingerprints of contrast enhancement are examined [86]. The feature is taken from the high frequency components of the DFT of the normalized pixel histogram. Let f_{ce} denote this feature. To distinguish $\{H_0, H_2\}$, if f_{ce} is greater than a threshold β_1, then the image is estimated as

H_2. Otherwise, it is estimated as H_0. Similar decision is applied for distinguishing $\{H_1, H_4\}$, whose threshold is denoted as β_2.

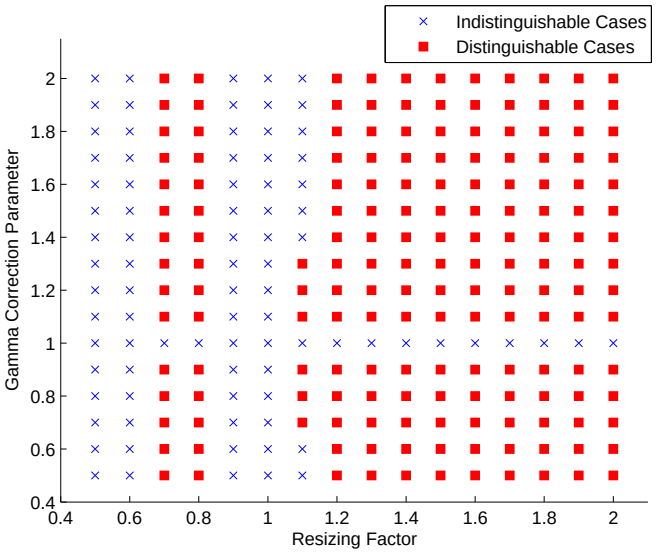
In summary, the estimation algorithm is as follows.

$$\hat{H} = \begin{cases} H_0, & \text{if } f_{rs} < \alpha \text{ and } f_{ce} < \beta_1, \\ H_1, & \text{if } f_{rs} \geq \alpha, f_{cers} < \lambda \text{ and } f_{ce} < \beta_2, \\ H_2, & \text{if } f_{rs} < \alpha \text{ and } f_{ce} > \beta_1, \\ H_3, & \text{if } f_{rs} \geq \alpha \text{ and } f_{cers} \geq \lambda, \\ H_4, & \text{if } f_{rs} \geq \alpha, f_{cers} < \lambda \text{ and } f_{ce} > \beta_2, \end{cases} \tag{4.34}$$

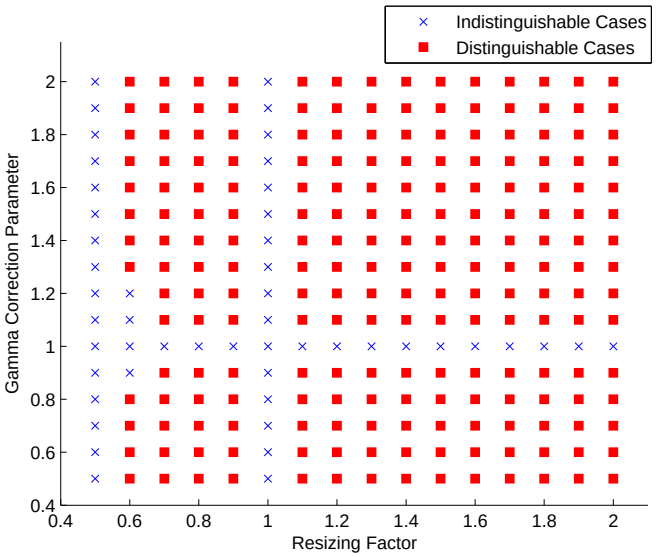
There are four tunable parameters and thus $\underline{\theta} = (\alpha, \lambda, \beta_1, \beta_2)$.

In order to know when we can and cannot detect the order of resizing and contrast enhancement, we use 1000 images from the UCID database to generate our test database. The rest 338 images are used to generate the training database for the resizing factor estimation step. We use gamma corrections with parameter γ to simulate the contrast enhancement operation [86]. For each $\gamma \in \{0.5, 0.6, ..., 2\}$ and $s \in \{0.5, 0.6, ..., 2\}$, the test database contains: 1000 unaltered images, 1000 resized images with scaling factor s, 1000 contrast enhanced images with gamma correction parameter γ, 1000 contrast enhanced then resized images, and 1000 resized then contrast enhanced images. Note that for $\gamma = 1$ or $s = 1$, contrast enhancement or resizing is not actually applied. Thus, we cannot distinguish all five hypotheses in these cases. To estimate the resizing factor, the training database for SVM contains 5070 (=338×15) images whose resizing factors are taken from $\{0.5, 0.6, ...0.9, 1.1, ..., 2\}$ [75].

We still consider two cases regarding the priors of the considered hypotheses.



(a)



(b)

Figure 4.9: Distinguishability test results of detecting the order of resizing and contrast enhancement by applying our information theoretical framework and criteria. (a) Priors are known and uniform. (b) Priors are unknown.

For uniform priors, the simulation results are shown in Fig. 4.9(a). While for the cases that we do not know priors, the results are shown in Fig. 4.9(b). As expected, when we do not have constraints on priors of the hypotheses, we have more distinguishable cases than that when uniform priors are assumed.

In [84], two examples of resizing factors and gamma correction parameters, $(s = 1.5, \gamma = 0.5)$ and $(s = 1.25, \gamma = 0.7)$, are examined in experiments. Specific estimation performance for these two pairs of parameters are plotted in five ROC curves. In both cases, authors in [84] have shown that the proposed estimator can successfully detect the order of resizing and contrast enhancement. To compare these results with those obtained by our framework, let us examine the uniform priors case. From Fig. 4.9(a), we can see that both $(s = 1.5, \gamma = 0.5)$ and $(s = 1.25, \gamma = 0.7)$ are distinguishable points. This shows that the results obtained by our approach match those in [84].

Besides the two example cases examined in [84], we obtain the detectability results for the whole range of resizing factors and gamma correction parameters. From these results, we have found that, though we can detect the order of resizing and contrast enhancement for most of the cases, there are a few indistinguishable cases. We examine these cases and use definition 5 to find which hypotheses are confused to make it indistinguishable. In addition, the reasons of why these hypotheses are confused is summarized as follows by definition 7.

- H_2 is confused with H_4 for the indistinguishable cases where $s = 1.1$. This means that the conditional fingerprints of resizing given resizing then contrast

enhanced do not exist in these scenarios. The effect of later applied contrast enhancement on the fingerprints of previously applied resizing is more obvious as the strength of contrast enhancement increases, i.e., for larger values of $|\gamma - 1|$.

- H_3 is confused with H_1 or H_4 when $s = 0.6, 0.9$. Given the tree structure of the estimation algorithm, this is due to the failure of distinguishing H_3 from H_1 and H_4 by the conditional fingerprints of contrast enhancement given contrast enhancement then resizing. This conditional fingerprints do not exist for these scenarios either because of the incorrect estimation of the resizing factor or due to the insufficient number of pixels extracted from the down-sampled image.

4.4.3 Detect the Order of Resizing and Blurring

By applying our information theoretical framework and criteria on double JPEG compression detection and the order detection of resizing and contrast enhancement, we have shown that the results obtained from our proposed framework match those in existing works. In this section, we examine the order detection of resizing and blurring and find when we can and cannot detect the order of these two operations.

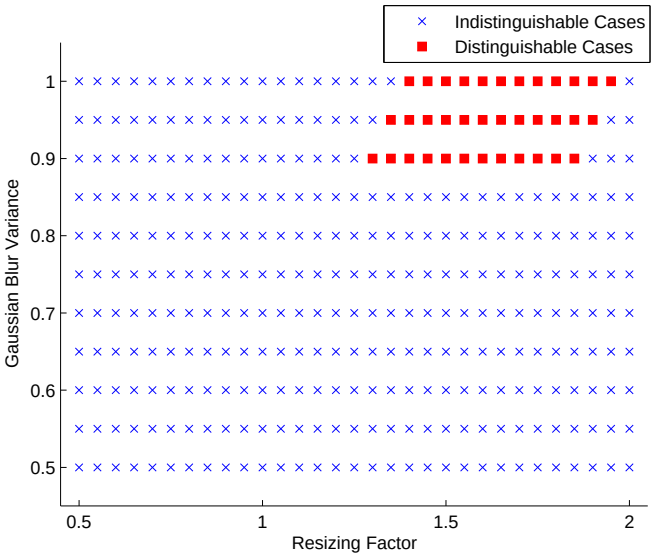
For forensic problems examined in previous sections 4.4.1 and 4.4.2, the considered hypotheses can be distinguished for most of the cases. However, as we have shown in Fig. 4.2, the order of resizing and blurring is not always detectable. Then, our framework and criteria can be used to determine when this order can and cannot

be detected.

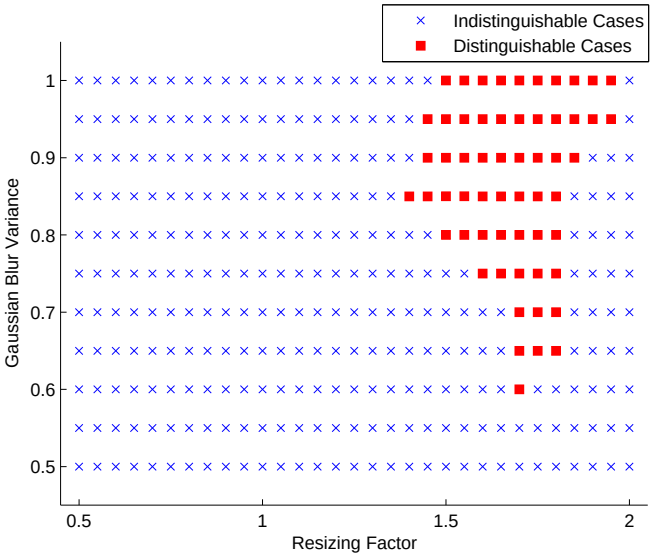
In this experiment, we use all 1338 unaltered images in UCID database to generate the test database. We use Gaussian blur with filter window 5 by 5 and variance ν to simulate the blurring operation. For each $s = \{0.5, 0.55, ..., 2\}$ and $\nu = \{0.5, 0.55, ..., 1\}$, the test database contains: 1338 unaltered images, 1338 resized images with scaling factor s, 1338 blurred images with Gaussian variance ν, 1338 blurred then resized images, and 1338 resized then blurred images. The reasonable range of $\nu \leq 1$ is obtained by calculating the distortion introduced by blurring using the structure similarity (SSIM) index [99] and setting the reasonable SSIM measure as greater that 0.9.

Based on the estimator proposed in section 4.3 with tunable parameters $\underline{\theta} = (\tau_1, \tau_2)$ (4.28), we use our information theoretical framework and criteria to obtain the distinguishable and indistinguishable cases for different pairs of s and ν. Fig. 4.10(a) and 4.10(b) shows results for the case of uniform priors and the case where priors are unknown, respectively. Due to the fewer constraints on hypothesis priors, Fig. 4.10(b) contains more distinguishable cases than Fig. 4.10(a) does.

To understand why we cannot detect the order of resizing and blurring in those indistinguishable cases, we examine the transition scenarios where distinguishable cases becomes indistinguishable. That is, we analyze the indistinguishable cases close to the range of distinguishable cases in Fig. 4.10. By definition 5, we have found that for most cases, the confusing hypotheses that makes the order indetectable are H_3 and H_4 in (4.1). Thus, by definition 7, the reason that we cannot detect the order of resizing and blurring in these cases is that the fingerprints of blurring then



(a)



(b)

Figure 4.10: Distinguishability test results of detecting the order of resizing and blurring by applying our information theoretical framework and criteria. (a) Priors are known and uniform. (b) Priors are unknown.

resizing and resizing then blurring are the same. This matches the example we have shown in Fig. 4.2 where the fingerprints in Fig. 4.2(b) and Fig. 4.2(c) are similar.

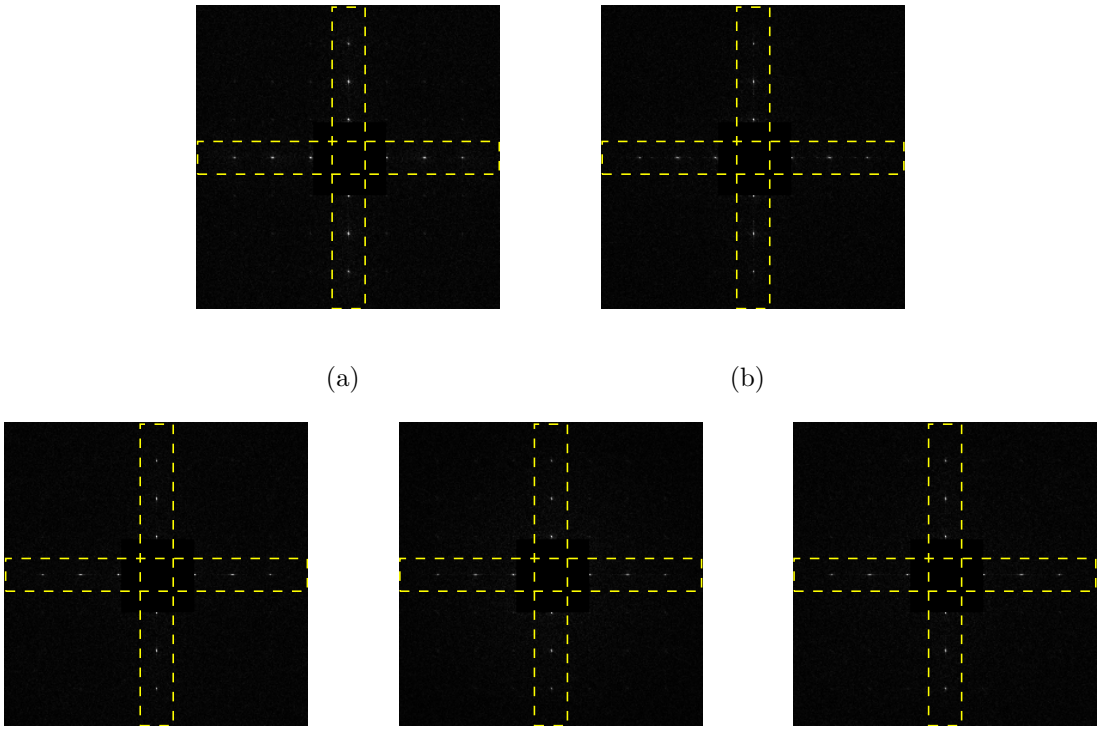
In addition, we consider a scenario where manipulations are applied on a compressed image. Then, more than two operations are involved in the analysis. In this scenario, investigators obtain a JPEG image, and want to distinguish the following hypotheses:

$$\begin{aligned} &H_0 : \text{It is single compressed,} \\ &H_1 : \text{It is double compressed interleaved by resizing,} \\ &H_2 : \text{It is double compressed interleaved by blurring,} \\ &H_3 : \text{It is double compressed interleaved by} \\ &\quad \text{blurring then resizing,} \\ &H_4 : \text{It is double compressed interleaved by} \\ &\quad \text{resizing then blurring.} \end{aligned} \tag{4.35}$$

Fig. 4.11 shows the DFT of the p-map for each of the hypotheses. Since the blocking artifact also results in peaks in the DFT of the p-map, both fingerprints of resizing and blurring are weakened by the last applied JPEG compression. Thus, these five hypotheses are easily confused with each other and may not be distinguishable based on p-map related features.

