

AGH University of Science and Technology
Faculty of Computer Science, Electronics and Telecommunications

Ph.D. Dissertation

Paweł Korus

Analysis of Image Reconstruction Schemes Based on Self-Embedding and Digital Watermarking

Supervisor:

Professor Andrzej Dziech, Ph.D. Eng.

AGH UNIVERSITY OF SCIENCE AND TECHNOLOGY
Faculty of Computer Science, Electronics and Telecommunications
Department of Telecommunications

al. Mickiewicza 30, 30-059 Kraków, Poland
tel. +48 12 6345582
fax +48 12 6342372

<http://www.agh.edu.pl>
<http://www.iet.agh.edu.pl>
<http://www.kt.agh.edu.pl>
<http://www.kt.agh.edu.pl/~korus/>



Copyright © Paweł Korus, 2012
All rights reserved

L^AT_EXtemplate by Rafał Stankiewicz

Acknowledgements

First of all, I would like to express my gratitude towards my supervisor, Professor Andrzej Dziech, who encouraged me to pursue my Ph.D. in the field of multimedia security. In particular, I would like to thank for providing excellent working conditions, and continuous support and encouragement to present my work at world-leading conferences, and in respected international journals.

Secondly, I have been fortunate to work in a friendly atmosphere among my colleagues at AGH University of Science and Technology. They were always ready to engage in stimulating discussions, and provide a critical outsider's perspective. In particular, I would like to thank Piotr Guzik for reading my dissertation with interest and providing constructive comments which helped me improve the presentation of the obtained results.

I would also like thank Professor Xinpeng Zhang, whom I had opportunity to visit at the Shanghai University, for fruitful discussions about various aspects of practical self-recovery schemes. I couldn't have imagined a warmer welcome in Shanghai.

This work would not have been possible without the support from my Family, and my beloved Magdalena. They managed to ensure me a perfect balance between the necessary focus on my research, and the time to relax.

Last but not least, I'd like to thank anonymous reviewers for their time and constructive feedback about my papers at the time of their submission.

Abstract

This dissertation deals with digital image authentication and reconstruction techniques based on fragile digital watermarking. The main focus of the presented analysis addresses the achievable reconstruction performance, the inherent restoration trade-offs, and the applicable operation bounds. The analysis is based on a new theoretical model of the self-embedding problem, and ultimately leads to the development of an efficient self-embedding scheme, capable of high-quality reconstruction even under extensive tampering. For a benchmark configuration of the proposed scheme, an average peak signal to noise ratio for 10,000 natural images is 37 dB, even when up to 50% of the image area becomes tampered.

The content reconstruction problem is modeled as an erasure communication channel, and implemented in practice with the use of digital fountain codes. With proper design of the self-embedding scheme, it is possible to exploit the remaining authentic image content, and overcome the limits of traditional erasure communication. It becomes possible to use reference streams significantly longer than the available embedding capacity.

The presented analysis is first performed for the case of uniform-fidelity reconstruction, and then extended to the case of adaptive self-embedding. The latter is an emerging research direction, which is of high importance in certain applications, e.g., in the authentication of closed circuit television footage. By controlling the reconstruction quality for each image block individually, a new degree of freedom is introduced. With such flexibility it becomes possible to bias the process towards better image quality, or better robustness against content modifications. In order to calculate the mapping between the image blocks and the reconstruction quality, a dedicated algorithm is proposed, which allows for providing guarantees on the quality of individual image fragments, and the achievable tampering rates.

The dissertation also discusses reconstruction quality optimization techniques. The commonly used uniform quantization procedure is replaced with optimal Lloyd-Max code-books, and the allocation of the reference payload is performed in

a formalized framework. Such an optimized self-embedding scheme is experimentally compared to 5 state-of-the-art alternative schemes in a common evaluation scenario.

All of the presented theoretical results are supported by exhaustive experimental evaluation. The obtained results lead to important insights on the design of efficient self-embedding schemes. For the sake of research reproducibility, the relevant Matlab code can be obtained at <http://kt.agh.edu.pl/~korus>.

Keywords image authentication, image reconstruction, self-embedding, self-recovery, digital watermarking, semi-fragile watermarking, erasure channel, digital fountain codes, Lloyd-Max quantization

Streszczenie

Tematyka niniejszej rozprawy dotyczy technik uwierzytelniania oraz rekonstrukcji treści obrazów cyfrowych bazujących na ulotnych cyfrowych znakach wodnych. W szczególności przedstawiona analiza dotyczy osiągalnej efektywności rekonstrukcji, właściwych jej kompromisów oraz granic poprawnej pracy. Podstawą przeprowadzonej analizy jest proponowany nowy teoretyczny model problemu samo-rekonstrukcji (ang. *self-embedding*), którego użycie prowadzi do efektywnego algorytmu, zdolnego do wysokiej jakości rekonstrukcji nawet w przypadku znaczących zmian w obrazie. Na zbiorze 10 000 naturalnych obrazów średnia wierność reprezentacji odtworzonej treści, wyrażona przez szczytowy współczynnik sygnału do szumu, wynosi 37 dB i jest osiągalna nawet gdy blisko 50% powierzchni kadru uległo zniszczeniu.

Problem rekonstrukcji przedstawiono jako komunikację przez kanał z wymazywaniem (ang. *erasure channel*). Opracowany w praktyce algorytm zaimplementowano w oparciu o kody fontannowe. Dzięki odpowiedniej konstrukcji algorytmu możliwe jest ponowne użycie pozostałej, autentycznej części obrazu, co pozwala na stosowanie strumieni referencyjnych o długości znacząco większej niż dostępna pojemność znaku wodnego.

Przeprowadzona analiza początkowo dotyczy wariantu algorytmu o jednolitej wierności rekonstrukcji. Jest ona następnie rozszerzona na wariant adaptacyjny. Adaptacyjna samo-rekonstrukcja jest nowym kierunkiem badań, a także istotnym elementem w niektórych zastosowaniach, np. w zabezpieczaniu obrazów pochodzących z telewizji przemysłowej. Dzięki możliwości sterowania wiernością rekonstrukcji na poziomie indywidualnych bloków obrazu, uzyskuje się dodatkowy stopień swobody pozwalający na dopasowanie procesu rekonstrukcji do potrzeb wyższej jakości, bądź odporności na bardziej rozległą podmianę treści. Niniejsza rozprawa przedstawia także algorytm przypisujący blokom obrazu parametry rekonstrukcji w taki sposób, aby spełnione były wymagania odnośnie poziomu jakości fragmentów obrazu, a także osiągalnej odporności na podmianę treści.

Przedmiotem dyskusji są także techniki optymalizacji wierności rekonstrukcji. Najczęściej stosowany jednorodny krok kwantyzacji został zastąpiony słownikiem Lloyd-Maxa, dopasowanym do rozkładu współczynników widma obrazów naturalnych. Precyzja, z jaką zapisywane są poszczególne współczynniki, wyznaczana jest automatycznie dla każdego z rozważanych przypadków przy użyciu dedykowanego algorytmu optymalizacyjnego. Wynikowy system rekonstrukcji obrazów, integrujący opisane usprawnienia przetestowano eksperymentalnie wraz z pięcioma najnowszymi systemami we wspólnym scenariuszu testowym.

Wszystkie przedstawione wyniki teoretyczne zweryfikowano eksperymentalnie. Uzyskane wyniki prowadzą do istotnych wniosków dotyczących projektowania efektywnych systemów zabezpieczania obrazów cyfrowych. W celu zapewnienia odtwarzalności przeprowadzonych badań, kod źródłowy wybranych eksperymentów można uzyskać po adresie <http://kt.agh.edu.pl/~korus>.

Słowa kluczowe uwierzytelnianie obrazów cyfrowych, rekonstrukcja obrazów cyfrowych, samo-rekonstrukcja, cyfrowe znaki wodne, ulotne znaki wodne, kanał z wymazywaniem, kodowanie fontannowe, kwantyzacja słownikiem Lloyd-Maxa

Contents

Contents	ix
List of Figures	xiii
List of Tables	xvii
Abbreviations	xix
Definitions	xxi
List of Symbols	xxiii
Introduction	1
1 Area of Research	7
1.1 Media Protection with Self-Embedding	7
1.2 Measuring the Reconstruction Performance	10
1.3 Related Work	14
1.3.1 Fundamental Reference Representations	14
1.3.2 Alternative Reference Representations	15
1.3.3 Flexible Self-Embedding	16
1.3.4 Adaptive Self-Embedding	18
1.3.5 Self-Embedding Summary	18
1.4 Limitations of Existing Schemes	22
2 Erasure Channel for Content Reconstruction	25
2.1 Formal Problem Statement	25
2.1.1 Success Bound Calculation	34

2.1.2	Optimistic Reconstruction Success Bound	35
2.1.3	Typical Reconstruction Success Bound	35
2.2	Experimental Evaluation	40
2.2.1	Reference Self-Embedding Scheme	40
2.2.2	Validation of the Reconstruction Demand Estimate	42
2.2.3	Validation of the Reconstruction Success Bounds	42
2.3	Conclusions and Practical Design Guidelines	47
3	Analysis of Content Adaptivity	49
3.1	Reconstruction Success Bound Analysis	50
3.1.1	Success Bound Types	51
3.1.2	Impact of Multiple Reconstruction Profiles	54
3.2	Automatic Design of Quality Descriptors	58
3.2.1	Descriptor Design Procedure	58
3.2.2	Experimental Evaluation of the Design Procedure	59
3.3	Conclusions	62
4	Evaluation of Reconstruction Quality Optimization Techniques	65
4.1	Reference Payload Allocation	67
4.1.1	Formal Problem Statement	67
4.1.2	Modeling the Objective Distortion	69
4.1.3	Solving the Reference Payload Allocation Problem	71
4.1.4	Optimization Results	74
4.2	Reconstruction Fidelity Improvement Techniques	75
4.2.1	Quantization Strategy Impact	75
4.2.2	Content Adaptivity Impact	80
4.2.3	Example Reference Images	85
5	Design and Evaluation of Practical Self-Embedding Schemes	95
5.1	Adaptive Self-Embedding Scheme	95
5.1.1	Definition of the Reconstruction Profiles	99
5.2	Experimental Evaluation	99
5.2.1	Reference Self-Embedding Scheme	100
5.2.2	Comparison with State-of-the-Art Schemes	106
5.2.3	Content-Adaptive Self-Embedding Scheme	120
5.2.4	Descriptor-Adaptive Self-Embedding Scheme	123
	Conclusions	129
	A Security and Protocols	135
	B Comparison with Reference Sharing	139

C Image Tests Sets	143
D Solutions to the Optimal Reference Allocation Problem	149
E Miscellaneous Practical Implementation Issues	159
References	161

List of Figures

1.1	Pro-active protection of digital images with self-embedding	8
1.2	Example recovery of a tampered video surveillance footage. . . .	13
2.1	M-ary symmetric erasure channel with probability of erasure p_e . .	27
2.2	Communication of the reconstruction reference and its corresponding auxiliary information between the encoder and the decoder. . .	29
2.3	Operation of the proposed content restoration approach	30
2.4	Operation of the proposed content restoration approach for perfect alignment between the reference blocks and symbols.	31
2.5	Decoding problem reduction in matrix representation of the random linear fountain code	33
2.6	Family of $1 - \gamma^{\alpha(\lambda)}$ functions for $\alpha = \frac{1}{\lambda}$	36
2.7	Impact of the misalignment between the reference blocks and symbols on the number of overlapping blocks per single symbol	37
2.8	Derived reconstruction success bounds.	41
2.9	Theoretical vs. empirical estimates of the shape parameter α	43
2.10	Experimental and theoretical dependency between the reconstruction demand and the tampering rate for selected reference rates . .	44
2.11	Experimental evaluation of the reconstruction success bounds γ_1 and γ_2 using Monte Carlo simulations	45
2.12	Experimental evaluation of the reconstruction success bounds γ_3 and γ_2 using Monte Carlo simulations	46
3.1	Graphical interpretation of the reconstruction success bound. . . .	51
3.2	Behavior of the reconstruction demand for pessimistic tampering. .	53
3.3	Graphical interpretation of the average tampering rate bound . . .	54

3.4	Impact of multiple reconstruction profiles on the pessimistic, and the average success bounds for configurations with two, and three reconstruction profiles.	57
3.5	Experimental validation of the descriptor design objective for 1,000 replications of the reconstruction process.	61
3.6	Behavior of the achievable tampering rates for successive iterations of the design procedure on two different images.	62
3.7	Successive descriptors during the design procedure, both without (b-f) and with an importance map (h-k).	63
4.1	Sample images from the <i>sipi</i> , and the <i>bows</i> data-sets	66
4.2	Example fitted distortion model for components v_1 , and v_2	70
4.3	Distribution of the coefficients and a corresponding Lloyd-Max quantization code-book for component v_2	76
4.4	Reconstruction quality improvement due to adoption of Lloyd-Max quantization	78
4.5	Images from the <i>bows</i> data-set with the highest negative impact of replacing the uniform with the Lloyd-Max quantizer.	79
4.6	Sample image blocks from the considered block classes	80
4.7	Histogram of Δ PSNR between the single-profile, and the 3-profile configurations	81
4.8	Reference images for the <i>baboon</i> image from the <i>sipi</i> test set.	86
4.9	Reference images for the <i>lena</i> image from the <i>sipi</i> test set.	88
4.10	Reference images for the <i>peppers</i> image from the <i>sipi</i> test set.	90
4.11	Reference images for the <i>3879</i> image from the <i>bows</i> test set.	92
5.1	Operation of the adaptive self-embedding scheme	96
5.2	Histogram of reconstruction PSNR on the <i>bows</i> data-set.	101
5.3	Reconstruction quality for various configurations of the reference scheme	102
5.4	Trade-off between the tampering rate and the reconstruction fidelity	103
5.5	Example reconstruction of maliciously tampered content	105
5.6	Example test images from the <i>bows</i> data-set.	108
5.7	Reconstruction PSNR for selected test images under varying tampering rates	109
5.8	Reconstruction results for state-of-the-art self-embedding schemes for image <i>6882</i> from the <i>bows</i> data set.	110
5.9	Reconstruction results for state-of-the-art self-embedding schemes for image <i>6882</i> from the <i>bows</i> data set tampered with rate $\tilde{\gamma} = 0.46$.	112
5.10	Reconstruction results for state-of-the-art self-embedding schemes for image <i>9011</i> from the <i>bows</i> data set.	114

5.11	Reconstruction results for state-of-the-art self-embedding schemes for image 131 from the <i>bows</i> data set.	116
5.12	Reconstruction results for state-of-the-art self-embedding schemes for image 4749 from the <i>bows</i> data set.	118
5.13	Reconstruction reference for image 7710 obtained with the reference, and the content-adaptive schemes.	121
5.14	Reconstruction reference for image 7787 obtained with the reference, and the content-adaptive schemes.	122
5.15	Validation of the descriptor design objective based on 1,000 random reconstruction attempts with the descriptor-adaptive scheme. . . .	124
5.16	Quality descriptors obtained for the descriptor-adaptive scheme for oblivious tampering without and with an importance map. . . .	125
5.17	Comparison of important content representation for the reference scheme, and the descriptor-adaptive scheme	126
5.18	Scatter plots of the reconstruction quality vs. the tampering rate for the reference, and the descriptor-adaptive schemes.	127
B.1	Probability of successful recovery of the reference sharing mechanism for different image sizes.	141
C.1	Training images from the <i>sipi</i> data set.	145
C.2	Training images from the <i>bows</i> data set.	146
C.3	Training images from the <i>ucid</i> data set.	147

List of Tables

1.1	Overview of reconstruction performance and restoration methods in state-of-the-art self-embedding schemes	20
1.2	Overview of reconstruction performance and restoration methods in flexible and adaptive self-embedding schemes	21
2.1	Reconstruction success bounds for selected reference rates.	48
4.1	Reconstruction quality for the uniform, and the Lloyd-Max code-books	77
4.2	Quality improvement due to the adoption of Lloyd-Max quantization.	79
4.3	Reconstruction quality improvement due to adoption of three reconstruction profiles for low, medium, and high texture blocks.	82
4.4	Impact of block texture levels on the solutions of the resource allocation problem for the <i>bows</i> data-set, $b = 40$ bpb, $L = 2$, and Lloyd-Max quantization.	83
4.5	Impact of block texture levels on the solutions of the resource allocation problem for the <i>bows</i> data-set, $b = 80$ bpb, $L = 2$, and Lloyd-Max quantization.	84
5.1	Reconstruction profiles for the descriptor-adaptive scheme.	99
5.2	Reconstruction quality for various configurations of the reference self-embedding scheme, measured as the PSNR on the <i>ucid</i> and <i>bows</i> test sets.. . . .	104
5.3	Comparison of the reconstruction performance with state-of-the-art self-embedding schemes.	107
5.4	Reconstruction performance for the content-adaptive scheme.	120

C.1	Summary of the considered data sets.	143
D.1	Solutions to the optimal resource allocation problem for the Lloyd- Max code-book, $L = 3$ and the <i>bows</i> training set.	150
D.2	Solutions to the optimal resource allocation problem for the uniform code-book, $L = 3$ and the <i>bows</i> training set.	152
D.3	Solutions to the optimal resource allocation problem for the Lloyd- Max code-book, $L = 2$ and the <i>bows</i> training set.	154
D.4	Solutions to the optimal resource allocation problem for the Lloyd- Max code-book, $L = 2$ and the <i>bows</i> training set.	156
E.1	Reconstruction success bounds in the presence of false positive classification errors.	160

Abbreviations

bpb	bits per block
bpp	bits per pixel
CRC	cyclic redundancy check
DCT	discrete cosine transform
DC	differential coding
DE	difference expansion
DEMUX	demultiplexer
DWT	discrete wavelet transform
ECC	error correction coding
GGD	generalized Gaussian distribution
HVS	human vision system
JPEG	joint photographic experts group
LSB	least significant bit
LSBS	least significant bit substitution
LSBM	least significant bit matching
MINLP	mixed-integer nonlinear programming
MLE	maximum likelihood estimation
MSE	mean square error

MUX	multiplexer
PSNR	peak signal to noise ratio
PST	pinned sine transform
QIM	quantization index modulation
RLF	random linear fountain
RoI	region of interest
SCS	scalar Costa scheme
SSIM	structural similarity index
VQ	vector quantization

Definitions

<i>allocation matrix</i>	assignment of per-block reference payload b to individual coefficients of a block-based DCT spectrum; defines the precision of coefficient representation
<i>authentication unit</i>	smallest image fragment defining the resolution of tampering localization; in this study, a 8×8 px block
<i>quality descriptor</i>	mapping between image blocks and reconstruction profiles; it controls the reconstruction quality on a per-block basis
<i>reconstruction demand</i>	portion of the reference stream, which is necessary for successful restoration for a given tampering rate
<i>reconstruction profile</i>	specification of a per-block reconstruction setting; defined in terms of allocation matrix and coefficient quantization code-books

<i>principal image content</i>	an embedding-invariant representation of the image content; in this study it refers to a low dynamic range version of the image, composed of a certain number of most significant bit-planes
<i>reconstruction reference</i>	compact representation of the principal content of the original image, which is embedded into the image itself, and used by a decoder for restoration
<i>reference block</i>	portion of the reconstruction reference, which refers to a single image block
<i>reference symbol</i>	portion of the reconstruction reference of length equal to the watermark capacity of a single image block

List of Symbols

\oplus	bit-wise exclusive disjunction (XOR)
$\operatorname{argmin}_x f(x)$	set of values of x for which $f(x)$ attains its lowest value
α	shape parameter, which controls the behavior of the reconstruction demand
$\hat{\alpha}$	empirical estimate of the shape parameter α
α_s	slope of s -th component of piecewise-linear representation of the reconstruction demand
\mathbf{a}	column vector with coefficient group cardinality
β_s	y-intercept of s -th component of piecewise-linear representation of the reconstruction demand
b	number of bits for the description of a single block content; either constant or $b = b(i)$
B	reference symbol length
c_q	scaling factor for outliers identification
$\mathbf{d}(\mathbf{v})$	distortion vector for allocation vector \mathbf{v} , where k -th component represents the MSE for the k -th coefficient group
$d(i, p)$	distortion map for i -th block and p -th reconstruction profile
$D(i)$	degradation map
δ	probability of fountain decoding failure

Δ_s	number of image blocks for degradation in a single design iteration
$\epsilon(\delta)$	random linear fountain code overhead, i.e., the number of additional symbols needed to obtain probability of decoding error δ
E	tampering (erasure) map
$f(\cdot)$	watermark embedding function
$f^{-1}(\cdot)$	watermark extraction function
$g_b(\cdot)$	reconstruction reference generation function, generates exactly b bits of reference information
$g_b^{-1}(\cdot)$	image block reconstruction function
γ	rate of authentic image blocks
$\tilde{\gamma}$	$= 1 - \frac{M}{N}$, tampering rate
$\tilde{\gamma}_{\text{target}}$	target tampering rate for the descriptor design procedure
$\tilde{\gamma}_{\text{min}}$	pessimistic tampering rate bound
$\tilde{\gamma}_{\text{ave}}$	average tampering rate bound
$\tilde{\gamma}_{\text{qd}}$	tampering rate bound for decoding the quality descriptor
\mathbf{G}	generator matrix of a digital fountain code
$h(\cdot)$	one way hash function
$ h $	number of hash bits
$hcf(a, b)$	highest common factor of a and b
H_i	$= h(I_i, i, Y_i, k)$ authentication hash for i -th image block
\hat{H}_i	extracted authentication hash H_i
I	cover image, 8-bit color depth
I_i	i -th block of the image I
$I_{x,y}$	pixel intensity of image I with coordinates x, y
$I^{(w)}$	watermarked image, 8-bit color depth
$I^{(r)}$	final restored image
$I^{(t)}$	tampered image

$I^{(\text{ldr})}$	$= \lfloor I/2^L \rfloor$ low dynamic range version of I
$I^{(\text{ref})}$	image recoverable from the reconstruction reference
$I^{(\text{ref} p)}$	image recoverable from the reconstruction reference for the p -th reconstruction profile
$I^{(\text{ref} \mathbf{V})}$	image recoverable from the reconstruction reference for the p -th reconstruction profile
κ	security context
K	total number of reference blocks
L	number of least significant bit planes used for watermark storage
λ	$= \frac{b}{B}$, reference rate, i.e., the rate of reference data length to available reference payload
λ_{ave}	weighted average reference rate
λ_{qd}	quality descriptor fountain coding rate
$\lambda(i)$	reference rate for the i -th image block
λ_k	k -th highest reference rate in the considered descriptor
Λ	the set of considered reconstruction reference code rates
M	number of authentic image blocks
N	total number of image blocks
N_t	number of training image blocks
N_x	image width
N_y	image height
\mathbf{N}	set of positive natural numbers
P	number of defined reconstruction profiles
\mathbf{p}_g	guard reconstruction profiles in the descriptor design procedure
$\phi(i)$	content importance map for the descriptor design procedure
$\psi_\alpha(\gamma)$	family of functions, which describe the behavior of the reconstruction demand in uniform-quality self-embedding