4.5 Summary

In this chapter, we proposed an information theoretical framework and mutual information based criteria to answer the question of when we can and cannot detect



(a) (b)

(c) (d) (e)

Figure 4.11: The DFT of the p-map of (a) a single JPEG compressed image with compression quality factor 75, and (b)-(e) double JPEG compressed images with compression quality factors 75 then 85 and interleaved by (b) reszing, (c) blurring, (d) blurring then resizing, and (e) resizing then blurring. The same image in Fig. 4.1(a) is examined in this example. Resizing factor is 1.5 and the variance of Gaussian blur is 1. Regions of interests are highlighted by dotted rectangles.

the order of operations. Specifically, we first formulated the order detection problems into multiple hypotheses estimation problems. Then, based on a certain set of estimators, mutual information based criteria were proposed to determine whether we can distinguish all considered hypotheses. To demonstrate the effectiveness of our proposed framework and criteria, we first apply them on two existing and detectable problems: double JPEG compression detection and the order detection of resizing and contrast enhancement. Simulations show that the results obtained by our framework match with those from existing literatures. Then, the case study of detecting the order of resizing and blurring is examined, where the order may not always be detectable. In this case study, we proposed an estimation technique to detect their order. Based on this estimator, we used our information theoretical framework and criteria to obtain the detectable cases and find the reasons for indetectable cases.

Chapter 5

Fundamental Tradeoffs in Compression Anti-forensics

Due to the wide availability of multimedia editing tools, the authenticity of multimedia content is often called into question. In order to verify the authenticity of this content, scientists have developed many forensic techniques to trace the processing histories of suspicious multimedia signals [24, 31–33, 50, 58, 59, 77, 86, 88, 94]. Among these techniques, tracing an images compression history has particular forensic significance. This is because detecting previous applications of JPEG compression in images that are currently stored in uncompressed formats can help the investigator to identify their origins [31,33]. Furthermore, double or multiple compression may occur when a compressed image is manipulated, then re-saved in the same format. As a consequence, detecting double compression or multiple compression can imply that editing has possibly been applied to the image, thus calling its authenticity into question. There are many forensic tools to detect double and multiple compressions [5, 7, 19, 35, 44, 55, 64, 66, 73, 74, 76].

Given the forensic significance of an image's compression history, anti-forensic techniques have been developed in order to confuse forensic detectors [4,26,30,57,81, 87]. These techniques enable a forger to fool forensic investigators through multiple ways. First, the forger can remove compression fingerprints completely so that the origin of the image cannot be detected. Furthermore, he/she can then recompress

the anti-forensically modified image using another quantization table to mislead the identification of its origin [87]. When double compression occurs while editing a compressed image, modifying the compression history can also reduce the possibility of the forgery being detected via compression fingerprints. Additionally, other anti-forensic techniques have been developed to create forensically undetectable forgeries [15, 49, 85, 101].

Studying anti-forensics and analyzing forgers' behavior are equally important for forensic purpose. Forensic investigators can use this information to improve existing detectors [40]. Furthermore, based on the specific fingerprints left by applying anti-forensics, investigators can develop new forensic detectors to reveal the use of anti-forensics [52, 56, 98]. Through either way, forensic invesitgators can make their detection system more robust by analyzing possible anti-forensic techniques.

Often, when anti-forensic techniques are applied, they introduce distortion to the multimedia content while concealing the fingerprints of manipulation [4, 26, 30, 87]. For example, the authors in [87] remove JPEG compression fingerprints by adding anti-forensic dither to each DCT coefficient to eliminate quantization fingerprints. Thus, as the fingerprints are removed, distortion is also introduced to the DCT coefficients through the dither. In [4, 26, 30], the fingerprints are concealed by optimizing a certain cost function under some constraints. While achieving the anti-forensic performance, the distortion is also introduced to the content, the amount of which depends on the constraints. In these cases, the forger must balance between the amount that fingerprints have been concealed and the distortion introduced by anti-forensic modification.

Similarly, anti-forensics may also increase the size of the multimedia content while concealing the fingerprints of manipulation. For example, in order to conceal the fingerprints of video frame deletion/addition, the authors in [85] increase the P-frame prediction error to eliminate the periodic characteristic of the fingerprints. As a consequence, this technique enlarges the file size of the anti-forensically modified video. In such a case, the forger needs to balance between the degree to which fingerprints are concealed and the data rate.

While anti-forensic techniques may introduce the two kinds of tradeoffs discussed above, there is no existing work formally studying either of these tradeoffs. In fact, when compressing an anti-forensically modified forgery, there is a tradeoff among how much manipulation fingerprints can be concealed, the data rate, and distortion introduced into the signal. The forger must balance all three factors to appropriately decide the strength of his/her operation.

In this chapter, we characterize the tradeoff discussed above. In order to measure the amount that manipulation fingerprints can be concealed, we define the effectiveness of concealing these fingerprints as *concealability*. To demonstrate this tradeoff in a real anti-forensic system, we introduce the concealability-rate-distortion (C-R-D) tradeoff in image double JPEG compression anti-forensics. In order to adjust concealability, we propose a flexible anti-forensic dither. To reduce the time and computational complexity associated with decoding a JPEG compressed image, applying anti-forensics, then recompressing it, we introduce an anti-forensic transcoder capable of efficiently performing these tasks in one step. Through a series of experiments, we have experimentally characterized the C-R-D tradeoff in

JPEG anti-forensic systems. We have found that this tradeoff results in two distinct C-R-D surfaces; one for if the forger uses a lower JPEG quality factor during the second compression and another for if the forger uses a higher quality factor during the second compression. Furthermore, we observe two surprising phenomena from these experiments.

It is worth pointing out the implication of introducing the rate-distortion tradeoff in the field of multimedia forensics and anti-forensics. The rate-distortion tradeoff has been well studied for image and video compression [70, 90]. Both empirical and theoretical results have been derived to characterize the optimal achievable rate under a certain distortion constraint. Given this tradeoff, one can choose the optimal compression method according to his/her demands.

Since compression is a necessary signal processing for storage and transmission, rate-distortion tradeoff has been involved in the analysis of many systems in different fields. For example, when implementing compression, complexity is an essential factor, and the rate-distortion-complexity tradeoff was studied [37]. When transmitting the compressed multimedia content through wireless communication systems, energy consumption needs to be considered, where power-rate-distortion tradeoff was analyzed [42]. For multimedia attackers, there are works on studying the risk-distortion tradeoff for video collusion attacks [18]. Many anti-forensic schemes also try to maximize their concealability under some distortion constraint.

However, there is no existing work that considered the rate-distortion tradeoff when the attack or manipulation was applied on compressed multimedia content, while this is usually the case when the size of the multimedia signal is big. Thus,

in this chapter, we introduce the rate-distortion tradeoff to the field of multimedia forensics and anti-forensics and characterize the C-R-D tradeoff using the double image compression anti-forensics as an example. We believe that the C-R-D tradeoff also exists for other forensic and anti-forensic systems, like the video frame deletion/addition anti-forensic system.

The rest of the chapter is organized as follows: first, we give an overview of image compression forensics and anti-forensics in Section 5.1. Then, in Section 5.2, we give the system model of double compression anti-forensics, and define the three tradeoff factors, concealability, rate and distortion. In Section 5.3, flexible anti-forensic dither is proposed for balancing the tradeoff between concealability, rate, and distortion. Section 5.4 introduces our anti-forensic transcoder, which combines decompression, flexible anti-forensic dither, and recompression into one process. Experimental results on the C-R-D tradeoff are shown and discussed in Section 5.5. Lastly, Section 5.6 summarizes this chapter.

5.1 Background

While our proposed C-R-D tradeoff exists in general image compression anti-forensic systems, we choose one of the most commonly used compression standards, JPEG, to characterize the tradeoff and show the effectiveness of our model. This section reviews the important concepts and techniques of JPEG compression forensics and anti-forensics which will be used in this case. Specifically, we start with a brief introduction of JPEG compression. Then, as an important set of fingerprints

in forensics, double JPEG compression fingerprints are discussed. Among those double JPEG compression forensic detectors, without loss of generality, we choose one of the most popular and effective techniques to review in the next subsection. At last, we review the compression anti-forensic technique, which will be a special case in our proposed flexible anti-forensic scheme.

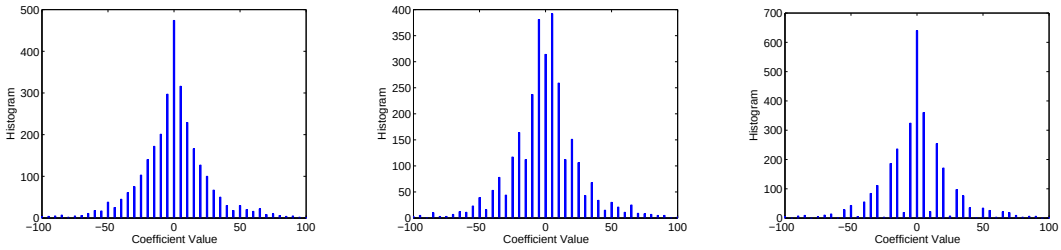
5.1.1 JPEG Compression

JPEG format is one of the most commonly used formats for images. We briefly overview the JPEG compression procedure as follows [72]: first, the image is separated into 8 by 8 blocks. Within each block, discrete cosine transform (DCT) is applied on the pixel values to obtain the DCT coefficients $x_{ij}, i, j = 0, 1, ..., 7$, where x_{ij} is the coefficient in subband (i, j). Then, quantization is applied on each DCT coefficient using a quantization table $\mathbf{Q}$, with each element denoted as q_{ij}. The quantized coefficients are

$$a_{ij} = \text{round}\left(\frac{x_{ij}}{q_{ij}}\right), \text{ for } i, j = 0, 1, \ldots, 7. \tag{5.1}$$

Finally, lossless entropy coding is applied on the quantized DCT coefficients to obtain the data ready for transmission or storage.

Decompression has the reverse procedure of compression. Yet, it cannot recover the original image due to the lossy quantization process of JPEG compression. Specifically, during dequantization, the quantized DCT coefficients a_{ij} will be multiplied by its quantization steps q_{ij} to obtain the dequantized coefficients $y_{ij} = a_{ij}q_{ij}$, which is different from x_{ij}. These dequantized coefficients will instead only have



(a) (b) (c)

Figure 5.1: Histograms of DCT coefficients subtracted from sub-band (0,2) of a natural image been (a) single compressed with specific quantization step 5, (b) doubly compressed with quantization step 3 followed by 5, and (c) doubly compressed with quantization step 7 followed by 5.

values of integer multiples of the quantization step. We use the commonly applied model, Laplace distribution, to model the DCT coefficients in a certain subband of an uncompressed image [54]. Then, the histogram of the DCT coefficients from a JPEG compressed image can be modeled as a quantized Laplace distribution. Fig. 5.1(a) shows an example of the DCT coefficient histogram of a single JPEG compressed image.

5.1.2 Double JPEG Compression Fingerprints

If a forger modifies a JPEG image, it may be saved as JPEG again after modification. In such a case, the image has undergone two instances of JPEG compressions. If the quantization tables used in these two JPEG compressions are not exactly the same, double JPEG compression fingerprints will be left in the image. Since double JPEG compression happens in most forgeries, detecting its

fingerprints is important in forensics to identify the existence of possible forgeries ever been applied on the image.

To see the double JPEG compression fingerprints, we examine the DCT coefficients of a double JPEG compressed image. Let $\mathbf{Q}^{(1)}$ and $\mathbf{Q}^{(2)}$ denote the quantization tables used in the first and second JPEG compressions, respectively. Then the quantized DCT coefficients of this double JPEG compressed image is

$$b_{ij} = \text{round}\left(\frac{y_{ij}}{q_{ij}^{(2)}}\right) = \text{round}\left(\text{round}\left(\frac{x_{ij}}{q_{ij}^{(1)}}\right)\frac{q_{ij}^{(1)}}{q_{ij}^{(2)}}\right), \tag{5.2}$$

where $q_{ij}^{(1)}$ and $q_{ij}^{(2)}$ are elements of $\mathbf{Q}^{(1)}$ and $\mathbf{Q}^{(2)}$, respectively. If we decompress the image, the DCT coefficients observed are $w_{ij} = b_{ij}q_{ij}^{(2)}$.