$q(i)$	quality descriptor
\mathbf{q}	bit-stream of the final quality descriptor used by the encoder
Q_i	i -th input symbol of the quality descriptor
$\rho(\lambda, \gamma)$	reconstruction demand
$\rho(\tilde{\gamma} \Lambda, \mathbf{w})$	reconstruction demand for adaptive self-embedding with a set of reference rates Λ with weights \mathbf{w}
\mathbf{r}	$= r_1, \dots, r_N$, reconstruction reference bit-stream
σ_i	standard deviation of pixel intensities in the i -th image block
S'	number of unique reference code rates
S	number of unique reference code rates, which are used in the considered quality descriptor
T	DCT transformation matrix
$\theta(\mathbf{v})$	objective distortion for given coefficient precision specified by means of the allocation vector \mathbf{v}
θ_i	average distortion for DCT coefficients in the i -th group
$\vartheta_{l,i}$	distortion model parameters
u_k	helper stub variable corresponding to component v_k of the allocation vector; used in an extended version of the resource allocation problem
$U(a, b)$	uniform probability distribution in range (a, b)
v_{\min}	minimum possible coefficient precision
\mathbf{V}	reference payload allocation matrix
\mathbf{v}	reference payload allocation vector
\mathbf{v}^+	non-negative components of \mathbf{v}
\mathbf{v}_{lin}	\mathbf{v} stemming from optimization with log-linear distortion model, starting from uniform \mathbf{v}_0
\mathbf{v}_{exp}	\mathbf{v} stemming from optimization with log-exponential distortion model, starting from uniform \mathbf{v}_0

\mathbf{v}_{ref}	\mathbf{v} stemming from optimization with log-exponential distortion model, starting from \mathbf{v}_{lin}
w_l	the weight of the l -th reference rate in the quality descriptor
W_i	embedding symbol for adaptive self-embedding, a multiplex of Y_i and Z_i
\hat{W}_i	extracted embedding symbol W_i
X_k	k -th reference symbol
$\{X_k\}$	set of all reference symbols
Y_i	embedding symbol for i -th image block
Z_i	i -th output symbol of a fountain-encoded quality descriptor
\mathbf{Z}	set of all integers

Introduction

With modern software it is now easier and cheaper than ever before to counterfeit digital documents, images, or any other data sets. The tools for adding or removing objects from digital images or videos are within reach of an average mainstream consumer. Features similar to *Content Aware Fill* [2] from Adobe® Photoshop®, are currently at our fingertips, implemented as applications on our smart-phones or tablets. An example application with this functionality, *Touch Retouch* [4], is available for all major mobile platforms.

The problem of authenticating the content of digital images can be approached from two perspectives, i.e., via *passive* forensic or *pro-active* protection techniques¹. The former involve sophisticated algorithms which analyze statistical properties of image features in search for typical inconsistencies resulting from content forgeries. Cues like lighting inconsistencies, transform-domain quantization artifacts, or sensor noise discrepancies can be exploited to detect counterfeit content. Unlike pro-active protection, passive techniques do not require the images to be previously prepared.

While not always applicable in practice, the ability to prepare an image at an early stage gives pro-active techniques an edge with respect to their authentication potential. The protection should be applied as close to the source of information as possible. Hence, such algorithms are sometimes integrated into digital cameras or close-circuit television cameras to serve as an additional protection mechanism for prospective key evidence [3]. Pro-active authentication schemes operate by embedding additional information into digital images by means of digital watermarks. The embedding process involves imperceptible modifications of the

¹Pro-active digital forensics is an emerging research and standardization direction not only for multimedia content [53]. Early standardization efforts are being seen for instance in the payment card industry, or to prevent misuse of public IT facilities. Regardless of the field, the goal is always to maximize the usage of digital evidence of forensic value.

content, often in the form of an additional noise-like component, somewhat similar to the sensor noise.

Proper design of the embedded information allows not only for efficient localization of the tampered fragments, but also for discrimination between allowed and disallowed content modifications [23], or even restoration of the original content of the tampered regions [24]. The reconstruction capability is one of the most compelling features of digital image authentication systems. In addition to content hashes for authentication purposes, an encoder embeds in the image a *reconstruction reference* which is used by a dedicated decoder to restore the tampered image fragments. The reconstruction reference is often a low-quality version of the original image. Hence the term *self-embedding* (also referred to as *self-recovery*), coined in the original publication [24].

A number of content reconstruction schemes have been proposed with various applications in mind, ranging from content authentication [24, 71, 77], through error concealment [6, 25, 67] to reversible privacy protection [37, 46]. Regardless of the application at hand, the reference information needs to be communicated to a decoder through an unreliable channel, i.e., the tampered digital image. Hence, designing a good self-embedding scheme is equivalent to designing an efficient communication architecture for the reconstruction reference.

Such a communication problem differs from typical channel models. Only certain fragments of the transmitted message need to be decoded for the process to be successful. The exact locations, and the amount of necessary message fragments depend on the observed conditions in the channel. The ability to adopt existing tools from communication theory to model and analyze the reconstruction performance is the main focus of this dissertation.

Scope and Thesis

This dissertation deals with digital image protection techniques based on self-embedding and digital watermarking. The content reconstruction problem is modeled as a communication over an erasure channel. The following thesis will be supported by both theoretical analysis and experimental verification:

It is possible to model the content reconstruction problem as an erasure communication channel. Such a model allows for accurate formal analysis of the success bounds, and the inherent restoration trade-offs. A self-embedding scheme based on the proposed communication model delivers superior performance, compared to state-of-the-art reconstruction algorithms. High-quality reconstruction is possible even under extensive tampering.

Based on the proposed communication model, I design a highly customizable content reconstruction method, which serves as a basis for three example self-embedding schemes. The schemes are implemented with the use of digital fountain codes, and are evaluated with respect to common performance measures. A benchmark configuration of a reference scheme is compared to 5 state-of-the-art alternatives in a common evaluation framework.

The dissertation is organized as follows. Chapter 1 introduces the area of research. It describes the relevant research problems, the currently known approaches, models, and techniques. It explains how the proposed solution fits in the history of the problem, and the improvement that can be achieved with its adoption. This chapter also addresses the limitations of existing schemes, and concludes with a brief discussion of the remaining challenges and the emerging research directions.

Theoretical foundations for the proposed content reconstruction model are presented in Chapter 2. It gives a formal statement of the problem, and emphasizes the differences with respect to traditional erasure communication. Based on the proposed model, I derive the reconstruction success bounds. The obtained theoretical results are validated experimentally by Monte Carlo simulations. Initially the analysis considers the most commonly used scenario, where the reconstruction fidelity is assumed to be identical for all image fragments. Extension to an adaptive variant, where the reconstruction quality can be controlled on a per-block basis is presented in Chapter 3.

The focus of both Chapter 2 and Chapter 3 is on the reconstruction success bounds. Techniques for optimization of the reconstruction quality are discussed in Chapter 4, which deals with the design of quantization code-books, and optimal resource allocation for generating the reconstruction reference. Content adaptivity is addressed from the image quality perspective.

Based on the derived reconstruction model, and the discussed optimization techniques, Chapter 5 presents the design of practical schemes, with emphasis on adaptive self-embedding. A reference scheme in a benchmark configuration is experimentally evaluated, and compared to state-of-the-art alternatives. Further evaluation addresses the efficiency of the trade-off between the reconstruction quality and the achievable tampering rates. The dissertation concludes with a summary of the proposed contributions, and the perspectives for further research.

The dissertation includes five appendices. Appendix A discusses popular attacks and security issues that arise when implementing self-embedding functionality in real-world systems. Most of the vulnerabilities are related to the content authentication phase, which directly precedes content reconstruction. Appendix B provides a detailed theoretical comparison of the asymptotic behavior of the proposed model, and the reference sharing mechanism from [76]. The performed analysis shows that the two methods are essentially different approaches to prac-

tical implementation of the same high-level paradigm. Appendix C describes the test sets of digital images which were used during the experiments. The penultimate Appendix D collects selected solutions to the coefficient precision optimization problem described in Chapter 4. The final Appendix briefly discusses miscellaneous practical implementation issues, including the handling of color images, and the impact of errors on the reconstruction performance. The latter is of principal importance for dealing with lossy-compressed image formats.

For the sake of research reproducibility, a complete implementation of the proposed contributions in the Matlab environment can be obtained at <http://kt.agh.edu.pl/~korus>. An interactive demonstration of the developed self-embedding system is available along with the experimental evaluation scripts.

Publications Related to the Dissertation

Selected research results have been presented on international conferences and in international journals. The list of relevant publications, in chronological order, is as follows:

- [37] Paweł Korus, Wojciech Szmuc, and Andrzej Dziech. A scheme for censorship of sensitive image content with high-quality reconstruction ability. In *Proc. of IEEE International Conference on Multimedia and Expo*, Singapore, 2010. doi: 10.1109/ICME.2010.5583410
- [39] Paweł Korus, Lucjan Janowski, and Piotr Romaniak. Automatic quality control of digital image content reconstruction schemes. In *Proc. of IEEE International Conference on Multimedia and Expo*, Barcelona, 2011. doi: 10.1109/ICME.2011.6011872
- [34] Paweł Korus and Andrzej Dziech. A novel approach to adaptive image authentication. In *Proc. of IEEE International Conference on Image Processing*, Brussels, 2011. doi: 10.1109/ICIP.2011.6116243
- [38] Paweł Korus, Jarosław Białas, Piotr Olech, and Andrzej Dziech. A high-capacity annotation watermarking scheme. In *Multimedia Communications, Services and Security*, volume 149 of *Communications in Computer and Information Science*, pages 1–9. Springer Berlin Heidelberg, 2011. ISBN 978-3-642-21512-4. doi: 10.1007/978-3-642-21512-4_1
- [40] Paweł Korus, Jarosław Białas, and Andrzej Dziech. A new approach to high-capacity annotation watermarking based on digital fountain codes. *Multimedia Tools and Applications*, pages 1–19, 2012. ISSN 1380-7501. doi: 10.1007/s11042-011-0986-8

- [35] Paweł Korus and Andrzej Dziech. Reconfigurable self-embedding with high quality restoration under extensive tampering. In *Proc. of IEEE International Conference on Image Processing*, Orlando, FL, 2012. doi: 10.1109/ICIP.2012.6467329
- [36] Paweł Korus and Andrzej Dziech. Efficient method for content reconstruction with self-embedding. *IEEE Transactions on Image Processing*, 22(3):1134–1147, March 2013. doi: 10.1109/TIP.2012.2227769

The paper [37] presents a reversible privacy protection scheme which features high-quality reconstruction. Selected regions of interest (RoIs) are blurred, and the reference information for their reconstruction is retained in the image by means of a digital watermark. The scheme operates in the discrete wavelet transform (DWT) domain, and is robust against JPEG2000 compression. The paper also briefly discusses the reconstruction quality related trade-offs in the application at hand. A detailed analysis follows in [39]. The paper proposes a prediction model for automatic selection of system parameters which maximize the overall image quality. The goal is to balance the reconstruction- and embedding-inflicted distortions.

Reversible privacy protection can also be seen as a generalization of adaptive self-embedding, where irrelevant background content is excluded from the reconstruction process. The paper [34] proposes an adaptive content reconstruction scheme, which defines several quality levels, and uses a quality descriptor to map the levels to image blocks. By controlling the reconstruction quality for each image block individually, it is possible to bias the scheme either towards better quality or greater tampering rates. The descriptor takes into account both local characteristics of the image content, and user-defined requirements. The presented concepts are extended in the dissertation in Chapter 3 and Chapter 4.

The next two papers, [38] and [40], describe an annotation watermarking scheme based on digital fountain codes. The former is a conference paper, which has been extended and published in a journal as the latter. The papers discuss the possibility of adopting the erasure communication model for embedding multiple annotations, associated with selected polygons in the image. The annotations are multiplexed in a carefully designed way which ensures robustness against cropping. The scheme is also robust against lossy JPEG compression.

The proposed self-recovery model is described in [36]. It is explained why such a model is a good fit for the problem, and how it can be implemented in practice using digital fountain codes. The presented theoretical analysis focuses on the applicable success bounds, and gives new insights into the inherent restoration trade-offs. Based on this model, [35] presents a customizable self-embedding scheme, which is then experimentally evaluated in a number of possible configurations. The

dissertation extends the performed analysis, and describes the model in detail in Chapter 2.

A patent claim *Sposób i układ do zabezpieczania dostępu do wrażliwych treści obrazów cyfrowych*, prepared by Andrzej Dziech, Andrzej Głowacz, Paweł Korus, and Wojciech Szmuc, has been submitted to the Polish patent office on 16 June 2011; the claim is filed under P-395 303.

The goal of this chapter is to survey both the applications, and algorithmic solutions for self-embedding systems. Different formulations of the reconstruction problem are presented, and classified according to the origin and use of the reference information. The chapter ends with a discussion of the limitations of state-of-the-art schemes, and the emerging research directions.

1.1 Media Protection with Self-Embedding

The focus of this dissertation is on pro-active protection techniques where the images can be prepared at an early stage of the information flow. A self-embedding system involves two entities: an *encoder* and a *decoder*, which implement the pre-processing and the reconstruction procedures, respectively. Fig. 1.1 shows a conceptual application of the protection framework. The encoder prepares a protected image given the input image, and a security context, within which the protected image should be considered as valid. The context information typically includes the date and time when the image was captured, and a secret key shared between the encoder and the decoder.

Prospective intermediate processing, which occurs between the encoder and the decoder is referred to as *tampering*, as the encoder is intended to yield final versions of the images. In the presented example, the tampering involves removal of the crashed car from the background. Given a protected image, and the security context, the decoder identifies the fragments of the image that have been tampered with, and recovers the original appearance of the modified content.

Operation of the encoder involves embedding additional information by means of digital watermarking, so that it could be exploited by the decoder to aid the reconstruction process. Digital watermarking is the enabling technology for the considered protection mechanism. It allows for encoding information in digital

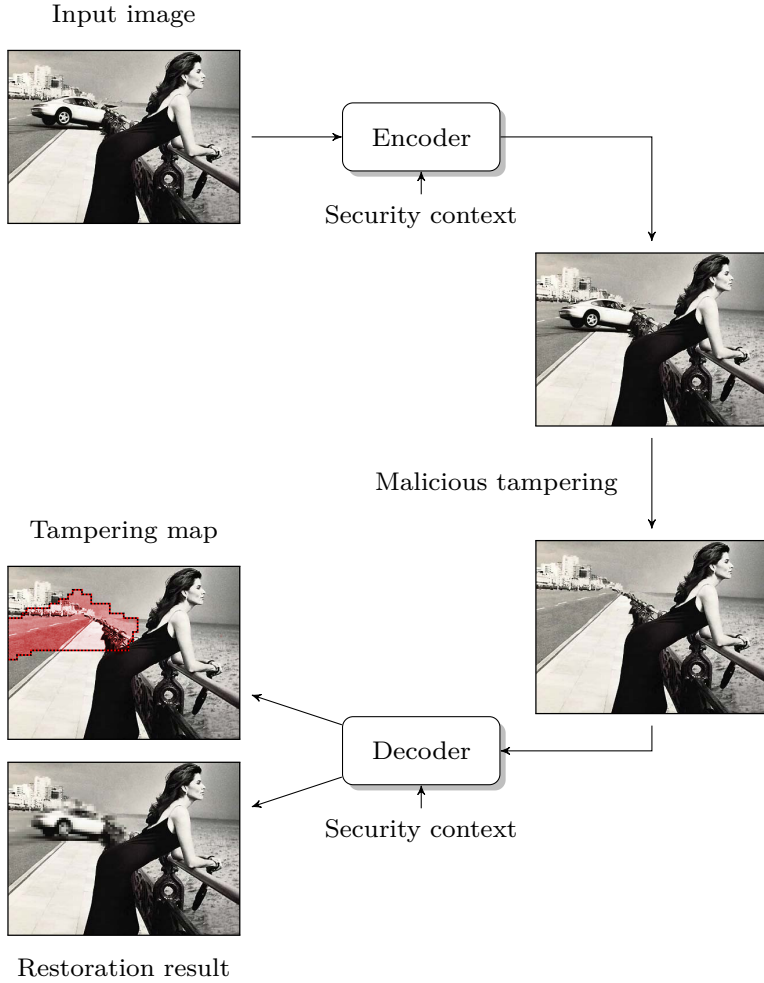


Fig. 1.1: Pro-active protection of digital images with self-embedding; the decoder allows for recovery of the crashed car, maliciously removed from the image.

signals by introducing imperceptible modifications of their samples. As a result, such information is typically more persistent than traditional meta-data supported by various image formats. Additionally, it can be accessed only with the use of a dedicated decoder.

The adopted embedding techniques fall in the class of *fragile watermarking*, intended for immediate destruction upon image modification. Once an image

fragment is maliciously tampered, its corresponding portion of the watermark instantly becomes void. In addition to the reference information, the encoder embeds in the image all necessary headers, e.g., specification of the applicable reconstruction methods, or restoration settings for individual image fragments. The encoder and the decoder typically also share some auxiliary information, which could include selected reconstruction parameters, encryption keys or settings, etc.

The described pro-active protection of digital images is used in three main applications: content authentication, reversible privacy protection, and error concealment, which slightly differ with respect to their individual requirements. Content authentication is the most common use for self-embedding [13, 24, 33, 34, 44, 55, 57, 71–77]. The images are divided into smaller fragments, e.g., blocks, which serve as authentication units. Verification of cryptographic hashes of their content allows for tampering localization. The goal of the decoder is twofold. First, it verifies the integrity of the image content, and yields a tampering map which assigns individual image blocks to either *authentic* or *tampered* states. Secondly, provided that sufficient amount of reference information can be extracted from authentic image fragments, the decoder will restore an approximate original content of the tampered regions.

The second application, privacy protection, uses content redaction techniques to limit access to individual regions of interest (RoI) of digital images [69, 70]. In reversible privacy protection schemes, the embedded reference information allows the decoder to restore their original appearance [37, 45, 46]. Authorized recipients can access the original content, which is restored for them on-demand on their personal devices. During transmission or storage, the protected images remain censored and safe for public distribution.

In reversible privacy protection, the tampering is a fully controlled operation in the information flow, and the affected regions are known from the beginning. Hence, content authentication is expendable and can be replaced with a binary mapping specifying the fragments for reconstruction, and a simple cyclic redundancy check (CRC) for watermark integrity verification. Compared to content authentication, this application usually requires higher restoration fidelity.

Alternative approaches also exist, where the RoIs in the image are simply encrypted with common cryptographic ciphers. However, the resulting redaction pattern is visually disturbing and automatically raises suspicion that some encryption mechanisms might be in use. Moreover, it has been shown that such an approach raises additional security concerns, resulting from the character of multimedia content [58]. Similar functionality can also be provided directly by image formats, e.g., JPEG2000 is capable of delivering various quality, or various image size for different users [22]. A possible use cases is to provide public access to the thumbnail, and limited access to a full scale image, e.g., upon payment.

The last of the applications, error concealment, is most commonly implemented

without the use of self-embedding. Multimedia transmission problems are detected on different levels, e.g., the network packet level, or the bit-stream syntax level. As a result, certain image fragments might turn out as damaged. The reconstruction usually takes form of interpolation, or inpainting. Adoption of additional reference information can significantly improve the efficiency of such schemes [16, 67]. In error concealment content authentication is obsolete, and it suffices to verify the integrity of the content with simple CRC codes.

1.2 Measuring the Reconstruction Performance

Performance of content reconstruction systems involves two principal aspects: the quality of the produced images, and the conditions of restoration. The quality is measured both from the perspective of watermarking-inflicted artifacts, and the reconstruction fidelity.

For two 8-bit gray-scale images of size $N_x \times N_y$ px, the distortion level in the image $I^{(w)}$, with respect to the original image I , is expressed by the peak signal to noise ratio (PSNR) [dB]:

$$\text{PSNR}(I, I^{(w)}) = 20 \cdot \log_{10} \frac{255}{\sqrt{\text{MSE}(I, I^{(w)})}}, \quad (1.1)$$

where MSE is the mean squared error calculated as:

$$\text{MSE}(I, I^{(w)}) = \frac{1}{N_x N_y} \sum_{x=1}^{N_x} \sum_{y=1}^{N_y} (I_{x,y} - I_{x,y}^{(w)})^2. \quad (1.2)$$

where $I_{x,y}$ denotes the pixel intensity of image I at coordinates (x, y) , and N_x and N_y represent image width and height, respectively.

The restoration conditions depend on the adopted formulation of the reconstruction problem. In many schemes, the reference information regarding a particular image block (i -th) is simply embedded into a different block (j -th). As a result, the i -th block can be recovered only if j -th is still authentic. Some schemes attempt to mitigate this *reconstruction dependency* problem by embedding multiple copies of the reference information [44, 68], or eventually using interpolation, or inpainting when the primary reconstruction attempt fails [68]. In such schemes it is not possible to express the robustness to tampering by a single concise parameter.

It is surprising that this problems cripples even recent schemes [44, 68], as such an approach is not only suboptimal, but also suffers from serious security vulnerabilities. In early self-embedding schemes, the embedding locations for the reference information were established with the use of a constant shift vector in the modulo arithmetic, which makes them susceptible to the synchronous

counterfeiting attack [29]. In more recent schemes the locations are typically selected pseudo-randomly, with the use of a chaotic system, or with a non-linear mapping. Selection of the best embedding locations for the reference information of individual image regions has also attracted attention as an independent research topic [30, 32].

The reconstruction dependencies can be eliminated by distributing the reference information of every single block over the entire image. This high-level concept has been recently implemented in practice in [73, 74]. In such schemes, the restoration condition is usually a threshold on *tampering rate* $\tilde{\gamma}$, i.e., the fraction of the tampered authentication units. When considering the tampering rate, it needs to be kept in mind, that this parameter is not necessarily equal to the actual extent of modifications. For instance, if a single pixel in every authentication unit is changed, the actual rate of modifications of the original data stream is significantly lower than the tampering rate, which equals 100%. However, in practice, malicious modifications often involve either adding or removing certain objects from the scene. Then the tampering rate is a good approximation of the actual extent of modifications. The approximation becomes more accurate for smaller image blocks. Very small blocks, e.g., 1×1 px, or 2×2 px, are actually discouraged due to unacceptable security vulnerabilities [17, 31]. Intermediate sizes, like 8×8 px blocks are usually a good compromise between the security and the tampering localization capability.

In addition to the reconstruction dependencies, another major problem with the design of efficient self-embedding schemes is the *reference waste*. Since it is not possible to determine *a priori* which regions will be tampered, the remaining portion of the reference stream eventually turns out as unnecessary, and contributes to the waste of the watermark's capacity. The problem can be mitigated by reusing the remaining authentic principal content in the tampered image.

There is a trade-off between the tampering rate, the watermarking-inflicted distortion, and the reconstruction fidelity. Initially, self-embedding schemes were designed without any explicit requirements towards the reconstruction quality. It was sufficient for the restoration result to be discernible for a human observer. It has recently been noticed that in certain applications the restoration fidelity is of paramount importance. The behavior of the restoration trade-offs has not received sufficient attention yet. Recently however, the issue has been addressed from two different perspectives.

The emerging *flexible* self-embedding schemes, adopt an approach where the reconstruction quality systematically deteriorates with the increasing tampering rate [74–76]. Such schemes allow for high-quality reconstruction, if the observed tampering affects small image areas. With the increasing tampering rate, the reconstruction fidelity deteriorates either gradually [75, 76], or through several levels [74].

An alternative is to design adaptive schemes [33, 34, 56, 57]. In this approach, the reconstruction quality is controlled for each image fragment individually, and does not change with the tampering rate. Typically, such a scheme defines several reconstruction profiles, characterized by different fidelity, or dedicated to different types of content. As a result, individual fragments can be represented in a way which best matches its content.