Although we still observe quantized DCT coefficients with step size $q_{ij}^{(2)}$ from double JPEG compressed images, these coefficients cannot be modeled as quantized Laplace. During the second quantization, uneven numbers of bins of the single quantized histogram are collected into the new bins. Thus, the magnitudes of the double quantized bins will present periodic peaks or zeros [74, 76]. These periodic characteristics of the DCT coefficient histogram are identified as the fingerprints of double JPEG compression.

For illustration, let us take a DCT subband where the quantization steps in two compressions are different. Let q_1 and q_2 denote the quantization steps in this subband during the first and second JPEG compressions, respectively. Fig. 5.1(b) and 5.1(c) show the double JPEG compression fingerprints for $q_1 < q_2$ and $q_1 > q_2$, respectively.

5.1.3 Double JPEG Compression Detection

Due to the forensic significance of double JPEG compression fingerprints, there are many forensic techniques to detect such trace [19, 35, 44, 55, 64, 66, 73, 74, 76]. Various features are used to identify the double compression fingerprints, such as the DCT histograms and their Fourier transforms [35,64,73,74,76], the histograms of the first digit of DCT coefficients [66], and the number of DCT coefficients changed when recompressing with the same quantization table [44]. Among them, we choose one of the most popular and best performing detectors in [74] to review and use in this work.

In [74], Pevný and Fridrich modeled the double JPEG compression detection problem as a classification of images between two classes:

$$C_1 : \text{The image is single compressed.} \tag{5.3}$$

$$C_2 : \text{The image is double compressed.} \tag{5.4}$$

Given the distinctive fingerprints of double JPEG compression in DCT coefficient histograms, they took the magnitudes of quantized bins in the histogram as the feature and fed them to a support vector machine.

Specifically, they chose the low frequency subbands where double JPEG compression fingerprints are most obvious. For each subband, the numbers of occurrences at integer multiples of q_2 were counted, where q_2 is the quantization step in the second compression. The feature vector was composed by concatenating the

data from all chosen subbands:

$$\underline{v} = \left\{ \frac{1}{c_{ij}} \big(h_{ij}(0), h_{ij}(1), \ldots, h_{ij}(15) \big) \Big| (i,j) \in \mathcal{L} \right\}, \tag{5.5}$$

where $h_{ij}(m)$ denotes the number of occurrences at $\pm m q_2$ in subband (i,j), and c_{ij} is a normalization constant, i.e., $c_{ij} = \sum_{m=0}^{15} h_{ij}(m)$. The set of low frequency subbands was chosen as

$$\mathcal{L} = \big\{ (1,0), (2,0), (3,0), (0,1), (1,1), (2,1), (0,2), (1,2), (0,3) \big\}. \tag{5.6}$$

Given the feature vector $\underline{v}$ described above, the classification was done by using a soft-margin support vector machine with the Gaussian kernel [10] $k(x,y) = exp(-\gamma||x-y||^2)$. $k(x,y)$, also known as radial basis function, is a popular kernel function used in SVM classification. It can be interpreted as a similarity measure between two feature vector samples x and y. γ is a free parameter, which defaultly equals to 1/num_features in LIBSVM open source machine learning library.

5.1.4 JPEG Compression Anti-Forensics

There are also anti-forensic techniques that can falsify the image compression history and confuse the forensic detectors [26, 30, 57, 87]. Among them, we choose one of the most popular techniques in [87], which can successfully attack the forensic detector in [74], for illustration in this work. Yet, the applicability of other anti-forensic techniques will also be discussed. In [87], single quantized DCT coefficients were added pre-designed dither so that the histogram will be smooth and look like the one from an uncompressed image. Then, when the forger modifies a JPEG

image, as long as the traces of the first compression are removed, the recompressed image will only present single compression fingerprints. In this way, the forger can escape the forensic detection of double JPEG compression.

We briefly review the anti-forensic scheme proposed in [87] as follows: let random variable X denote the DCT coefficient of a certain sub-band (i, j) from an uncompressed image. $f(x, \lambda)$ is the modeled Laplace distribution of X with parameter λ, i.e.,

$$\mathbb{P}(X = x) = f(x, \lambda) = \frac{\lambda}{2} e^{-\lambda|x|}. \tag{5.7}$$

After JPEG compression, let Y denote the DCT coefficient of a JPEG compressed image and its distribution will be a quantized Laplace:

$$\mathbb{P}(Y = kq) = \begin{cases} 1 - e^{-\lambda q/2} & \text{if } k = 0, \\ e^{-\lambda|kq|} \sinh(\frac{\lambda q}{2}) & \text{otherwise,} \end{cases} \tag{5.8}$$

where q is the quantization step and $k \in \mathbb{Z}$. Then, in order to remove the fingerprints of JPEG compression, an anti-forensic dither, denoted as D, is added on the DCT coefficients of the JPEG compressed image. The resulting anti-forensically modified coefficients are $Z = Y + D$. Given a carefully designed anti-forensic dither, the distribution of Z can be equal to that of X. The distribution of the anti-forensic dither D in [87] is given by

$$\mathbb{P}(D = d|Y = kq) = \frac{f(kq + d, \hat{\lambda})}{\int_{(k-\frac{1}{2})q}^{(k+\frac{1}{2})q} f(x, \hat{\lambda}) dx} \mathbb{1}(-\tfrac{q}{2} \leq d < \tfrac{q}{2}), \tag{5.9}$$

where $\hat{\lambda}$ is the estimated parameter using coefficients Y and $\mathbb{1}(\cdot)$ is an indicator function.

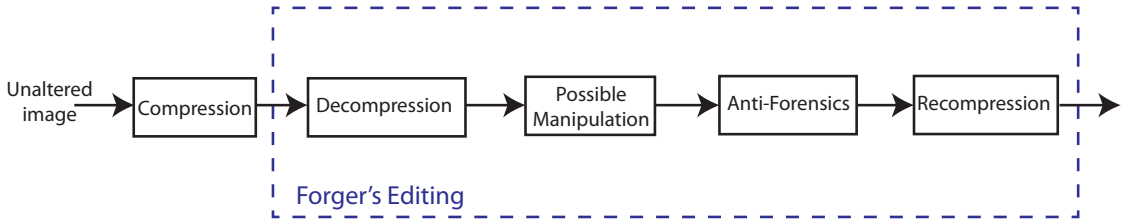


Figure 5.2: The system model considered in this chapter.

5.2 Concealability-Rate-Distortion Tradeoff

In this chapter, we assume that the forger wishes to recompress an image that has previously been JPEG compressed. This may happen under a variety of scenarios. For example, a forger may wish to falsify the content of the image. In this case, the forger must decompress the image, perform some manipulation, then recompress the image. Alternatively, if the forger does not wish to alter the content of the image but just wishes to falsify its origin, they must recompress the image using the quantization matrix used by the target camera [87]. In both scenarios, standard recompression will cause double JPEG fingerprints to occur.

To analyze both of these scenarios, we adopt the following system shown in Fig. 5.2. First, the forger receives a JPEG compressed image, which we refer to as the unaltered image. The forger will then decompress the image and perform any desired image manipulation. After this, they will apply anti-forensics to remove JPEG compression fingerprints, then recompress the image using their desired compression parameters. During this process, the forger is able to adjust the strength with which they apply anti-forensics, as well as the quality factor or quantization tables used during compression. Because we are interested primarily in characterizing

the tradeoff among rate, distortion and the amount of double JPEG compression fingerprints that can be concealed, we neglect any effects caused by other possible manipulations for the purposes of this work.

Intuitively, when a forger applies anti-forensic techniques, he/she must balance a tradeoff between the amount of double JPEG compression fingerprints that can be concealed and the distortion introduced by anti-forensic modification. The forger can vary the anti-forensic strength to adjust the amount of modification caused by the anti-forensic technique, and thus balance this tradeoff. When recompressing the forgery, there is a well-known tradeoff between the data rate and the distortion. In addition, since anti-forensics modifies the distribution of the DCT coefficients, it is possible that it can also affect the data rate during recompression. On the other hand, the performance of double JPEG compression detection depends on the relationship between the primary and the secondary quality factor. Thus, the secondary quality factor may also affect the possibility that the double JPEG compression will be detected. In other words, the amount of double JPEG compression fingerprints that can be concealed is also affected by the secondary quality factor.

Therefore, the amount of double JPEG compression fingerprints that can be concealed, the data rate, and the distortion are all related in the system. Adjusting either the strength of anti-forensics or the quality factor in recompression process will result in change of all three factors. Therefore, in order to achieve a certain requirement, the forger must balance the tradeoff among these three factors.

We note that, given the existence of many compression anti-forensic detectors, i.e., counter anti-forensic schemes, [52, 56, 98], our system model can be extended

to include their effect in the following ways: 1) generalize the definition of concealability by including the amount of anti-forensic fingerprints that can be concealed 2) introduce another dimension in the tradeoff to reflect the detectability of anti-forensic techniques.

In order to characterize the tradeoff between how much the double JPEG compression fingerprints can be concealed, the data rate, and the distortion, we first define the term *concealability* as the measure of how much the fingerprints can be concealed. Since the accuracy of a detector is one measure of how well the fingerprints have been concealed, we define concealability in terms of the detection rate.

When detecting manipulation fingerprints, a simple hypothesis test is often used, where two hypotheses are defined as

$$H_0 : \text{Manipulation fingerprints do not present.}$$

$$H_1 : \text{Manipulation fingerprints do present.}$$

A forensic investigator will apply a certain decision rule to a suspicious signal to determine which hypothesis it belongs to. The decision rule results in a probability that the fingerprints are correctly detected, which is called the detection rate; and a probability that an unmanipulated signal is identified as a falsified one, which is called the false alarm rate. Different decision rules often results in different pairs of detection rates and false alarm rates. A receiver operating characteristic (ROC) curve plotting all reachable pairs of detection rates and false alarm rates characterizes the overall performance of the detector.

We define concealability as follows: let I denote the image edited by the forger. Let function $m(\cdot)$ be the modification made by the forger. Then, $m(I)$ is the forger modified image. In the system describe in Fig. 5.2, I represents the single compressed JPEG image and $m(I)$ represents the double JPEG compressed and anti-forensically modified image. For a given detector and a certain false alarm rate P_f, there is a corresponding decision rule $\delta_{P_f}(\cdot)$. Then the concealability of the forger edited image $m(I)$ is defined as

$$C(m, P_f) = \min\left(\frac{1 - \mathbb{P}\Big(\delta_{P_f}\big(m(I)\big) = H_1\Big)}{1 - P_f}, 1\right). \tag{5.10}$$

We explain the definition of concealability by using ROC curves, as it is shown in Fig. 5.3. When no anti-forensics has been applied, the best performance of a forensic detector is perfect detection. That is, the detector can achieve detection rate of 100% at false alarm rate of 0%. Under this scenario, manipulation fingerprints can be detected without any error, and we say that the fingerprints have been fully exposed to the investigators. Thus, the concealability in this case will be its minimum value 0.

On the other hand, if anti-forensics are applied, it will reduce the accuracy of the forensic detector and increase the false alarm rate. Such degradation reaches its maximum when the detection rate becomes the same as the false alarm rate. In this case, the detector will act as an equal probability random decision process, i.e., the decision is made equivalently to randomly flipping a coin. Under this scenario, forger edited images will have no difference with those that have no been edited by the forger. Thus, we say that manipulation fingerprints have been fully concealed to

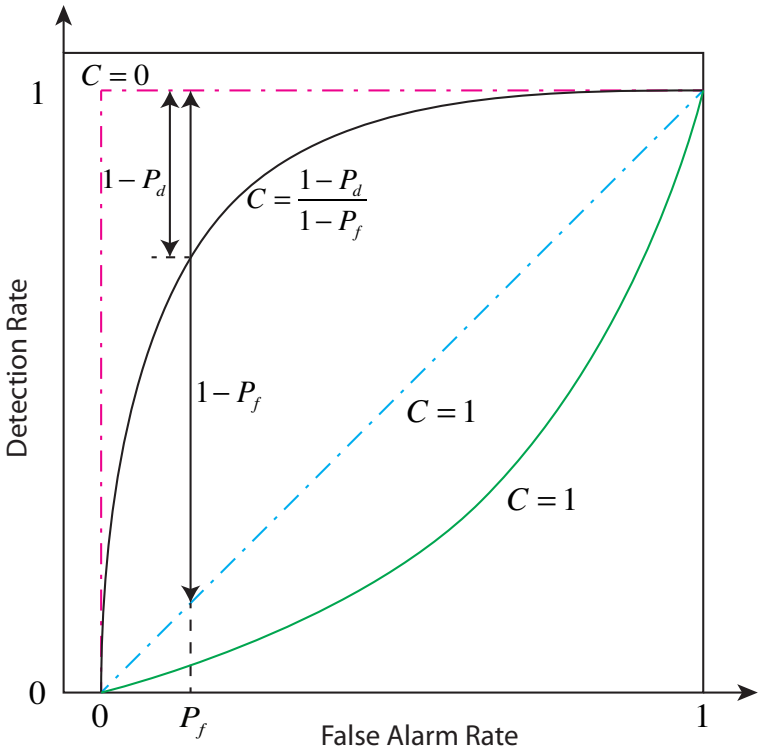


Figure 5.3: Examples of concealabilities related to ROC curves. When the detector achieves perfect detection, the forger has concealability of the fingerprints as 0. When the ROC curve is at or below the random decision line, we say that the forger has achieved concealability as 1. Then for those ROC curves between perfect detection and random decision, the concealability ranges from 0 to 1 and depends on a certain false alarm rate.

the forensic investigators. We define the concealability in this case as its maximum value 1. We note that since the forensic detection strategy is determined regardless of the possible existence of anti-forensic technique, one may obtain a ROC curve below the random decision line, where detection rates equal to false alarm rates. However, because this scenario also implies that the forger has fully concealed the fingerprints, we define the concealability in such case also as 1.