Suitability of either the flexible, or the adaptive approach depends on the application at hand. While the flexible schemes can guarantee high-quality reconstruction for low tampering rates, the adaptive schemes can guarantee high-quality reconstruction for selected image fragments. Hence, adaptive self-embedding is particularly well suited for video surveillance applications. Fig. 1.2 shows an example recovery of a tampered footage. The original, and the watermarked images are shown in (a) and (b), respectively. A content importance map is shown in (c). Important details of the captured scene, i.e., the license number plates, and the wind-screen of the frontal car, are recovered with high-quality (e). Low-priority background content is restored either with low quality, or not restored at all. The forged image, and the locations of malicious tampering are shown in (d) and (f), respectively.

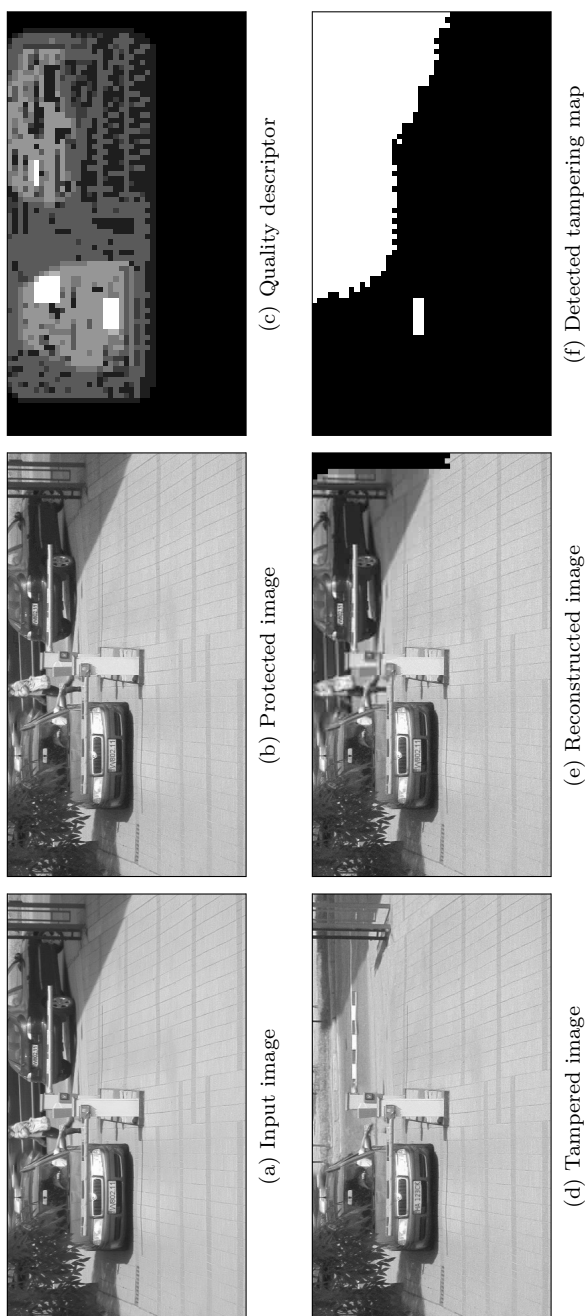


Fig. 1.2: Example recovery of a tampered video surveillance footage; adaptive self-embedding is used to guarantee high-quality reconstruction of important details.

1.3 Related Work

This section presents an exhaustive review of the current state-of-the-art in self-embedding systems. The description begins with a brief introduction of the fundamental techniques, and approaches. Then, it highlights some alternative reference information representations. Finally, it discusses the emerging research directions of flexible, and adaptive self-embedding.

1.3.1 Fundamental Reference Representations

In the most common approach, the reconstruction reference is generated as a reduced-quality version of the original image. The reference information can be extracted either in the spatial, or in a transform domain, where it is common to adopt traditional techniques from transform image coding. The image is divided into non-overlapping blocks, and the coefficients of a block-based transform are quantized according to the given payload constraints. Such an approach is adopted in the original publication on self-embedding [24], where the process resembles standard JPEG compression. Numerous later schemes, also adopt a similar approach, e.g., [33, 73].

The reference generation can also be constructed in a lossy manner directly in the spatial domain, e.g., with the use of vector quantization (VQ) [68]. In this approach, the reference information for a single image block consists of an identifier which points to a dictionary of vectors. The dictionary is obtained during a training phase, and represents a variety of possible block content. The reconstruction fidelity will directly depend on the size of the dictionary, which needs to be shared between the encoder and the decoder. The scheme [68] embeds the identifiers in random positions in the image, and suffers from the reconstruction dependency problem. In order to mitigate this issue, the scheme tries to estimate the content of unrecoverable blocks based on their neighbors. As a result, there is no explicit maximal tampering rate limit. As the amount of modified content increases, the restoration fidelity drops to the order of 15 dB, and the result becomes practically indiscernible.

When highest possible reconstruction fidelity is needed, a self-embedding system should adopt a reversible watermarking technique. The scheme from [71] aims at loss-less reconstruction with bit-level accuracy. The reconstruction reference is constructed from a complete 8-bit deep pixel intensity information, and the watermark is embedded using difference expansion (DE) [62]. The resulting watermarked images are highly distorted, with PSNR of approximately 29 dB, but the original image can be perfectly recovered by the decoder. The main limitation of this scheme is the maximum tampering rate of 3.2%.

In [12, 13] the authors use a dithered binary version of the original image as a reconstruction reference. During restoration, the decoder extracts this binary

reference, and applies inverse halftoning to obtain the final reconstruction result. The watermark is embedded in the discrete wavelet transform (DWT) domain with the use of bit substitution. Prospective tampering of the image content leads to bit recovery errors, which gradually destroy the reference image by introducing random noise instead of the content. Hence, growing amount of tampered image fragments lowers the reconstruction quality at first, but quickly renders the reconstruction result indiscernible. Similarly to [77], the maximum feasible tampering rate stems from the reconstruction fidelity drop. No specific values are reported, though.

Some of the techniques adopt exhaustive search for content reconstruction [7–9, 72]. The methods presented in [7–9] embed two watermarks: a block-based watermark for tampering localization, and a pixel-based watermark for both reconstruction, and tampering localization enhancement. The latter is constructed as a cryptographic hash of most significant bits of randomly permuted image pixels. The process may be repeated with a different permutation seed to further improve the reconstruction capability. The reconstruction is performed in an iterative manner, and the decoder considers all feasible corrections until the hash value is correct.

The achievable performance varies between the schemes, but the best results are obtained under low tampering rates. Depending on the number of embedded pixel-based watermarks, the scheme [7] performs a one- or two-step reconstruction. The restoration capability is highly dependent on the amount of observed tampering within the image blocks. Provided that the reconstruction is meaningful up to 50% of successfully recovered tampered pixels, the scheme can survive tampering rates up to 10%. More recent schemes [8, 9] improve upon the original reconstruction procedure, and the process is now performed in multiple iterations. This results in an increase of the supported tampering rates, with meaningful restoration results up to 32% and 24% of the image area.

An important unique feature of the schemes [7–9] is their robustness against cropping. This functionality is integrated with the block-based tampering localization component, and is independent from the content recovery capability.

1.3.2 Alternative Reference Representations

In alternative reference representations, the reference information does not have a direct interpretation of an image. In [42, 43] it is constructed from the redundancy provided by traditional error correction codes. A certain number of the most significant bit-planes is considered as payload and remains unchanged. The least significant bit-planes are replaced by the generated redundancy. Hence, the reconstruction process resembles a distributed source coding problem [61, 66]. The authors do not report the maximum supported tampering rate.

In [77], content reconstruction is modeled as an irregular sampling problem, and the restoration is performed by iterative projections onto convex sets. The reference data is obtained by logical exclusive disjunction on cosine transform coefficient polarity information and pseudo-random bit sequences. It is embedded in the image by modulating middle frequency components of the pinned sine transform. The scheme operates on sub-blocks and macro-blocks of the images, used for authentication and restoration, respectively. The method is an example of *semi-fragile* watermarking, as it allows for certain classes of image processing operations after embedding the watermark. The scheme is robust against lossy JPEG compression, Gaussian filtering, unsharpening, and contrast changes. The main limitations are the low reconstruction fidelity, and low tampering rates. Provided that the tampered areas are sufficiently small, and that no global attacks are present, the resulting distortion in the restored regions can reach 27 dB. The typical fidelity varies between 12 dB and 24 dB. An exact value for the maximum tampering rate is not reported.

The reconstruction reference can also be constructed with the use of fractal coding [65]. The image is divided into non-overlapping range blocks, and their reference information is obtained as a parametrized transformation of one of overlapping domain blocks. In the described scheme, only selected RoIs are recovered with the use of fractal coding, and the achievable reconstruction quality reaches 41 dB. The remaining background content is restored with the use of inpainting, which is not a suitable choice for authentication purposes. The maximal tampering rates are not reported.

1.3.3 Flexible Self-Embedding

One of the key contributions of the recent self-embedding research was to adopt reference information spreading mechanisms, which distribute the reference information over the image. Such an approach eliminates the reconstruction dependencies. In [73], the authors address this problem by grouping image blocks in pairs, randomly scattered over the whole image. Then, they quantize the coefficients of a block-based DCT, and extract a 54-bit representation of each image block. A pseudo-random linear projection of the resulting 108 bits yields 320 reference bits, which are then scattered over the image. Hence, the scheme performs local spreading, and global scattering of the reference information. The scheme uses 3 least significant bit-planes to embed the watermark, which consists of 32 bits of authentication information, and 160 bits of reference payload in a single image block. The reconstruction is possible, if the number of uncertain reference bits within each block pair is below a certain threshold, determined with the use of the binomial distribution. The resulting maximum tampering rate is 59%. The

typical reconstruction quality is approximately 28 dB, while the quality of the watermarked images is 37.9 dB.

The authors extended the concept in [76], where an analogous mechanism is used to construct two self-embedding schemes: (A) with a constant, and (B) with a flexible restoration fidelity. In the former, the reference information is obtained from randomly ordered 5 most significant bit-planes by extracting the individual bits of pixel intensity representation. The latter uses a pyramidal decomposition of the image blocks, and defines a three-part scalable reference stream. The remaining least significant bits are used for watermark embedding, which results in the embedding distortion of 37.9 dB.

In both schemes, the stream is further divided into subsets, which are locally spread using pseudo-random binary matrices. Concatenated reference information from all of the subsets is then scattered over the whole image. In the flexible scheme, the reconstruction quality exhibits distinct levels, depending on the extractable parts of the reference stream. Depending on the tampering rate, either all of the reference levels, or just the most coarse of them will be successfully recovered. The applicable tampering rate thresholds are 24%, 48%, and 72% for the successive parts of the scalable stream. Scheme A uses only the first of the three levels, and the expected reconstruction quality is 40.7 dB. In scheme B, the reconstruction quality averages to 32 dB for low tampering rates, and drops down to around 25 dB for the higher ones.

In [74], the authors propose a self-embedding scheme, which uses compressive sensing for content restoration [20]. The algorithm involves a similar reference distribution mechanism as [76]; the pseudo-random projection is performed using Gaussian matrices, and is applied to block-based DCT coefficients of 16 randomly scattered image blocks. The resulting samples are fully democratic in the sense that they all carry information about the whole block group [19]. During content reconstruction, the necessary DCT coefficients are recovered either by compositive reconstruction or compressive sensing, depending on the resulting problem being over- or under-determined. The reconstruction fidelity drops with the number of samples, i.e., with increasing tampering rate.

The watermark is embedded in 3 least significant bit-planes, and the remaining 5 planes are used to obtain the reference information. The expected watermarking-inflicted distortion is 37.9 dB. The maximum achievable reconstruction fidelity averages to 38 dB for low tampering rates, and gradually drops to approximately 28 dB for the tampering rate of 40%. The maximum tampering rate is 60%.