For scenarios between these extreme cases, the concealability is defined as a measure dependent on the false alarm rate. Since it is inversely proportional to the detection rate and the value is limited between 0 and 1, we use a normalized decreasing function of the detection rate $\frac{1-P_d}{1-P_f}$ to characterize the concealability at a certain false alarm rate.

To evaluate the distortion, we define a measure that is based on the mean structural similarity (MSSIM) between the image that has not been edited by the forger and the one after the forger's editing [99]. MSSIM is a popular similarity measure between 0 and 1 that matches well with human perception. In order to let the distortion equal to zero when the two images are identical, i.e., when the similarity is 0, we define the distortion between I and $m(I)$ as

$$D(m) = 1 - \mathrm{MSSIM}\big(I, m(I)\big). \tag{5.11}$$

We note that similar results can be obtained for other measures of distortion such as mean square error (MSE), which will be shown in simulation results.

Lastly, we use bits per pixel as the measure of rate. Specifically, the rate is calculated by examining the size of the forger edited image and dividing it by the

number of pixels in that image:

$$R(m) = \frac{\text{number of bits of } m(I)}{\text{number of pixels in } m(I)}. \tag{5.12}$$

5.3 Flexible Anti-Forensic Dither

In order to balance the tradeoff of concealability and distortion during the anti-forensic process, the forger needs to vary the strength of anti-forensics. Though there exists anti-forensic techniques to fully conceal the fingerprints of double JPEG compression [4, 26, 30, 87], these techniques do not provide the flexibility to control the strength of anti-forensics. However, in order to characterize the C-R-D tradeoff and find the best choice, flexible anti-forensic schemes are necessary. In this section, we propose a *flexible anti-forensic dither* for the technique in [87] that enables the forger to adjust the strength of anti-forensics. Similar concept can be applied on other anti-forensic techniques, which we will discuss in the end of this section.

As we discussed in section 5.1, double JPEG compression fingerprints are presented in DCT coefficients. Thus, in order to remove the fingerprints, our flexible anti-forensic dither will also be applied on DCT coefficients. To develop flexible dither, let us examine the procedure that a DCT coefficient in a certain subband of an image will go through during the whole process described in Fig. 5.2. First of all, the unaltered image will go through its first JPEG compression. Let q_1 denote the quantization step of the examined subband used in this compression. Then, the DCT coefficient of the single compressed image is obtained by

$$Y = q_1 \,\text{round}(X/q_1). \tag{5.13}$$

We assume that X obeys a Laplace distribution (5.7). Thus, Y will be distributed as a quantized Laplace distribution with quantization step size q_1.

Secondly, the flexible anti-forensic dither is applied on Y. Let α denote the *anti-forensic strength*. We define that $0 \leq \alpha \leq 1$. The corresponding flexible anti-forensic dither is denoted as D_α. Thus, the anti-forensically modified DCT coefficient becomes to

$$Z_\alpha = Y + D_\alpha. \tag{5.14}$$

Lastly, after recompressing Z_α with a quantization step q_2, the double JPEG compressed and anti-forensically modified DCT coefficient is

$$W_\alpha = q_2 \,\mathrm{round}(Z_\alpha / q_2). \tag{5.15}$$

If no anti-forensics has been applied, which means that the anti-forensic strength is 0, then $W_0 = q_2 \,\mathrm{round}(Y/q_2)$. The histogram of W_0 will present the fingerprints of double JPEG compression, as it is shown in Fig. 5.1(b) or Fig. 5.1(c). The periodic peaks or zeros in the histogram distinguish W_0 from those of single compressed images, who have quantized Laplace distribution shape as shown in Fig. 5.1(a). Thus, by measuring the distance between the normalized histogram of W_0 and the quantized Laplace distribution, forensic analysts can detect double JPEG compression.

If anti-forensics are fully applied, as it is the case in [87], the anti-forensic strength is 1, and the distribution of D_1 is the same as (5.9) with q substituted with q_1. Then, the distribution of Y will be the same as that of X. Consequently, the distribution of W_1 will be a quantized Laplace distribution. In such a case, the double

JPEG compressed and anti-forensically modified image is hard to be distinguished from single JPEG compressed images through DCT coefficient histograms.

When anti-forensic strength is not applied in full, we can reduce the anti-forensic distortion by sacrificing the exposure of fingerprints to the forensic detector. That is, the histogram of W_α will be less like a quantized Laplace distribution when less anti-forensic strength is applied. By examining (5.9), we can see that the distribution of the dither D_1 has a bounded support $[-q_1/2, q_1/2)$. The shape of this distribution is a normalized and shifted version of the target distribution $f(x, \hat{\lambda})$ on support $\big[(k-1/2)q_1, (k+1/2)q_1\big)$ with left shifting of kq_1. Such design is to make the conditional probability $\mathbb{P}(Z_1 = z|Y = kq_1)$ be the same as $f(z, \hat{\lambda})$ normalized by $\mathbb{P}(Y = kq_1)$ with $z \in [kq_1 - q_1/2, kq_1 + q_1/2)$. Then, with Y taken all integer multiplies of q_1, the distribution of Z_1 will be the same as $f(z, \hat{\lambda})$.

When $\alpha < 1$, we shrink the support of the anti-forensic dither to decrease distortion. Meanwhile, the similarity between the distribution of Z_α and $f(z, \hat{\lambda})$ will be reduced. We note that because of the shrink of the dither's support, the anti-forensically dithered coefficients will not spread out the entire quantization interval. Consequently, the support of the histogram of the anti-forensically modified image before recompression will not match the support of the histogram of an uncompressed image. Nevertheless, the image will be recompressed, where all coefficients are requantized to integer multiples of the new quantization step. The use of anti-forensic dither can cause some coefficients that would normally get quantized to lq_2 to instead be mapped to $(l-1)q_2$ or $(l+1)q_2$. In this way, the strength of the double compression fingerprints are weakened by the anti-forensic dither.

Let $S_\alpha^{(k)}$ denote the support of Z_α given $Y = kq_1$, which means that the support of D_α is $S_\alpha^{(k)}$ left shifted by kq_1. Then the range of $S_\alpha^{(k)}$ will be decreased when less anti-forensic strength is applied, i.e., α decreases. We will give the explicit expression of $S_\alpha^{(k)}$ in later paragraphs. We still take the shape of the dither's distribution to be a normalized and shifted version of $f(x, \hat{\lambda})$. The distribution of the flexible anti-forensic dither is proposed as

$$\mathbb{P}(D_\alpha = d|Y = kq_1) = \frac{f(kq_1 + d, \hat{\lambda})}{\int_{S_\alpha^{(k)}} f(x, \hat{\lambda})dx} \mathbb{1}\big(kq_1 + d \in S_\alpha^{(k)}\big). \tag{5.16}$$

We define $S_1^{(k)}$ as

$$S_1^{(k)} = \big\{t \in \mathbb{R}\big|(k - \tfrac{1}{2})q_1 \leq t < (k + \tfrac{1}{2})q_1\big\}, \tag{5.17}$$

then (5.9) becomes a special case of (5.16). By our definition, $S_0^{(k)}$ is the support of Z_0 given $Y = kq_1$, which results in W_0. However, due to the second compression described by (5.15), there are multiple choices of $S_0^{(k)}$ which can lead to the same W_0 after requantization. Specifically, let lq_2 be the quantized bin that $Y = kq_1$ will be mapped into during the second compression, i.e.,

$$l = \text{round}\left(\frac{kq_1}{q_2}\right). \tag{5.18}$$

Then, any dither within the range $\big[(l - 1/2)q_2, (l + 1/2)q_2\big)$ will be mapped into the same bin lq_2. We define $S_0^{(k)}$ as the one that has the largest range while any dither within this support will be mapped into the same $W_0 = lq_2$. In addition, the property of $S_\alpha^{(k)}$ needs to be satisfied, i.e., $S_0^{(k)} \subseteq S_1^{(k)}$. Thus, the expression of $S_0^{(k)}$ is given as

$$S_0^{(k)} = \big\{t \in S_1^{(k)}\big|(l - \tfrac{1}{2})q_2 \leq t < (l + \tfrac{1}{2})q_2\big\}. \tag{5.19}$$

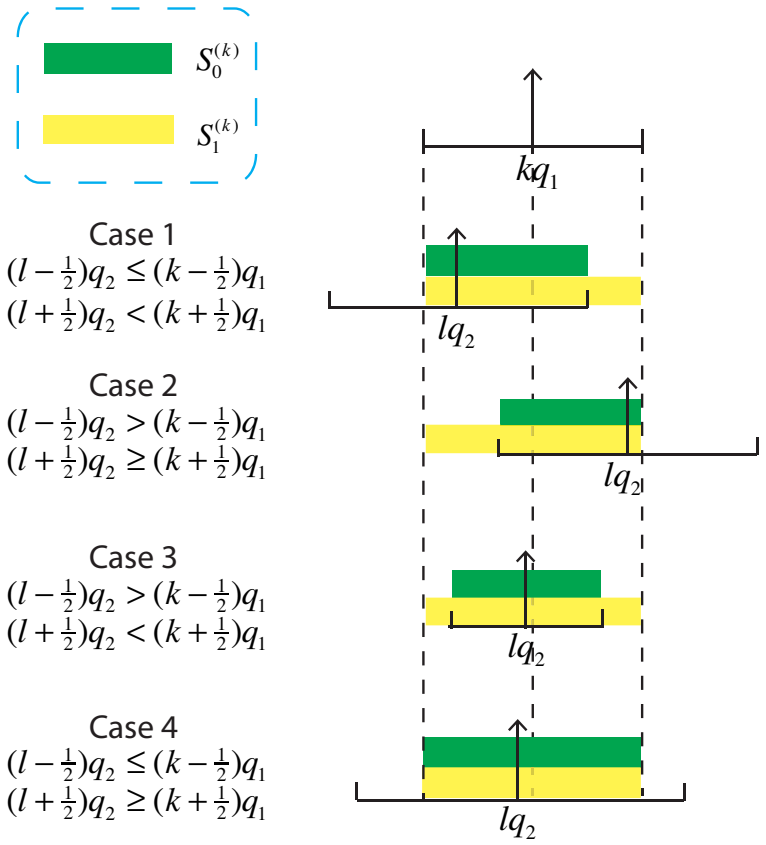


Figure 5.4: An illustration of how to determine $S_0^{(k)}$ and $S_1^{(k)}$ for a certain value of $Y = kq_1$. The vertical arrows denote the position of a certain quantized bin in the coefficient histogram. The horizontal line segment at the bottom of each arrow represents the quantization interval where all values within this range will be mapped into the quantized bin indicated by the arrow. lq_2 is the quantized bin that kq_1 will be mapped into during the recompression. According to different positions of lq_2 and its quantization intervals, there are four cases for $S_0^{(k)}$, while $S_1^{(k)}$ keeps the same for the same kq_1.

Fig. 5.4 shows an illustration of how to find $S_0^{(k)}$ and $S_1^{(k)}$ for a certain quantized bin $Y = kq_1$. Four cases are listed in the figure regarding the relative positions of the quantization intervals of Y in the first compression and $W_0 = lq_2$ in the second compression. Basically, $S_0^{(k)}$ is the intersection between the intervals $\big[(k-\frac{1}{2})q_1, (k+\frac{1}{2})q_1\big)$ and $\big[(l-\frac{1}{2})q_2, (l+\frac{1}{2})q_2\big)$.

Given the extreme cases of $S_\alpha^{(k)}$ when $\alpha = 0$ and $\alpha = 1$, we pick up $S_\alpha^{(k)}, 0 < \alpha < 1$, from the convex hull of the supports of $S_0^{(k)}$ and $S_1^{(k)}$. Formally, let $b_{\alpha,1}$ and $b_{\alpha,2}$ be the lower and upper bounds of support set $S_\alpha^{(k)}$, respectively. We have the extreme cases

$$\begin{aligned} b_{0,1} &= \max\big((k-\tfrac{1}{2})q_1, (l-\tfrac{1}{2})q_2\big), \;\; b_{1,1} = (k-\tfrac{1}{2})q_1, \\ b_{0,2} &= \min\big((k+\tfrac{1}{2})q_1, (l+\tfrac{1}{2})q_2\big), \;\; b_{1,2} = (k+\tfrac{1}{2})q_1. \end{aligned} \tag{5.20}$$

Then, $S_\alpha^{(k)}, 0 < \alpha < 1$ is defined as

$$S_\alpha^{(k)} = \big\{t \in \mathbb{R} \big| b_{\alpha,1} \le t < b_{\alpha,2}\big\},$$

where

$$b_{\alpha,j} = (1-\alpha)b_{0,j} + \alpha b_{1,j}, \quad \text{for } j = 1, 2. \tag{5.21}$$

Using (5.21) and (5.16), our flexible anti-forensic dither can be generated from this pre-determined distribution.

The flexible anti-forensic scheme can be summarized as follows:

1. Obtain DCT coefficients by decompressing the single compressed image for all subbands.

2. In each subband, estimate the parameter $\hat{\lambda}$ of the Laplace distribution function $f(x, \hat{\lambda})$ using Y statistics [87].

3. For a certain anti-forensic strength α, calculate $S_\alpha^{(k)}$ and $\mathbb{P}(D_\alpha = d|Y = kq_1)$ for each kq_1 using (5.21) and (5.16).

4. For each $Y = kq_1$, randomly generate a value of D_α from the distribution function (5.16), and add it to Y to obtain Z_α.

5. Obtain the anti-forensically modified image by modifying all coefficients in all subbands and mapping them to pixel domain.

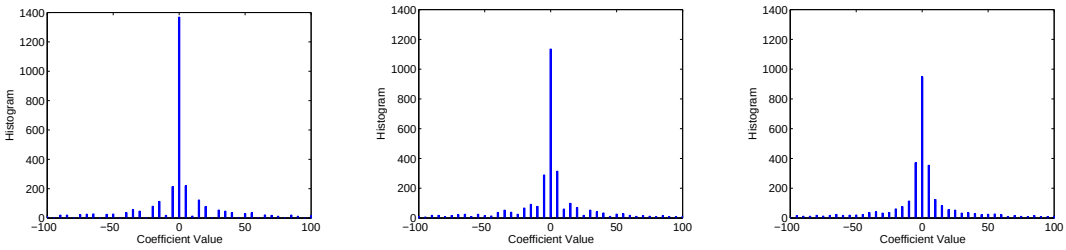
We note that the concealability-distortion tradeoff also occurs in other anti-forensic techniques, where the forger can vary the anti-forensic strength to balance them [4, 26, 30, 57]. In [30], the authors modified pixel values of an image to conceal JPEG compression fingerprints. Specifically, they minimized the total variance and variance difference between boundary areas and interior areas of blocks while limiting the modified DCT coefficients in a distortion constraint set. The smaller the minimized function is, the higher the concealability will be. Then, by shrinking the range of the constraint set, less distortion is allowed to be introduced to the image, but a larger minimized function will be obtained, and thus concealability decreases. Techniques in [4, 26] concealed manipulation fingerprints by modifying the manipulated histogram to one that is closest to an unaltered histogram under some distortion constraints. Similarly, by varying the distortion constraints, the forger is able to vary how close the anti-forensically modified histogram is to an unaltered one, and thus vary the concealability. Lastly, in [57], the fingerprints of double

JPEG compression with the same quantization table were concealed by modifying the DCT coefficients in textural regions. Then, the less the DCT coefficients were modified, the less distortion it introduces to the image. However, the fingerprints are less concealed, and thus concealability becomes smaller. In all cases, the tradeoff between concealability and distortion exists and flexible anti-forensic techniques can be applied to characterize them.

5.4 Anti-Forensic Transcoder

As a post-processing technique, anti-forensic dither can be used whenever a forgery needs to be recompressed without leaving double JPEG compression fingerprints. Yet, there are some cases, for example when modifying the quantization table of the compressed image, where the forger simply wants to recompress the JPEG image without performing other manipulations. In such cases, the forger do not need to decompress the JPEG image, apply anti-forensic dither, and then recompress the image. Instead, the forger can use an integrated anti-forensic transcoder to directly falsifies the DCT coefficients from the JPEG file and transcodes them into the coefficients associated with another quantization table while no double compression fingerprints will be detected. In this section, we propose this anti-forensic transcoder to reduce the time and computational complexity associated with decompressing a JPEG image, applying anti-forensics, then recompressing it, and efficiently perform all these tasks in one step.

To propose this anti-forensic transcoder, let us review the modifications of



(a) (b) (c)

Figure 5.5: Histograms of DCT coefficients of an anti-forensically modified and double compressed image with anti-forensic strength (a) $\alpha = 0$, (b) $\alpha = 0.4$, and (c) $\alpha = 1$.

DCT coefficients made by the anti-forensic dither and recompression. As described in Section 5.3, the decompressed DCT coefficient Y will be added with the anti-forensic dither D_α to obtain the anti-forensically modified coefficient Z_α. This modification dithers each $Y = kq_1$ to some nearby values. When we examine the coefficients' histogram, we will see that the anti-forensic dither spreads each quantized bin within a certain range. Then, Z_α will be mapped into pixel domain where recompression is applied. During recompression, Z_α is again transformed into DCT domain and then quantized. In quantization process, some of the dithered values will be mapped into one bin while some of them may be mapped into other bins. Thus, even though these dithered coefficients are all coming from the same value of $Y = kq_1$, they will be mapped into different values of $W_\alpha = jq_2, j_{\min} \leq j \leq j_{\max}$. If we figure out what portions of coefficients valued as $Y = kq_1$ will be mapped into $W_\alpha = jq_2, j_{\min} \leq j \leq j_{\max}$, we can then directly map some of the coefficients $Y = kq_1$ to one of W_α without the intermediate state of Z_α. Different anti-forensic strengths will affect these portions and also the range that kq_1 will be mapped into,

i.e., $j_{\min}$ and $j_{\max}$.

Fig. 5.5 shows the transition of the histograms of W_α when increasing the anti-forensic strength. When no anti-forensics is applied, each $Y = kq_1$ can only be mapped into one bin valued as $W_0 = lq_2$ during the second quantization. Without loss of generality, we consider the case where $q_1 > q_2$. Then, some integer multiples of q_2 may even not have corresponding coefficients. This results in those nearly zero bins in Fig. 5.5(a). With anti-forensics applied, some of the coefficients valued as kq_1 can be mapped into nearby bins other than the lq_2 bin. Thus, those nearly zero bins can be gradually filled up by its neighboring bins to finally obtain the quantized Laplace shape histogram, as it is shown in Fig. 5.5(b) and Fig. 5.5(c).

We derive the direct map between Y and W_α using the intermediate state Z_α described in Section 5.3. First, we decide the range that kq_1 can be mapped into. Recall that $S_\alpha^{(k)}$ is the support of Z_α given $Y = kq_1$. Thus, when quantizing Z_α to obtain $W_\alpha = jq_2$, all candidates of j will be bounded by

$$\begin{aligned} j_{\min} &= \text{round}\left(\frac{b_{\alpha,1}}{q_2}\right), \\ j_{\max} &= \text{round}\left(\frac{b_{\alpha,2}}{q_2}\right). \end{aligned} \tag{5.22}$$

Next, we let γ_{kj} denote the probability that the anti-forensic transcoder maps a coefficient valued as kq_1 to jq_2. Then, we can describe the mapping of the anti-forensic transcoder on DCT coefficients by using the following transition probability function,

$$\mathbb{P}(W_\alpha = jq_2 | Y = kq_1) = \begin{cases} \gamma_{kj} & \text{if } j_{\min} \le j \le j_{\max}, \\ 0 & \text{otherwise.} \end{cases} \tag{5.23}$$

The value of γ_{kj} depends on the extent that the anti-forensic dither spreads the single bin kq_1, which is also determined by the anti-forensic strength. From (5.16) and (5.14), we have

$$\mathbb{P}(Z_\alpha = z|Y = kq_1) = \frac{f(z,\hat{\lambda})}{\int_{S_\alpha^{(k)}} f(x,\hat{\lambda})dx}\mathbb{1}\big(z \in S_\alpha^{(k)}\big). \tag{5.24}$$

When quantizing Z_α, those values belonging to the range $\big[(j-\frac{1}{2})q_2, (j+\frac{1}{2})q_2\big)$ will be mapped to value $W_\alpha = jq_2$. Let R_j denote this quantization interval for $W = jq_2$, i.e.,

$$R_j = \big\{t \in \mathbb{R} | (j-1/2)q_2 \le t < (j+1/2)q_2\big\}. \tag{5.25}$$

Then, we have

$$\begin{aligned}
\gamma_{kj} &= \mathbb{P}(W_\alpha = jq_2|Y = kq_1) \\
&= \int_{R_j} \mathbb{P}(Z_\alpha = z|Y = kq_1)dz \\
&= \frac{\int_{S_\alpha^{(k)}\cap R_j} f(z,\hat{\lambda})dz}{\int_{S_\alpha^{(k)}} f(x,\hat{\lambda})dx}
\end{aligned} \tag{5.26}$$

Given $j_{\min}$, $j_{\max}$, and γ_{kj} well defined by (5.22) and (5.26), the anti-forensic transcoder can be described as follows: Let U be a uniformly distributed random variable within $[0,1)$. Then, for a coefficient valued as kq_1, the anti-forensic transcoder with anti-forensic strength α will map it to

$$W_\alpha = \sum_{j=j_{\min}}^{j_{\max}} jq_2\mathbb{1}\Big(\sum_{t=j_{\min}}^{j-1}\gamma_{kt} \le U < \sum_{t=j_{\min}}^{j}\gamma_{kt}\Big), \tag{5.27}$$

where $\sum_{t=j_{\min}}^{j-1}\gamma_{kt} = 0$ when $j = j_{\min}$.

We summarize the anti-forensic transcoder as follows:

1. Obtain DCT coefficients by directly reading the JPEG file.

2. In each subband, estimate the parameter $\hat{\lambda}$ of the Laplace distribution function $f(x, \hat{\lambda})$ using Y statistics [87].

3. For a certain anti-forensic strength α, calculate $j_{\min}$, $j_{\max}$, and γ_{kj} using (5.22) and (5.26).

4. For each $Y = kq_1$, transcode it to W_α according to equation (5.27).

5. Apply lossless entropy coding similar as that used in JPEG compression to obtain the undetectable double JPEG compressed file.

We note that, for a certain anti-forensic strength and recompression quantization table, by either applying the anti-forensic dither and then recompressing, or directly applying the anti-forensic dither, the forger can obtain the same double JPEG compressed and anti-forensically modified image file.

5.5 Simulation Results and Analysis

In order to characterize the C-R-D tradeoff, we set up an experiment to obtain the reachable C-R-D values. We used the flexible anti-forensic dither to apply anti-forensics with adjustable strength. During the experiment, different strengths of anti-forensics and different quality factors of the recompression were used. Then, based on the data, we characterized the tradeoff using polynomial surfaces. Two surprising results were found during the analysis of the simulation results.

5.5.1 Two C-R-D Tradeoffs Revealed From Simulation

To experimentally characterize the C-R-D surface, we compressed, then anti-forensically modified and recompressed a set of images using a variety of JPEG quality factors and anti-forensic strengths. We then measured the concealability, rate, and distortion of each pairing of quality factor and anti-forensic strength, and used the resulting data to characterize the C-R-D surface.

We set up the simulation database based on the 1300 natural unaltered images from UCID database [82]. We examine the behavior of the forger, who can vary the anti-forensic strength and the quality factor of the recompression. So we fixed the first quality factor $Q_1 = 75$, and varied the secondary quality factor Q_2 from 60 to 90 with incremental interval 1. Then, we took 1000 unaltered images from the UCID database and JPEG compress each one using quality factors Q_2 to build the single compressed image database for training. The training database of double compressed images were obtained by compressing the same 1000 unaltered images using quality factor 75 and then recompressing them using secondary quality factors Q_2. Thus, the training database in our simulation contained $1000 \times 31 \times 2 = 62000$ images. Our testing database involved single compressed images, double compressed images and double compressed but anti-forensically modified images. The single compressed images for testing were composed by compressing the rest 300 unaltered images from the UCID database using quality factors Q_2. The double compressed images and double compressed but anti-forensically modified images were obtained by first compressing the same 300 unaltered images using quality factor

75, then applying anti-forensic dithers with strengths taken from range $[0, 1]$, and lastly recompressing them using secondary quality factors Q_2. We used 11 different anti-forensic strengths from $[0, 1]$ for each secondary quality factor. Therefore, we finally built up a testing database containing $300 \times (31 + 31 \times 11) = 111600$ images. The numbers of images used in our experiment are summarized in Table 5.1.

Table 5.1: Numbers of images in (a) training database and (b) testing database that were used in our experiment.

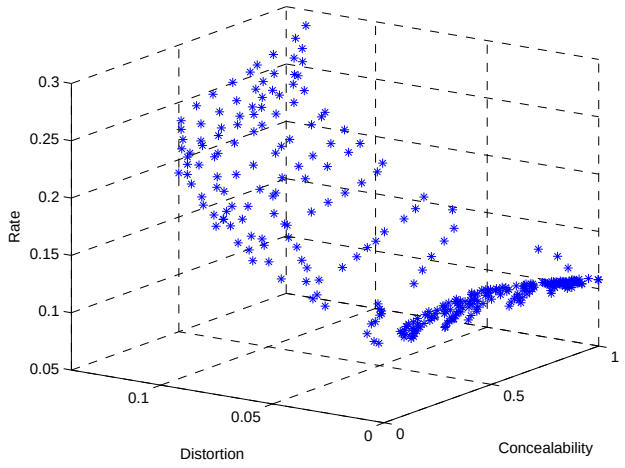
(a)

	# of different image content	# of different Q_2	Total # of images
H_0	1000	31	31000
H_1	1000	31	31000

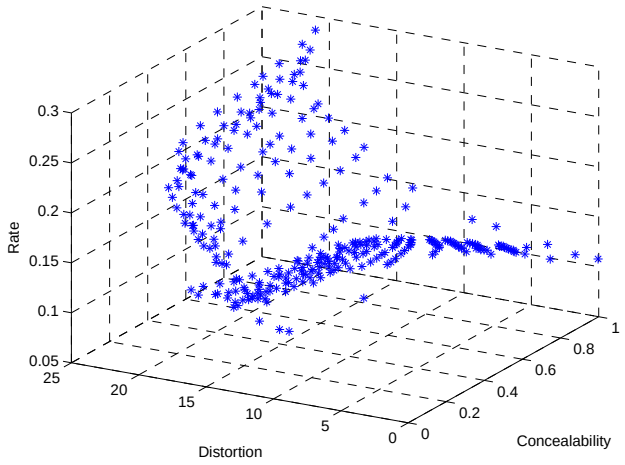
(b)

	# of different image content	# of different Q_2	# of different α	Total # of images
H_0	300	31	1	9300
H_1	300	31	11	102300

In order to characterize the C-R-D tradeoff, we calculate the concealability, rate, and distortion for each pair of anti-forensic strength and secondary quality factor. The detection technique described in Section 5.1.3 developed by Pevný et al.