In the flexible reconstruction system from [75], 5 most significant bits of two randomly selected pixels are combined with exclusive disjunction, then grouped with other pixel pairs and embedded into 3 least significant bit-planes of randomly selected image blocks. Depending on the authenticity of individual pixels, the combined 5 bits of a pixel pair can be either fully or partially recovered. The

remaining uncertainty is resolved by exploiting local pixel correlations. The maximum achievable restoration fidelity stems from the number of the most significant bit-planes retained in the reconstruction reference, and reaches 40 dB. The quality decreases with the growing tampering rate; on average it drops below 30 dB for tampering rates above 40%. The maximum tampering rate is 54%.

1.3.4 Adaptive Self-Embedding

In certain applications, it is desirable to control the reconstruction fidelity on a per-block basis. For instance, in video surveillance, it is of paramount importance to reconstruct selected details with higher fidelity than the background. Examples of such high-priority content might include license number plates, human faces, etc. This motivates the emerging research on adaptive self-recovery [15, 33, 34].

Most commonly, adaptive self-embedding is seen as a technique for reducing the length of the reference bit-stream by allowing flat image blocks to be represented by shorter streams [54, 56]. Some of the profiles can have a marker role, which designates their blocks for reconstruction with a different method, e.g., using inpainting [54], or excludes them from the reconstruction process [34]. The resulting shorter bit-stream is essentially a means of improving the achievable tampering rates.

In my opinion, the primary benefit of employing adaptivity features is the additional degree of freedom, which allows to precisely control the inherent restoration trade-offs. The scheme from [34] defines several quality levels, which are mapped to individual image blocks with a quality descriptor. By controlling the reconstruction quality for each image block individually, it is possible to bias the scheme either towards better quality or greater tampering rates. The descriptor takes into account both local characteristics of individual image blocks, and user-defined requirements.

The resulting variable-length reconstruction reference is encoded into a constant-length payload with the use of the LT code [50]. This allows for full exploitation of the watermarking capacity. Additionally, the LT code spreads the information about each image fragment over the whole image. The scheme uses 3 least significant bit-planes for watermark embedding, which results in the embedding-inflicted distortion of 37.9 dB. The maximum achievable reconstruction quality reaches 37 dB, and the maximum tampering rate reaches 72%. This scheme fails to exploit the remaining authentic image content, and is limited by traditional bounds of general erasure communication.

1.3.5 Self-Embedding Summary

A summary of the approaches, and the achievable reconstruction performance for state-of-the-art self-embedding systems, is collected in Table 1.1 and Table 1.2. The

former addresses the schemes with a uniformly distributed reconstruction quality. The latter addresses the schemes with either flexible or adaptive reconstruction.

The originally reported restoration fidelity results cannot be directly compared. They were obtained on different test images, and in different conditions, e.g., for differently selected regions for modification. Hence, this summary contains fuzzy distortion scores, which stem from the typical restoration fidelity, expressed by PSNR as follows:

$$\text{Fuzzy restoration fidelity} = \begin{cases} \textit{Low}, & \text{PSNR} < 30 \text{ dB}, \\ \textit{Medium}, & 30 \leq \text{PSNR} < 35 \text{ dB}, \\ \textit{High}, & 35 \leq \text{PSNR} < 40 \text{ dB}, \\ \textit{V. High}, & \text{PSNR} \geq 40 \text{ dB}, \\ \textit{Loss-less}, & \text{PSNR} = \infty \text{ dB}. \end{cases}$$

An exhaustive experimental evaluation of the state-of-the-art self-embedding schemes, performed in a common scenario on a representative test set of natural images, is presented in Chapter 5.

Table 1.1: Overview of reconstruction performance and restoration methods in state-of-the-art self-embedding schemes: AR - authentic content reused.

Scheme	Embedding Method	PSNR	Restor. fidelity	Tamp. rate	Reference rate	Restoration approach	Reference data	Spreading approach	AR
Schemes with uniform reconstruction quality									
[24] - I	LSBS	44 dB	Low	N/A	N/A	read permuted data	T/8 MSB	N/A	N
[24] - II	DC	33 dB	Low	N/A	4 bpp	read permuted data	P/4 MSB	N/A	N
[71]	DE	29 dB	Loss-less	3.2	8 bpp	solve linear systems for random content groups	P/8 MSB	local-spreading, global-scatter	Y
[73]	LSBS	38 dB	Low	59	54 bpb	solve linear systems for random content groups	T/5 MSB	local-spreading, global-scatter	Y
[72]	LSBS	38 dB	V. High	6.6	320 bpb	brute-force MSB search until hash valid	P/5 MSB	local-spreading, global-scatter	N
[77]	PST ¹	35 dB	Low	N/A	N/A	irregular sampling, projection on convex sets	T/8 MSB	N/A	N/A
[13]	DWT ²	42 dB	Low	N/A	1 bpp	inverse halftone of binary watermark	1-bit dither	N/A	N
[76] - A	LSBS	38 dB	V. High	24	320 bpb	solve linear systems for random content groups	P/5 MSB	local-spreading, global-scatter	Y
[7], $u = 3$	LSBS	38 dB	N/A ³	10^4	2 bpp	exhaustive search until hash valid	P/5 MSB	two-level random permutation	N/A
[8]	LSBS	44 dB	Low	33	1 bpp	exhaustive search until hash valid	P/6 MSB	iterative random permutation	N/A
[9]	LSBS	38 dB	V. High	23	2 bpp	exhaustive search until hash valid	P/5 MSB	iterative random permutation	N/A
Different configurations of the proposed scheme									
$\lambda = 1$	LSBS	38 dB	High	50	160 bpb	self-recovery, fountain coding	T/5 MSB	global spreading	Y
$\lambda = 2$	LSBS	38 dB	V. High	33	320 bpb	self-recovery, fountain coding	T/5 MSB	global spreading	Y

¹ custom quantization in PST domain

² bit substitution in DWT domain

³ reported reconstruction performance includes the percentage of the recovered pixels only

⁴ for at least 50% of correctly recovered pixels

Table 1.2: Overview of reconstruction performance and restoration methods in flexible and adaptive self-embedding schemes: AR - authentic content reused.

Scheme	Embedding Method	PSNR	Restor. fidelity	Tamp. rate	Reference rate	Restoration approach	Ref. data	Spreading approach	AR
Flexible or adaptive schemes									
[34]	LSBS	38 dB	Medium ³	$\leq 72^3$	variable	general erasure communication	T/5 MSB	global-spreading	N
[15]	LSBS	45 dB	High ⁴	N/A	27 - 101 bpb, variable	read non-linearly mapped data	T/8 MSB	global-spreading	N
[54]	LSBS	51 dB	Medium	N/A	variable	read permuted data, error correction coding	T/7 MSB	global-scatter	N
[56]	LSBS	38 dB	High	35	variable	solve linear system for random content groups	T/5 MSB	local-spreading, global-scatter	N
[75]	LSBS	38 dB	High ⁴	54	160 bpb	<i>pixel xor</i> + bit estimation from local pixel correl.	P/5 MSB	global-scatter	Y
[74]	LSBS	38 dB	High ⁴	60	161 bpb	compressive sensing / compressive reconstr. for random content groups	T/5 MSB	local-spreading, global-scatter	Y
[76] - B	LSBS	38 dB	Medium ⁵	24-66	20 + 45 + 102 bpb	solve linear systems for random content groups	P/5 MSB	local-spreading, global-scatter	Y
[44]	LSBM	41 dB	Medium ⁴	95 ⁶	N/A	read permuted data, two chances	P/5 MSB ⁷	global-scatter	N
[68]	LSBM	41 dB	High ⁴	95	8 bpb	read permuted VQ indices, four chances	P/8 MSB	global-scatter	N
Different configurations of the proposed scheme									
$\lambda = 1$	LSBS	38 dB	High	50	160 bpb	self-recovery	T/5 MSB	global spreading	Y
$\lambda = 2$	LSBS	38 dB	V. High	33	320 bpb	self-recovery	T/5 MSB	global spreading	Y

³ restoration fidelity exchangeable for the tampering rate

⁴ gradual drop of the restoration fidelity

⁵ level-wise drop of the restoration fidelity

⁶ manual help might be needed

⁷ down-scaled image used as reference

1.4 Limitations of Existing Schemes

Despite the variety of papers on self-embedding, there still remain certain fundamental problems, which prevent this technology from its widespread adoption. The emerging new research directions towards flexible and adaptive schemes provide better insights into the inherent restoration trade-offs, and the achievable reconstruction performance. Although, there exist a variety of reconstruction approaches, which cover practically every conceivable application scenario, the presented schemes serve mainly as proof-of-concept prototypes.

Detailed examination of the available literature shows that virtually all of the schemes use the least significant bit substitution (LSBS), or at best one of its variations. Such an approach guarantees low embedding distortion, and high embedding rates. It also features low computational complexity, and can be implemented in a straightforward manner. However, it fails to deliver the most fundamental property for practical systems, i.e., robustness against attacks.

Robustness is inherently connected with limited embedding rates. The use of impractically high rates renders many existing schemes nontransferable to real-world environments with lossy-compressed digital images. Some of the enhancements to the original LSBS strategy have proven to be dead-end directions, e.g., least significant bit matching (LSBM) which slightly improves the PSNR scores of the watermarked images, but prevents the crucial reuse of remaining authentic image content. More importantly, this approach does not solve the fundamental problem with non-existing robustness.

The minority of the schemes, which do not directly use the LSBS, e.g., [13, 77] provide only limited improvement, and in the end turn out to be camouflaged versions of bit substitution, either in a transform domain, or in form of quantization index modulation (QIM) [14, 21], which is in fact a generalization of the original concept. The main advantage of adopting these techniques is the inherent limitation of the embedding rate, and the prospect of achieving reasonable robustness. If a scheme is designed with limited embedding rates in mind, there is a better chance of delivering satisfactory performance in practical scenarios.

To the best of my knowledge, there exist only a handful of schemes, which are capable of working with JPEG-compressed images [13, 47, 63, 77], and they deliver very limited reconstruction performance. The schemes presented in [13, 77] are actually general schemes with mild tolerance for non-destructive image processing. The remaining [47, 63] were designed specifically for use with JPEG. The restoration fidelity is at best of order of 25 dB, even at extremely low compression ratios, e.g., quality levels around 90-95. Hence, the design of practical self-embedding, capable of working with popular lossy-compressed image formats, remains the most important challenge.

Further improvement can also be achieved with respect to robustness against

other popular image processing, e.g., cropping. However, the approach presented in [8, 9] clearly shows that this functionality is actually independent from the self-embedding problem, and can be realized with a properly designed content authentication component.

This chapter describes the considered content reconstruction problem in terms of an erasure communication channel. First, I introduce a generic self-embedding framework, and explain why such a model is a good fit for the problem at hand. I emphasize the differences between the communication process in a general, and in the considered scenario, and derive formulas for the applicable reconstruction success bounds. The presented theoretical results are verified experimentally using Monte Carlo simulations.

2.1 Formal Problem Statement

Operation of many self-embedding systems can be summarized in terms of a generic framework, regardless of the assumed formulation of the content reconstruction problem. There are three fundamental properties which differentiate such systems: the reference generation and restoration method, the payload encoding method, and the data embedding scheme. As it will be demonstrated later, all of the three aspects need to be designed properly, to meet certain cooperation requirements.

Let I denote the original, unprotected image and $I_i : i = 1, \dots, N$ the i -th image block in the raster scan order. Let $g_b(I_i)$ denote a reconstruction reference generation function for a single image block which generates exactly b bits of reference information. In general, $b = b(i)$, but at this point, it is assumed that $b = \text{const}$. Thus, $g_b(\cdot)$ generates a complete Nb -bit reconstruction reference:

$$\mathbf{r} = r_1, \dots, r_N = g_b(I_1), \dots, g_b(I_N). \quad (2.1)$$

An inverse function $g_b^{-1}(\cdot)$ restores image blocks from relevant fragments of the reference bit-stream: $I_i^{(\text{ref})} = g_b^{-1}(r_i)$.