(a)



(b)

Figure 5.6: Concealability, rate, and distortion triples for all tested anti-forensic strengths and secondary quality factors with distortion defined based on (a) MSSIM in (5.11) and (b) MSE.

was used to perform double JPEG compression detection. Different detectors were trained for each secondary quality factor using images from the training database described in the above paragraph. The false alarm rate is taken as 5%. Rate and distortion are calculated as the mean values of all the testing images with the same anti-forensic strength and secondary quality factor. Besides using (5.11) to calculate distortion, we also calculated mean square errors as an illustration of the results by applying other distortion measures. Based on the concealabilities, rates, and distortions obtained for different anti-forensic strengths and secondary quality factors, we plot each triple of concealability, rate, and distortion as a point in three dimensional figures in Fig. 5.6. Fig. 5.6(a) shows the tradeoff for using our definition of distortion in (5.11), and Fig. 5.6(b) is the tradeoff when we measure the distortion using the mean square error.

We find that, in both figures of Fig. 5.6, the points are separated into two surfaces. The lower surface is composed by the points where the secondary quality factor is lower than the primary quality factor. We call the tradeoff described by them as the lower quality factor tradeoff. The higher surface contains the points where the secondary quality factor is higher than the primary quality factor. This tradeoff is called the higher quality factor tradeoff. We note that the authors in [98] have found the similar phenomenon about separated cases for lower quality factors and higher quality factors when they studied the counter detector of the anti-forensic dither. Yet, they only considered the change on distortion, while our work characterizes the whole C-R-D tradeoff. We will study these two tradeoffs separately in the following two subsections. For the sake of space limitation, we

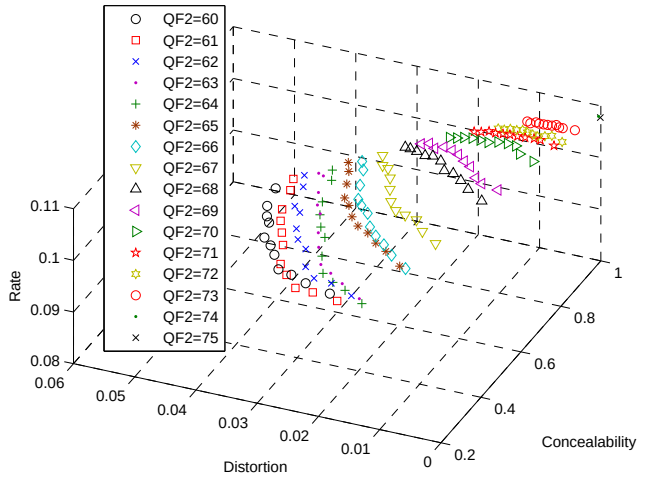
only give the detailed analysis to Fig. 5.6(a), while the other one can be analyzed similarly.

5.5.2 C-R-D Tradeoff for Lower Secondary Quality Factors

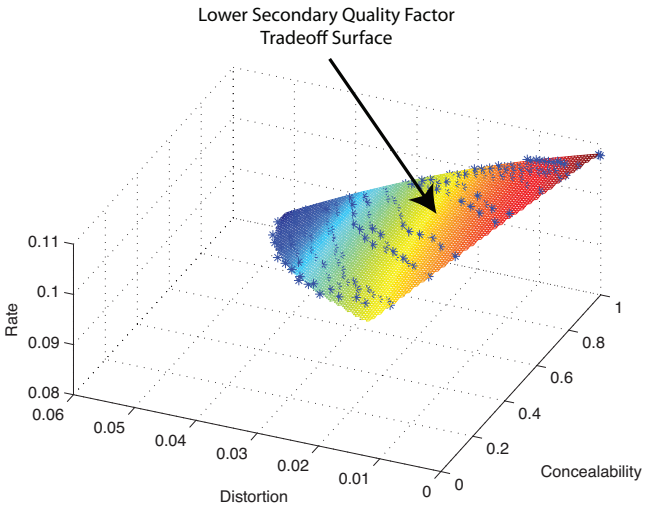
To characterize the C-R-D tradeoff for lower secondary quality factors, we plot those triple points obtained by using lower secondary quality factors in Fig. 5.7(a). Different markers represent different secondary quality factors. Each marker has several points obtained by using different anti-forensic strengths. Among them, the one with higher concealability implies that more anti-forensic strength has been applied to get this point. It is easy to see that increasing anti-forensic strength will increase concealability but also introduce more distortion.

Since anti-forensic dither adds noise to DCT coefficients, and typically a noisy signal is harder to be compressed, we would expect to get a higher rate when applying anti-forensics. However, we surprisingly find that, in the case of a lower secondary quality factor, applying anti-forensics will actually decrease the rate. We use a 2-D figure to more explicitly present this surprising result in Fig. 5.8.

This phenomenon happens due to the entropy coding procedure of JPEG compression. When quantization table is fixed, the rate of the compressed image depends on the entropy of the DCT coefficients. Since the coefficient histogram describes its probability density function, we can use the normalized histogram to compare the entropy. Furthermore, when the normalized histogram is closer to the uniform distribution, it implies a higher entropy of the coefficient. With anti-forensics applied



(a)



(b)

Figure 5.7: Tradeoff of concealability, rate, and distortion for the case where the second quality factor is smaller than the first one. (a) plots the reachable (C,R,D) points, where the points with the same marker and color are those who have the same secondary compression quality factor but have been applied different anti-forensic strengths. The higher the concealability, the more the anti-forensic strength. (b) is the polynominal fitting surface of (a).

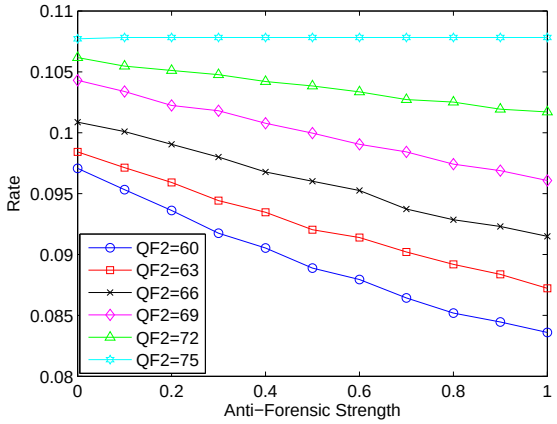


Figure 5.8: Rate changes with anti-forensic strength for lower secondary quality factor case.

to the double compressed image, it gradually changes the coefficient histogram from a double compressed histogram to a single compressed one. Thus, we can compare the entropies of these two cases to see how does anti-forensics affect the rate. Recall the typical coefficient histograms for single compressed and double compressed images shown in Fig. 5.1. It is easy to see that the entropy of the single compressed coefficient (histogram is shown in Fig. 5.1(a)) is less than that of the double compressed one for lower quality factor case (histogram is shown in Fig. 5.1(b)), where $q_2 > q_1$, i.e., $Q_2 < Q_1$. Thus, when anti-forensics change the histogram from the double compressed one to the single compressed one, it decreases the rate. However, similar argument implies that the result will be reversed for higher secondary quality factor scenario.

Next, we characterize the lower secondary quality factor tradeoff using a poly-

nomial surface, as it is shown in Fig. 5.7(b). The expression for the surface is

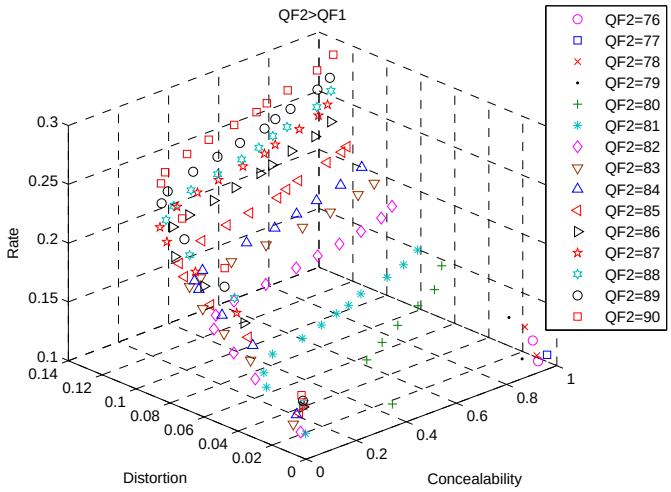
$$R = 0.1018 + 0.0088C - 0.238D - 0.0025C^2 - 0.1037CD - 2.771D^2, \quad (5.28)$$

where C, R, and D are concealability, rate, and distortion calculated from (5.10), (5.12), and (5.11), respectively. We obtain this equation by modeling R as a polynomial function of C and D. Then, we varied the degrees of freedom on both C and D to obtain the best fitting that yielding the minimum fitting error. We used the curve fitting toolbox in Matlab to implement this process. Similar approaches will be applied to obtain the tradeoff surfaces for the higher secondary quality factor case.

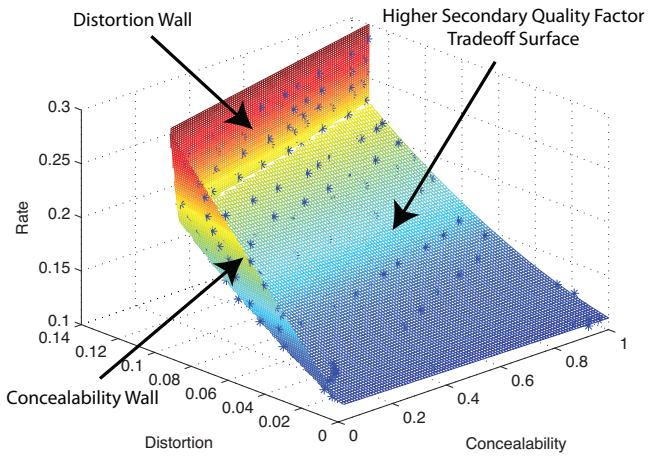
In (5.28), for a fixed C, R decreases with D, which matches the property of conventional $R - D$ curve. The $C - D$ tradeoff for a certain R is that increasing C will increase D. When D is fixed, by a little calculation on (5.28) we find that for most of the cases where $D < 0.037$, R increases with C. In this case, there exists a $R - C$ tradeoff, where increasing concealability will increase the rate. We note that this $R-C$ tradeoff is different from our previously mentioned surprising result, where increasing anti-forensic strength results in increase on the concealability and decrease on the rate. The former is a tradeoff for a certain distortion value, while the latter implies changes on distortion with the increase of anti-forensic strength.

5.5.3 C-R-D Tradeoff for Higher Secondary Quality Factors

To characterize the C-R-D tradeoff for higher secondary quality factors, we plot the rest triple points obtained by using higher secondary quality factors in Fig.



(a)



(b)

Figure 5.9: Tradeoff of concealability, rate, and distortion for the higher secondary quality factor case. (a) plots the R-D-C points. Points with the same marker and same color are those obtained by using the same secondary quality factor but different anti-forensic strengths. (b) is the polynominal fitting surfaces of (a).

5.9(a). Again, different markers represent different secondary quality factors. Each marker has several points obtained by using different anti-forensic strengths. Among them, the one with higher concealability implies that more anti-forensic strength has been applied to get this point. In this tradeoff, the reachable points of concealability, rate, and distortion depict three surfaces, which we use polynomial surfaces to fit.

As it is shown in Fig. 5.9(b), the main tradeoff surface for higher secondary quality factor is expressed as

$$R = 0.1146 - 0.0038C + 0.5474D - 0.15CD + 3.738D^2. \tag{5.29}$$

In this tradeoff, the $R - D$ tradeoff for a certain C is that the increase rate will also increase distortion. It is inconsistent with the conventional $R - D$ tradeoff, where distortion is reduced by the increase of data rate. This phenomenon happens due to the fact that, in higher secondary quality factor case, anti-forensic modification introduces much more distortion than recompression. Specifically, when using higher secondary quality factors, as the quality factor increases, double compression fingerprints will be more obvious and harder to conceal. Thus, more anti-forensic modification is needed to achieve the expected concealability. This results in the increase of distortion for higher secondary quality factor and consequently higher rate. From the expression, we can find the $R - C$ tradeoff for a fixed D is that increasing C will decrease R. This is also a result due to the distortion of the anti-forensic modification: when C increases, it implies that more anti-forensic strength has been applied, and thus more distortion has been introduced by anti-forensic modification. Then, in order to keep D unchanged, the distortion from recompression must be re-

duced, which means the secondary quality factor should be closer to the first quality factor. Since $Q_2 > Q_1$, it results in a lower R. Additionally, when we fix R, D will increase with higher C.

Besides the higher secondary quality factor tradeoff surface, there are two walls along the concealability axis and distortion axis. Which we call the concealability wall:

$$R = 0.1378 - 2.0084C + 2.9504D, \tag{5.30}$$

and the distortion wall:

$$R = 39.7255 + 118.4314C - 392.1569D. \tag{5.31}$$

The concealability wall is generated for small anti-forensic strengths. Specifically, because the double compression fingerprints for $Q_2 > Q_1$ is very distinctive, when anti-forensic strength is small, the increase on anti-forensic strength hardly changes C. However, the distortion introduced by anti-forensic modification increases proportionally with the strength, and thus it leads to the increase of R and D. Therefore, while R and D are increasing, the little change on C results in the concealability wall. The distortion wall happens for much higher quality factors, where recompression distortion decreases with finer quantization, i.e., higher quality factor, but anti-forensics distortion increases with higher quality factor. Thus, the summation of these two distortions results in the little change on overall distortion and the distortion wall appears.

When comparing the higher secondary quality factor tradeoff with the lower secondary quality factor tradeoff, we notice that the lower secondary quality factor

tradeoff locates entirely below the higher secondary quality factor tradeoff, as it is shown in Fig. 5.6. This implies that using a lower secondary quality factor can achieve the same concealability and distortion as the one obtained by using a higher quality factor, while the rate is lower. Note that we consider the data rate in this chapter, which is inversely proportional to the compression rate. Thus, such phenomenon induces the forger to choose a lower secondary quality factor rather than a higher one to obtain a lower rate without increasing the distortion or decreasing the concealability. This surprising behavior happens because that the anti-forensic modification introduces much more distortion in higher secondary quality factor case than in lower secondary quality factor case. Since double compression fingerprints are more obvious in higher secondary quality factor case than in the lower one, in order to achieve the same concealability, anti-forensic modification will introduce much more distortion when the forger decides to use a higher secondary quality factor. Thus, to achieve a certain concealability, using higher secondary quality factors will not only results in more distortion but also higher rate than the case of using lower secondary quality factors. As a consequence, the forger will always tend to use a lower secondary quality factor rather than a higher one.

5.6 Summary

In this chapter, we proposed a concealability-rate-distortion tradeoff in anti-forensic systems. Specifically, we defined concealability and characterized the C-R-D tradeoff in double JPEG compression anti-forensics. To obtain the tradeoff, we

proposed a flexible anti-forensic dither to vary the strength of anti-forensics. We also provided an anti-forensic transcoder to more efficiently accomplish the tasks of anti-forensics and recompression. We then experimentally characterized the C-R-D tradeoff by polynomial surfaces regarding whether the secondary quality factor is lower or higher than the first one. From the experimental results, we found two surprising results. The first one is that if the forger recompresses using a lower secondary quality factor, applying anti-forensics with greater strength will decrease the data rate. The second one is that the forger is always incentivized to recompress using a lower secondary quality factor. This is because our results have shown that, for any pairing of concealability and distortion values achieved by a higher secondary quality factor, the forger can choose a lower secondary quality factor that will achieve the same concealability and distortion values yet at a lower data rate.

Chapter 6

Conclusions and Future Work

6.1 Conclusions

In this dissertation, we studied the fundamental limits in multimedia forensics and anti-forensics. First, by acknowledging the limit of existing forensic techniques when new technologies emerge, we proposed a set of forensic tools to conquer this scenario. Furthermore, we explored the capability of forensic investigators when estimating the processing history of multimedia content. Information theoretical frameworks have been proposed to formulate conventional forensic problems, where mutual information based measurements and criteria were used to determine the limit of what we can do. By studying the behavior of forgers, we proposed and characterized the fundamental tradeoff in compression anti-forensics. More specifically, this dissertation contributes to the following aspects of multimedia forensics and anti-forensics.

In chapter 2, we proposed a set of forensic techniques to identify if a given signal was captured by compressive sensing. To do this, we categorized feasible signals that can be compressively sensed into three classes and found the fingerprints of compressive sensing in each class of signals. Then, depending on the amount of knowledge we know about the signal model, zero ratio detector and distribution-based detector were proposed to distinguish compressively sensed signals from traditionally sensed

signals. Specifically for images, because wavelet-based compressed images can be easily confused with compressively sensed images by existing forensic techniques, we proposed a two step detector to distinguish compressively sensed images from both traditionally sensed images and traditionally sensed but wavelet-based compressed images. In addition, we proposed a technique to estimate the number of compressive measurements used to acquire a compressively sensed signal. Experimental results have shown the effectiveness of our proposed techniques on a wide variety of signals.

In chapter 3, we proposed information theoretical frameworks for operation forensics to find how many operations we can detect at most. We have introduced the concept of forensicability as the maximum information that extracted features contain about the considered multimedia states. We used mutual information between features and multimedia states to quantify forensicability. Given forensicability, we obtained the lower bound of error probabilities for all estimators based on certain features. Then, using this lower bound, we proposed the concept of expected perfect detection to determine when we cannot detect any more operations. A case study of multiple JPEG compression detection has been examined in this chapter. By applying our framework and based on experimental results, we have found that, under typical forensic settings, the maximum number of JPEG compressions we can detect is 4. In addition, optimal strategies for both investigators and forgers have been discussed based on forensicability.

In chapter 4, we studied the forensic problem of detecting the order of operations and proposed information theoretical criteria to determine when we can and cannot detect the order. The problem of order detection has been formulated as

a multiple hypotheses estimation problem. Then, the mutual information between estimated hypotheses and true hypotheses has been used as a measure to evaluate the best estimation performance. Based on this criterion, conditions of when we can and cannot detect the order of operations have been proposed. Furthermore, by introducing the concept of conditional fingerprints, we were able to analyze the reason of why we cannot detect the orders. Three case studies have been involved in demonstrating the effectiveness of our proposed framework and criteria. The existing problems of double JPEG compression detection and order detection of resizing and contrast enhancement were examined to show that the results obtained by our framework match those from existing literatures. In addition, we examined the order detection of resizing and blurring. A forensic technique was proposed to detect their order. Based on our framework, we have found the regions of detectable cases for the order detection of resizing and blurring.

In chapter 5, we explored the fundamental tradeoff in compression anti-forensics. We have found that, in anti-forensic systems where compression is applied in the end for storage or transmission, there are three factors that forgers concern: concealability, rate and distortion. We defined the fundamental measurement of anti-forensic, concealability, as how effective the anti-forensic technique can conceal the manipulation fingerprints. Then, a case study of double JPEG compression anti-forensics was used to demonstrate the fundamental tradeoff between concealability, rate and distortion. In order to characterize this tradeoff, we proposed a flexible anti-forensic dither and anti-forensic transcoder to vary the strength of anti-forensics. Then, we experimentally characterize the tradeoff using polynomial surfaces. Two separated

tradeoffs have been discovered regarding the relationship between two compression quality factors. In addition, we have found two surprising phenomena from the analysis. One is that when forgers recompress using a lower secondary quality factor, increasing anti-forensic strength will decrease the data rate. The other one is that forgers are incentivized to recompress an image with a lower secondary quality factor because it can achieve lower data rate with the same concealability and distortion.

6.2 Future Work

As we have shown in this dissertation, exploring fundamental limits is an inevitable trend in multimedia forensics and anti-forensics. Various theoretical frameworks will be needed to formulate different types of forensic problems and obtain their own fundamental limits or constraints. In this dissertation, we have examined the fundamental limits in operation forensics, order forensics, and compression anti-forensics. There are many other challenging forensic and anti-forensic scenarios where I will continue to explore their fundamental limits.

In our proposed information theoretical framework for operation forensics, a single set of features are considered. Thus, the analysis involves only one abstract channel between the features and the multimedia states. Similarly, in order forensics, we considered one set of estimators in the framework and analyzed a single abstract channel between estimated hypotheses and true hypotheses. However, as forensic techniques develop, different features or different sets of estimators have been found and used for the same forensic purpose [19, 55, 76]. While our frameworks provide a

way to compare the fundamental performance of forensic techniques using different features or estimators, it would be more interesting to find the optimal performance of combining all feasible features or estimators. This involves in introducing multiple abstract channels in the analysis and formulating the relationships among these channels. Based on appropriate frameworks, we want to find the fundamental limit of fusing different features or estimators towards the same forensic purpose and find the optimal fusion algorithm for the best detection performance.

In this dissertation, we have examined anti-forensics where forgers use anti-forensic techniques to confuse forensic investigators. With the development of anti-forensic techniques, counter anti-forensic schemes have been developed by forensic investigators to detect the trace of anti-forensics [40, 98, 102]. By understanding counter anti-forensic techniques, forgers may, again, developing corresponding anti-forensics to attack these new forensic schemes. As such interactions between forensics and anti-forensics continue, it is interesting to know whether they will come to an equilibrium and when it will happen. To answer these, we need to formulate the relationship between forensicability and concealability, the two fundamental measurements we proposed in this dissertation for multimedia forensics and anti-forensics respectively. By appropriately choosing theoretical frameworks, we want to know how forensicability and concealability change during the interplay between investigators and forgers. Furthermore, we would like to explore the maximum number of interactions needed to achieve the equilibrium, if any, and what the ultimate state would be.

Bibliography

[1] Bell labs invents lensless camera. *MIT Technology Review*, May 2013.

[2] E. G. Allstot, A. Y. Chen, A. M. R. Dixon, D. Gangopadhyay, and D. J. Allstot. Compressive sampling of ECG bio-signals: Quantization noise and sparsity considerations. *Proc. IEEE BioCAS*, pages 41–44, Nov. 2010.

[3] M. Barni and A. Costanzo. A fuzzy approach to deal with uncertainty in image forensics. *Signal Processing: Image Communication*, 27(9):998 – 1010, 2012.

[4] M. Barni, M. Fontani, and B. Tondi. A universal technique to hide traces of histogram-based image manipulations. In *Proceedings of the on Multimedia and Security*, MM&Sec '12, pages 97–104. ACM, 2012.

[5] T. Bianchi and A. Piva. Image forgery localization via block-grained analysis of JPEG artifacts. *IEEE Trans. on Information Forensics and Security*, 7(3):1003–1017, Jun. 2012.

[6] T. Bianchi and A. Piva. Reverse engineering of double jpeg compression in the presence of image resizing. In *IEEE International Workshop on Information Forensics and Security*, pages 127–132, Dec. 2012.

[7] T. Bianchi, A. De Rosa, and A. Piva. Improved dct coefficient analysis for forgery localization in JPEG images. *Proc. IEEE ICASSP*, pages 2444–2447, May 2011.

[8] J. Bobin, J.-L. Starck, and R. Ottensamer. Compressed sensing in astronomy. *IEEE Journal of Selected Topics in Signal Processing*, 2(5):718–726, Oct. 2008.

[9] D. H. Brainard. Hyperspectral image data. `http://color.psych.upenn.edu//hyperspectral/`.

[10] C. J. C. Burges. A tutorial on support vector machines for pattern recognition. *Data Mining and Knowledge Discovery*, 2(2):121–167, 1998.

[11] E. J. Candes, J. Romberg, and T. Tao. Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information. *IEEE Transactions on Information Theory*, 52(2):489–509, Feb. 2006.

[12] E. J. Candes and T. Tao. Near-optimal signal recovery from random projections: Universal encoding strategies? *IEEE Transactions on Information Theory*, 52(12):5406–5425, Dec. 2006.

[13] G. Cao, Y. Zhao, and R. Ni. Edge-based blur metric for tamper detection. pages 20–27, Jan. 2010.

[14] G. Cao, Y. Zhao, R. Ni, and X. Li. Contrast enhancement-based forensics in digital images. *IEEE Transactions on Information Forensics and Security*, 9(3):515–525, Mar. 2014.

[15] G. Cao, Y. Zhao, R. Ni, and H. Tian. Anti-forensics of contrast enhancement in digital images. In *Proc. 12th ACM workshop on Multimedia and security*, pages 25–34, 2010.

[16] M. Chen, J. Fridrich, M. Goljan, and J. Lukáš. Source digital camcorder identification using sensor photo-response non-uniformity. *IEEE Signal Processing Magazine*, 26(2):16–25, Mar. 2009.

[17] S. S. Chen, D. L. Donoho, and M. A. Saunders. Atomic decomposition by basis pursuit. *SIAM Journal on Scientific Computing*, 20(1):33–61, 1998.

[18] Y. Chen, W. S. Lin, and K. J. R. Liu. Risk-distortion analysis for video collusion attacks: A mouse-and-cat game. *IEEE Transactions on Image Processing*, 19(7):1798–1807, Jul. 2010.

[19] Y. L. Chen and C. T. Hsu. Detecting doubly compressed images based on quantization noise model and image restoration. In *IEEE International Workshop on Multimedia Signal Processing*, pages 1–6, Oct. 2009.

[20] X. Chu, Y. Chen, M. C. Stamm, and K. J. R. Liu. Information theoretical limit of compression forensics. In *IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 2689–2693, May 2014.

[21] X. Chu, M. C. Stamm, Y. Chen, and K. J. R. Liu. Concealability-rate-distortion tradeoff in image compression anti-forensics. In *Proc. IEEE ICASSP, 2013*, pages 3063–3067, May 2013.

[22] X. Chu, M. C. Stamm, and K. J. R. Liu. Compressive sensing forensics. *IEEE Transactions on Information Forensics and Security*, (99), Mar. 2015.

[23] X. Chu, M. C. Stamm, and K. J. Ray Liu. Forensic identification of compressively sensed signals. *Proc. IEEE ICIP*, pages 257–260, Sept. 2012.

[24] Xiaoyu Chu, M.C. Stamm, W.S. Lin, and K.J.R. Liu. Forensic identification of compressively sensed images. In *Acoustics, Speech and Signal Processing (ICASSP), 2012 IEEE International Conference on*, pages 1837–1840, March 2012.

[25] P. Comesaña. Detection and information theoretic measures for quantifying the distinguishability between multimedia operator chains. In *IEEE Workshop on Information Forensics and Security*, Tenerife, Spain, 2012.

[26] P. Comesana-Alfaro and F. Perez-Gonzalez. Optimal counterforensics for histogram-based forensics. *Proc. IEEE ICASSP*, May 2013.

[27] InView Company. *http://inviewcorp.com*.

[28] T. M. Cover and J. A. Thomas. *Elements of Information Theory, second edition*. John Wiley & Sons, Inc., Hoboken, NJ, USA, 2006.

[29] D. L. Donoho. Compressed sensing. *IEEE Transactions on Information Theory*, 52(4):1289–1306, Apr. 2006.