Let $h(\cdot)$ denote a hashing function, which generates a cryptographic hash from the image block content I_i , the block payload Y_i , the block number i , and

a security context κ . The security context typically contains a secret key, and some context for authentication, e.g., image identifier or time-stamp. This issue is addressed in Appendix A. The number of hash bits produced is denoted as $|h|$. For brevity, let H_i denote the resulting hash for i -th image block:

$$H_i = h(I_i, Y_i, i, \kappa). \quad (2.2)$$

Let also $f(\cdot)$ and $f^{-1}(\cdot)$ denote the embedding, and the blind watermark extraction functions:

$$f(I_i, Y_i, H_i) \rightarrow I_i^{(w)}, \quad (2.3a)$$

$$f^{-1}(I_i^{(w)}) \rightarrow \hat{Y}_i, \hat{H}_i. \quad (2.3b)$$

Both the watermark embedding, and extraction functions are key controlled, e.g., by pseudo-random order of the selection channel. We omit this obvious dependency for the sake of notation simplicity. The capacity of the watermarking scheme is $B + |h|$ bits per block.

The decoder performs content authentication by comparing the extracted hashes \hat{H}_i with the ones that can be generated from the received watermarked image:

$$H_i = h(I_i^{(w)}, \hat{Y}_i, i, \kappa). \quad (2.4)$$

This comparison yields an erasure map E with binary elements $e_i \in \{0, 1\}$ for image blocks $i = 1, \dots, N$:

$$e_i = \begin{cases} 1 & \text{if } H_i = \hat{H}_i, \\ 0 & \text{otherwise,} \end{cases} \quad (2.5)$$

which indicates which blocks contain altered content or watermark payload.

Communication of the reconstruction reference to the decoder clearly resembles the erasure channel (Fig. 2.1). Each image block carries a single symbol of the watermark payload. Since the erasure (tampering) localization information is intrinsically available after content authentication, the decoder sees the transmitted symbols either as correctly transmitted or erased, for authentic and tampered blocks, respectively. Since the payload is authenticated along with the image content, transition probabilities between different symbols are negligible. In this study, I consider the channel to be symmetric, i.e., each block has the same probability of being tampered.

The best known family of codes for the erasure channel are the digital fountain codes [52]. Given a stream of input symbols, they produce a potentially limitless stream of same-length output symbols. The output symbols are computed by bit-wise exclusive disjunction on randomly selected input symbols. The transmitted

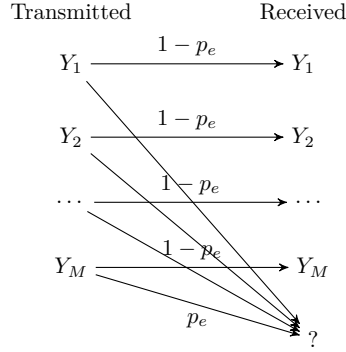


Fig. 2.1: M-ary symmetric erasure channel with probability of erasure p_e .

message can be successfully decoded from arbitrary portions of the resulting stream provided that sufficient amount of correct symbols is collected. In this study, I use the random linear fountain (RLF) code.

In the proposed approach, the reconstruction reference \mathbf{r} consists of N b -bit *reference blocks*. The stream is then divided into constant length B -bit *reference symbols* $X_k : k = 1, \dots, K$ which are then encoded with the RLF code to produce N *embedding symbols* Y_i . In case an image block is tampered, the decoder marks the corresponding embedding symbol as erased and continues with reference data decoding. The code rate, denoted as $\lambda = \frac{K}{N}$, reflects the rate of the effective payload with respect to the available watermark capacity. It will later be shown, that with the assumed formulation of the problem, it is possible to work with reference streams significantly longer than the watermark capacity, i.e., $\lambda > 1$.

Operation of the framework is described in Algorithm 1. In earlier schemes, certain operations from this framework are either skipped completely or trivialized. For instance, often the payload encoding phase consists of simple permutations of the reference stream blocks, leading to reconstruction dependencies.

The key insight to distinguish the cases of general communication and self-recovery communication over the erasure channel is the success criterion. For the reconstruction process to succeed, the decoder needs only selected fragments of the transmitted message, corresponding to the factually needed reference blocks. The remaining portion of the reference stream should be recoverable from the remaining authentic image fragments. This leads to a concise definition of the communication process.

Definition 1. *The self-recovery communication problem is a special case of an erasure channel which:*

1. *uses the media content as a communication channel,*

Algorithm 1 Operation of the generic self-recovery framework.

Require: I, κ **Require:** $h(), f(), g(), f^{-1}(), g^{-1}()$ **Require:** $b, B : b \leq B$ **for** $i = 1 \rightarrow N$ **do** $r_i \leftarrow g_b(I_i)$ **end for** $\mathbf{r} \leftarrow [r_1, \dots, r_N]$ *Encode \mathbf{r} to obtain watermark payload $Y_i : i = 1, \dots, N$* **for** $i = 1 \rightarrow N$ **do** $H_i \leftarrow h(I_i, Y_i, i, \kappa)$ $I_i^{(w)} \leftarrow f(I_i, Y_i, H_i)$ **end for**Tamper selected image blocks : $I_i^{(w)} \rightarrow I_i^{(t)}$ **for** $i = 1 \rightarrow N$ **do** $\hat{Y}_i, \hat{H}_i \leftarrow f^{-1}(I_i^{(t)})$ $H_i \leftarrow h(I_i^{(t)}, \hat{Y}_i, i, \kappa)$ Generate the tampering map : $e_i \leftarrow H_i = \hat{H}_i$ **end for**Discard $\tilde{Y}_i : e_i \neq 1$ *Regenerate $r_i : e_i = 1$* *Remove the dependencies on $r_i : e_i = 1$ from \hat{Y}_i* Recover \mathbf{r} from remaining $\hat{Y}_i = Y_i$ // Only $r_i : e_i \neq 1$ Reconstruct $I_i^{(w)} : e_i \neq 1$ as $I_i^{(\text{ref})}$

-
2. carries the message which describes the media content itself, i.e., the reconstruction reference,
 3. aims to recover only the reference data of the altered (erased) media fragments,
 4. allows to recover an identical reconstruction reference from image fragments with embedded payload, provided that they are authentic.

The latter condition means that the reconstruction reference generation algorithm must be invariant to the utilized embedding scheme, i.e., $g_b(I_i) = g_b(I_i^{(w)})$. Upon computing the tampering map, the decoder can recalculate the reference blocks for authentic image areas, and remove these dependencies from the correctly recovered symbols. Hence, the decoder can make the reconstruction reference forget what it knows about the authentic blocks, and reduce the problem to tampered blocks only. This selective decoding capability implies the necessity for

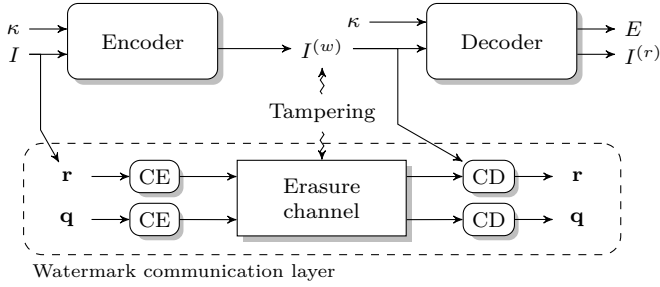


Fig. 2.2: Communication of the reconstruction reference and its corresponding auxiliary information between the encoder and the decoder.

random access to the reference stream. This property can be guaranteed either by using $b = \text{const}$, or by adopting the quality descriptor (Chapter 3).

Communication of the relevant information between the encoder and the decoder is summarized in Fig. 2.2. CE and CD represent a channel encoder, and a channel decoder, respectively. Decoding of the reference information is aided with the use of the remaining authentic content. If necessary, auxiliary information required for reconstruction, denoted here with \mathbf{q} , is communicated through a traditional erasure channel. The decoder, when supplied the same security context as the encoder, yields a tampering map E , and if the reconstruction is possible, a restored image $I^{(r)}$.

Fig. 2.3 illustrates the proposed reconstruction model in detail. In addition to the information processed in successive steps of the algorithm, it also demonstrates the impact of the misalignment between the reference blocks and symbols. A single erased reference block might invalidate multiple reference symbols and thus, limit the prospective reduction of the decoding problem. This effect can be minimized, or even eliminated, by proper choice of b and B . If the boundaries of reference blocks and symbols are aligned, the rate of the necessary reference symbols will be identical to the rate of authentic image blocks. The situation is illustrated in Fig. 2.4. A mathematical explanation will follow later on.

The reduction of the inherent reference decoding problem can also be explained by examining a matrix representation of digital fountain codes. The generator matrix:

$$\mathbf{G}_{N \times K} = [G_{i,k}] : i = 1, \dots, N; k = 1, \dots, K; G_{i,k} \in \{0, 1\} \quad (2.6)$$

fully represents the code by assigning value 1 to elements i, k iff the k -th reference symbol X_k is included in the i -th embedding symbol Y_i . An exemplary random code for $K = 7$ and $N = 9$ is shown in Fig. 2.5a. The rank of the matrix is 7 and all of the reference symbols can be decoded from this set of embedding symbols.

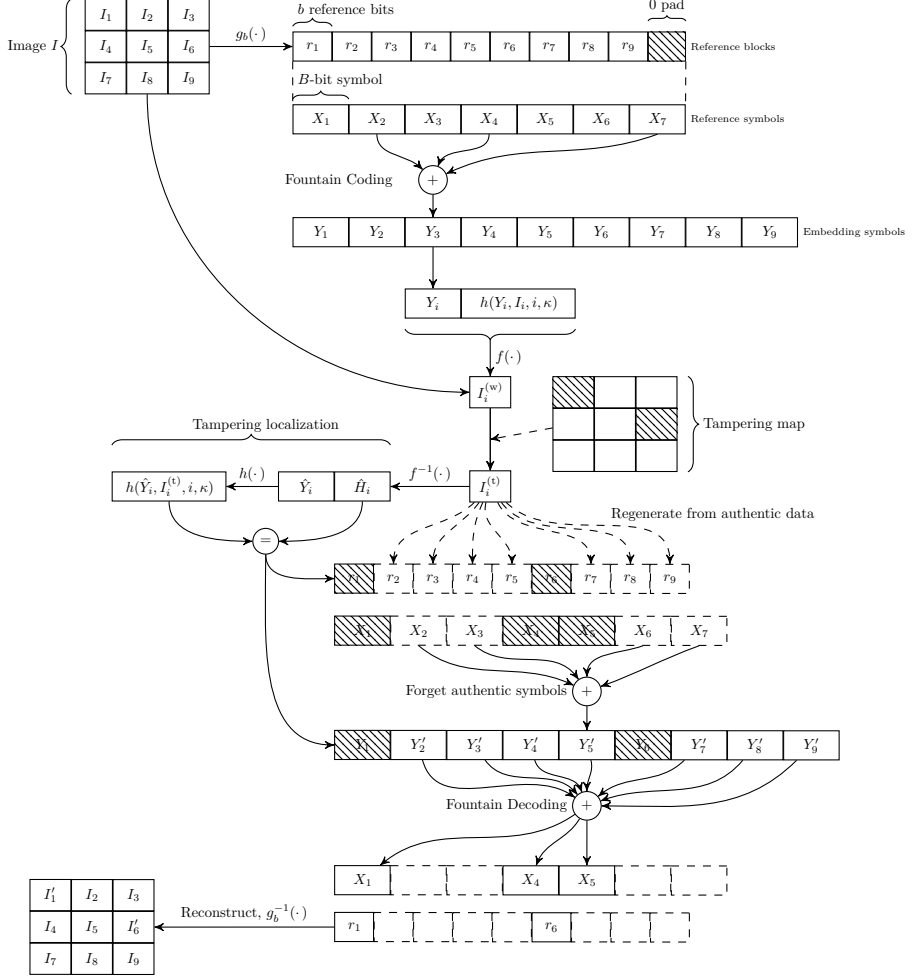


Fig. 2.3: Operation of the proposed content restoration approach - reduction of the decoding problem, and the impact of prospective misalignment between the reference blocks and symbols.