[30] W. Fan, K. Wang, F. Cayre, and Z. Xiong. A variational approach to JPEG antiforensics. *Proc. IEEE ICASSP*, May 2013.

[31] Z. Fan and R. L. de Queiroz. Identification of bitmap compression history: JPEG detection and quantizer estimation. *IEEE Transactions on Image Processing*, 12(2):230 – 235, 2003.

[32] H. Farid. Image forgery detection. *IEEE Signal Processing Magazine*, 26(2):16–25, Mar. 2009.

[33] H. Farid. Digital image ballistics from JPEG quantization. Dept. of Computer Science, Dartmouth College, Tech. Rep. TR2006-583, 2006.

[34] H. G. Feichtinger. Atomic characterizations of modulation spaces through gabor-type representations. *Rocky Mountain Journal of Mathematics*, 19(1):113–125, 1989.

[35] X. Feng and G. Doërr. JPEG recompression detection. In *Proc. of SPIE, Media Forensics and Security II*, volume 7541, pages 0J1–0J10, Feb. 2010.

[36] P. Ferrara, T. Bianchi, A. De Rosa, and A. Piva. Reverse engineering of double compressed images in the presence of contrast enhancement. In *IEEE 15th International Workshop on Multimedia Signal Processing*, pages 141–146, Sep. 2013.

[37] B. Foo, Y. Andreopoulos, and M. van der Schaar. Analytical rate-distortion-complexity modeling of wavelet-based video coders. *IEEE Transactions on Signal Processing*, 54(2):797–815, Feb. 2008.

[38] R. Garg, A. L. Varna, and M. Wu. "seeing" enf: natural time stamp for digital video via optical sensing and signal processing. In *Proceedings of the 19th ACM international conference on Multimedia*, MM '11, pages 23–32, New York, NY, USA, 2011. ACM.

[39] Thomas Gloe and Rainer Böhme. The 'dresden image database' for benchmarking digital image forensics. In *ACM Symposium on Applied Computing*, volume 2, pages 1584–1590, 2010.

[40] Miroslav Goljan, Jessica Fridrich, and Mo Chen. Sensor noise camera identification: countering counter-forensics. *Proc. SPIE*, 7541:75410S–75410S–12, 2010.

[41] J. Haupt, W. U. Bajwa, M. Rabbat, and R. Nowak. Compressed sensing for networked data. *IEEE Signal Processing Magazine*, 25(2):92–101, Mar. 2008.

[42] Z. He, Y. Liang, L. Chen, I. Ahmad, and D. Wu. Power-rate-distortion analysis for wireless video communication under energy constraints. *IEEE Transactions on Circuits and Systems for Video Technology*, 15(5):645–658, May 2005.

[43] M. A. Herman and T. Strohmer. High-resolution radar via compressed sensing. *IEEE Transactions on Signal Processing*, 57(6):2275–2284, Jun. 2009.

[44] F. Huang, J. Huang, and Y. Q. Shi. Detecting double JPEG compression with the same quantization matrix. *IEEE Transactions on Information Forensics and Security*, 5(4):848 –856, Dec. 2010.

[45] C. Huo, R. Zhang, and D. Yin. Compression technique for compressed sensing hyperspectral images. *International Journal of Remote Sensing*, 33(5):1586–1604, 2012.

[46] S. Ji, Y. Xue, and L. Carin. Bayesian compressive sensing. *IEEE Transactions on Signal Processing*, 56(6):2346 –2356, June 2008.

[47] O. Katz, Y. Bromberg, and Y. Silberberg. Compressive ghost imaging. *Applied Physics Letters*, 95(13):131110, 2009.

[48] M. Kirchner. Fast and reliable resampling detection by spectral analysis of fixed linear predictor residue. In *Proceedings of the 10th ACM Workshop on Multimedia and Security*, MM&Sec '08, pages 11–20, New York, NY, USA, 2008. ACM.

[49] M. Kirchner and R. Bohme. Hiding traces of resampling in digital images. *IEEE Trans. on Information Forensics and Security*, 3(4):582 –592, Dec. 2008.

[50] M. Kirchner and J. Fridrich. On detection of median filtering in digital images. *Media Forensics and Security II, Proc. of SPIE-IS&T Electronic Imaging, SPIE*, 7541, 754110, 2010.

[51] H. W. Kuhn and A. W. Tucker. Nonlinear programming. *Proc. 2nd Berkeley Symposium on Mathematical Statistics and Probability*, pages 481–492, 1951.

[52] S. Lai and R. Bohme. Countering counter-forensics: The case of JPEG compression. In *Information Hiding*, volume 6958 of *Lecture Notes in Computer Science*, pages 285–298. Springer Berlin Heidelberg, 2011.

[53] S.-Y. Lai and R. Bohme. Block convergence in repeated transform coding: JPEG-100 forensics, carbon dating, and tamper detection. In *Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on*, pages 3028–3032. IEEE, 2013.

[54] E. Y. Lam. A mathematical analysis of the DCT coefficient distributions for images. *IEEE Transactions on Image Processing*, 9(10):1661–1666, Oct. 2000.

[55] B. Li, Y. Q. Shi, and J. Huang. Detecting doubly compressed jpeg images by using mode based first digit features. In *Multimedia Signal Processing, 2008 IEEE 10th Workshop on*, pages 730–735, Oct 2008.

[56] H. Li, W. Luo, and J. Huang. Countering anti-JPEG compression forensics. In *Image Processing (ICIP), 2012 19th IEEE International Conference on*, pages 241–244, Sept 2012.

[57] Haodong Li, Weiqi Luo, and Jiwu Huang. Anti-forensics of double jpeg compression with the same quantization matrix. *Multimedia Tools and Applications*, pages 1–16, 2014.

[58] 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*, 4(3):460–475, Sep. 2009.

[59] J. Lukáš, J. Fridrich, and M. Goljan. Digital camera identification from sensor pattern noise. *IEEE Transactions on Information Forensics and Security*, 1(2):205–214, Jun. 2006.

[60] Da Luo, Rui Yang, and Jiwu Huang. Detecting double compressed amr audio using deep learning. In *Acoustics, Speech and Signal Processing (ICASSP), 2014 IEEE International Conference on*, pages 2669–2673, May 2014.

[61] W. Luo, Y. Wang, and J. Huang. Detection of quantization artifacts and its applications to transform encoder identification. *IEEE Transactions on Information Forensics and Security*, 5(4):810–815, Dec. 2010.

[62] M. Lustig, D. Donoho, and J. M. Pauly. Sparse MRI: The application of compressed sensing for rapid MR imaging. *Magnetic Resonance in Medicine*, 58:1182–1195, 2007.

[63] M. Lustig, D. L. Donoho, J. M. Santos, and J. M. Pauly. Compressed sensing MRI. *IEEE Signal Processing Magazine*, 25(2):72–82, Mar. 2008.

[64] B. Mahdian and S. Saic. Detecting double compressed JPEG images. *3rd International Conference on Crime Detection and Prevention*, pages 1–6, Dec. 2009.

[65] S. Mallat. *A Wavelet Tour of Signal Processing, Third Edition: The Sparse Way*. Academic Press, 3rd edition, 2008.

[66] S. Milani, M. Tagliasacchi, and S. Tubaro. Discriminating multiple jpeg compression using first digit features. In *IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 2253–2256. IEEE, 2012.

[67] T. K. Moon. The expectation-maximization algorithm. *IEEE Signal Processing Magazine*, 13(6):47–60, Nov. 1996 http://www.cs.ubc.ca/labs/scl/spgl1/.

[68] S. X. Ng and L. Hanzo. On the MIMO channel capacity of multidimensional signal sets. *Vehicular Technology, IEEE Transactions on*, 55(2):528–536, March 2006.

[69] J. F. O'Brien and H. Farid. Exposing photo manipulation with inconsistent reflections. *ACM Transactions on Graphics*, 31(1):4:1–11, January 2012.

[70] A. Ortega and K. Ramchandran. Rate-distortion methods for image and video compression. *IEEE Signal Processing Magazine*, 15(6):23–50, Nov. 1998.

[71] V. M. Patel, G. R. Easley, D. M. Healy, and R. Chellappa. Compressed synthetic aperture radar. *IEEE Journal of Selected Topics in Signal Processing*, 4(2):244 –254, Apr. 2010.

[72] W. Pennebaker and J. Mitchell. *JPEG : Still Image Data Compression Standard*. Van Norstrand Reinhold, 1993.

[73] T. Pevný and J. Fridrich. Estimation of primary quantization matrix in double compressed JPEG images. In *Proc. of Digital Forensic Research Workshop*, Cleveland, Ohio, Aug. 2003.

[74] T. Pevný and J. Fridrich. Detection of double-compression in JPEG images for applications in steganography. *IEEE Trans. on Information Forensics and Security*, 3(2):247–258, Jun. 2008.

[75] S. Pfennig and M. Kirchner. Spectral methods to determine the exact scaling factor of resampled digital images. In *5th International Symposium on Communications Control and Signal Processing*, pages 1–6, 2012.

[76] A. C. Popescu and H. Farid. Statistical tools for digital forensics. In *6th International Workshop on Information Hiding*, Toronto, Canada, 2004.

[77] A. C. Popescu and H. Farid. Exposing digital forgeries by detecting traces of re-sampling. *IEEE Transactions on Signal Processing*, 53(2):758–767, Feb. 2005.

[78] A. C. Popescu and H. Farid. Exposing digital forgeries in color filter array interpolated images. *IEEE Transactions on Signal Processing*, 53(10):3948–3959, Oct. 2005.

[79] Rice Single-Pixel Camera Project. *http://dsp.rice.edu/cscamera*.

[80] J. Provost and F. Lesage. The application of compressed sensing for photo-acoustic tomography. *IEEE Transactions on Medical Image*, 28(4):585–594, Apr. 2009.

[81] Zhenxing Qian and Xinpeng Zhang. Improved anti-forensics of JPEG compression. *Journal of Systems and Software*, 91(0):100 – 108, 2014.

[82] G. Schaefer and M. Stich. UCID: An uncompressed color image database. *Proc. SPIE: Storage and Retrieval Methods and Applications for Multimedia*, 5307:472480, 2004.

[83] D. Schneider. New camera chip captures only what it needs. *IEEE Spectrum Magazine*, Mar. 2013.

[84] M. C. Stamm, X. Chu, and K. J. R. Liu. Forensically determining the order of signal processing operations. In *IEEE International Workshop on Information Forensics and Security*, pages 162–167, Nov. 2013.

[85] M. C. Stamm, W. S. Lin, and K. J. R. Liu. Temporal forensics and anti-forensics for motion compensated video. *IEEE Transactions on Information Forensics and Security*, 7(4):1315–1329, 2012.

[86] M. C. Stamm and K. J. R. Liu. Forensic detection of image manipulation using statistical intrinsic fingerprints. *IEEE Transactions on Information Forensics and Security*, 5(3):492–506, Sep. 2010.

[87] M. C. Stamm and K. J. R. Liu. Anti-forensics of digital image compression. *IEEE Transactions on Information Forensics and Security*, 6(3):1050–1065, Sep. 2011.

[88] M. C. Stamm, M. Wu, and K. J. R. Liu. Information forensics: An overview of the first decade. *IEEE Access*, 1:167–200, 2013.

[89] B. Su, S. Lu, and C. L. Tan. Blurred image region detection and classification. In *Proceedings of the 19th ACM International Conference on Multimedia*, MM '11, pages 1397–1400, New York, NY, USA, 2011. ACM.

[90] G. J. Sullivan and T. Wiegand. Rate-distortion optimization for video compression. *IEEE Signal Processing Magazine*, 15(6):74–90, Nov. 1998.

[91] T. Sun and K. Kelly. Compressive sensing hyperspectral imager. *Proc. Computational Optical Sensing and Imaging*, page CTuA5, Oct. 2009.

[92] A. Swaminathan, M. Wu, and K. J. R. Liu. A component estimation framework for information forensics. *in Proc. IEEE 9th Workshop on Multimedia Signal Processing*, pages 397–400, Oct. 2007.

[93] A. Swaminathan, M. Wu, and K. J. R. Liu. A pattern classification framework for theoretical analysis of component forensics. In *in Proc. IEEE ICASSP*, pages 1665–1668, March 2008.

[94] A. Swaminathan, M. Wu, and K. J. R. Liu. Component forensics. *IEEE Signal Processing Magazine*, 26(2):38–48, Mar. 2009.

[95] R. Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B*, 58:267–288, 1996.

[96] H. Tong, M. Li, H. Zhang, and C. Zhang. Blur detection for digital images using wavelet transform. In *Multimedia and Expo, 2004. ICME '04. 2004 IEEE International Conference on*, volume 1, pages 17–20 Vol.1, Jun. 2004.

[97] J. A. Tropp and A. C. Gilbert. Signal recovery from partial information via orthogonal matching pursuit. *IEEE Transactions on Information Theory*, 53(12):4655–4666, Dec. 2007.

[98] G. Valenzise, M. Tagliasacchi, and S. Tubaro. Revealing the traces of JPEG compression anti-forensics. *Information Forensics and Security, IEEE Transactions on*, 8(2):335–349, Feb 2013.

[99] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4):600–612, Apr. 2004.

[100] D. Wu, W.-P. Zhu, and M. N. S. Swamy. A compressive sensing method for noise reduction of speech and audio signals. *Proc. IEEE MWSCAS*, pages 1–4, Aug. 2011.

[101] Z.-H. Wu, M. C. Stamm, and K. J. R. Liu. Anti-forensics of median filtering. *IEEE International Conference on Acoustics, Speech and Signal Processing*, May 2013.

[102] H. Zeng, T. Qin, X. Kang, and L. Liu. Countering anti-forensics of median filtering. In *Proc. IEEE ICASSP*, May 2014.