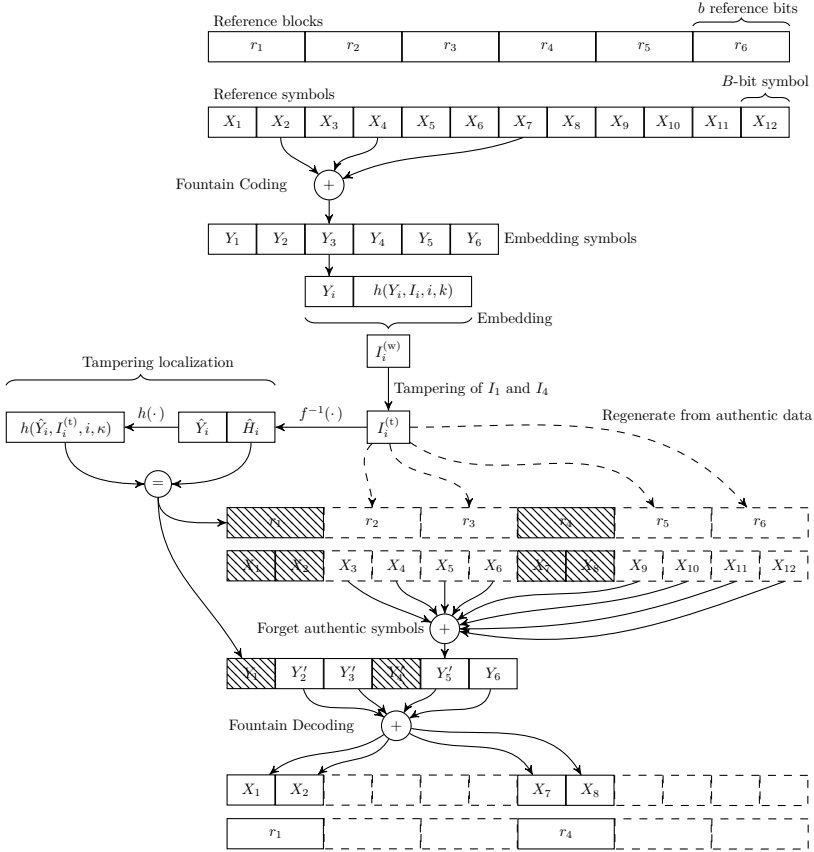


Fig. 2.4: Operation of the proposed content restoration approach for perfect alignment between the reference blocks and symbols.

Consider now a tampering pattern from Fig. 2.3 where Y_1 and Y_6 are erased. The resulting code is represented by a matrix (Fig. 2.5b) of rank 6 and it is no longer possible to successfully decode all X_j symbols. We can however, regenerate the reconstruction reference of known blocks, restore the reference symbols $\{X_2, X_3, X_6, X_7\}$, and remove these dependencies from the original Y_i to produce Y'_i (Fig. 2.5c). The rank of the resulting matrix is 3, and it is possible to recover the remaining 3 unknown reference symbols $\{X_1, X_4, X_5\}$ (Fig. 2.5d).

(a) Full reconstruction problem									
rank 7	X_1	X_2	X_3	X_4	X_5	X_6	X_7		
Y_1	1	0	0	1	0	0	1		
Y_2	0	1	1	1	0	0	1		
Y_3	1	0	1	0	0	1	1		
Y_4	1	1	0	0	1	0	1		
Y_5	0	0	0	1	0	1	1		
Y_6	1	1	1	0	0	0	1		
Y_7	0	0	0	1	1	0	1		
Y_8	0	1	0	1	1	0	1		
Y_9	1	0	1	0	0	1	1		

(b) Erased Y_1 and Y_6									
rank 6	$\overline{X_1}$	X_2	X_3	$\overline{X_4}$	$\overline{X_5}$	X_6	X_7		
Y_2	0	1	1	1	0	0	1		
Y_3	1	0	1	0	0	1	1		
Y_4	1	1	0	0	1	0	1		
Y_5	0	0	0	1	0	1	1		
Y_7	0	0	0	1	1	0	1		
Y_8	0	1	0	1	1	0	1		
Y_9	1	0	1	0	0	1	1		

(c) Removed dependencies on known X_i									
rank 3	$\overline{X_1}$	$\overline{X_4}$	$\overline{X_5}$						
$Y'_2 = Y'_3 \oplus X_2 \oplus X_3 \oplus X_7$	0	1	0						
$Y'_3 = Y_3 \oplus X_3 \oplus X_6 \oplus X_7$	1	0	0						
$Y'_4 = Y_4 \oplus X_2 \oplus X_7$	1	0	1						
$Y'_5 = Y_5 \oplus X_6 \oplus X_7$	0	1	0						
$Y'_7 = Y_7 \oplus X_7$	0	1	1						
$Y'_8 = Y_8 \oplus X_2 \oplus X_7$	0	1	1						
$Y'_9 = Y_9 \oplus X_3 \oplus X_6 \oplus X_7$	1	0	0						

(d) Solved reconstruction problem									
rank 3				$\overline{X_1}$	$\overline{X_4}$	$\overline{X_5}$			
$Y'_3 = Y'_3 \oplus X_3 \oplus X_6 \oplus X_7$				1	0	0			
$Y'_2 = Y_2 \oplus X_2 \oplus X_3 \oplus X_7$				0	1	0			
$Y'_2 \oplus Y'_7 = Y_7 \oplus X_7 \oplus Y_2 \oplus X_2 \oplus X_3$				0	0	1			

Fig. 2.5: Decoding problem reduction in matrix representation of the random linear fountain code: the full code generator matrix \mathbf{G} allows to recover all reconstruction symbols (a), after erasing two symbols, the whole message cannot be decoded (b) but after eliminating the dependencies on known symbols (c), the remaining necessary symbols can be easily decoded (d).

2.1.1 Success Bound Calculation

The bound on the allowed tampering rate can be calculated analytically. Let M denote the number of authentic image blocks, and automatically the embedding symbols Y_i . Then $\gamma = \frac{M}{N}$ denotes the block survival rate and $\tilde{\gamma} = 1 - \gamma$ the tampering rate. Given the probability of decoding error $\delta > 0$, the bound on the reconstruction success for a general erasure channel and a random linear fountain code is [52]:

$$M \geq K + \epsilon(\delta), \quad (2.7a)$$

$$\gamma \geq \lambda + \frac{\epsilon(\delta)}{N}, \quad (2.7b)$$

where $\epsilon(\delta)$ is the number of additional symbols needed for a given δ . Hence, $\epsilon(\delta)$ represents the overhead or imperfectness of the code. For an ideal code, the decoder would always be capable of successful decoding if the number of received symbols is equal to the number of input symbols, i.e., $\epsilon(\delta) \equiv 0$. For the RLF, the overhead is bounded by [10, 52]:

$$\epsilon(\delta) \leq \log_2 \frac{1}{\delta}. \quad (2.8)$$

In practice, for long messages (large images) $\frac{\epsilon}{N} \approx 0$, e.g., assuming the satisfactory decoding error probability is 10^{-6} , for a 512×512 px image and 8×8 px blocks, the overhead is below 0.487%. Hence, in the succeeding derivations this term will be disregarded. The equality in (2.7b) defines the bound on the maximum allowed tampering rate, referred to as the γ_1 bound.

By exploiting the described properties of the self-recovery communication problem, the number of necessary reference symbols becomes reduced. The exact number depends on the observed tampering. For a random choice of image blocks for modification, the success condition becomes:

$$\gamma \geq \lambda \rho(\lambda, \gamma), \quad (2.9a)$$

$$\rho(\lambda, \gamma) : \mathbf{R}^+ \times [0, 1] \rightarrow [0, 1], \quad (2.9b)$$

where $\rho(\lambda, \gamma)$ is the *reconstruction demand*, i.e., the expected value of the fraction of reference symbols X_i which need to be decoded from the remaining embedding symbols $\{\hat{Y}_i : e_i = 1\}$ for a given tampering rate $\tilde{\gamma}$. In general, behavior of the reconstruction demand depends on the character of the image tampering pattern. In the next two sections I provide analytical expressions for an optimistic upper bound, and for an oblivious random tampering pattern.

2.1.2 Optimistic Reconstruction Success Bound

The best possible reconstruction capability is achieved when the tampering in a new reference block r_i yields minimal impact on the reference symbols X_j . Such a situation occurs when the end of each reference block coincides with the end of a reference symbol, when the erasure pattern is continuous over r_i , e.g., when image blocks $i = 1, \dots, N - M$ are tampered. Then:

$$\rho(\lambda, \gamma) = \left(1 - \frac{M}{N}\right) = (1 - \gamma) = \tilde{\gamma}. \quad (2.10)$$

Hence, from (2.9a):

$$\gamma \geq \lambda \left(1 - \frac{M}{N}\right) = \lambda(1 - \gamma), \quad (2.11a)$$

$$\gamma \geq \frac{\lambda}{\lambda + 1}. \quad (2.11b)$$

The inequality (2.11b) represents the condition of successful reconstruction, and a corresponding equality defines the optimistic reconstruction success bound, referred to as the γ_2 bound. What is important, is that this optimistic bound is relevant not only for continuous erasure patterns. It can also be reached for $\lambda \in \mathbf{N}$. This phenomenon will be explained in more detail later on.

2.1.3 Typical Reconstruction Success Bound

In general, linear growth of the reconstruction demand with the tampering rate cannot be safely assumed. The tampering locations can be distributed over the image, and the misalignment between the boundaries of reference blocks and symbols will cause a more precipitous increase of the reconstruction demand. An example such situation has already been shown in Fig. 2.3 where the damage of one image block - I_6 - implicates the necessity to decode two reference symbols - X_4 , and X_5 .

In this section, I derive the reconstruction success bound for a random tampering pattern, where the image blocks for modification are chosen randomly. The main factor that influences the reconstruction demand is the misalignment between the reference blocks and symbols. Assuming that both the number of reference bits per block b , and the symbol length B are longer than one bit, i.e., $b, B > 1$, two boundary cases can be distinguished:

1. $hcf(b, B) = b$ or $hcf(b, B) = B \Leftrightarrow \frac{1}{\lambda} \in \mathbf{N}$ or $\lambda \in \mathbf{N}$,

2. $hcf(b, B) = 1$.

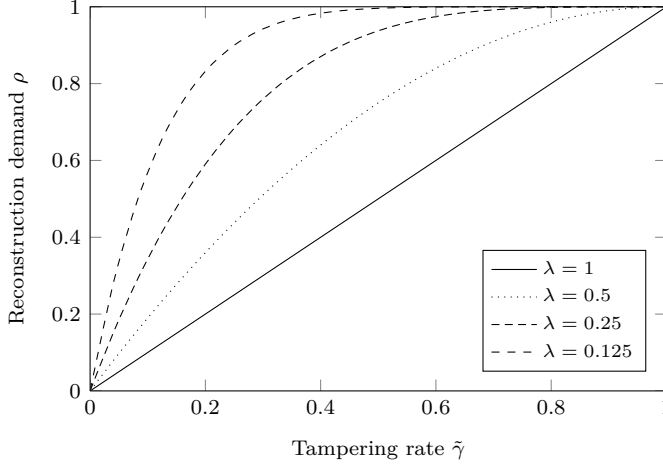


Fig. 2.6: Family of $1 - \gamma^{\alpha(\lambda)}$ functions for $\alpha = \frac{1}{\lambda}$.

$hcf(\cdot)$ is the highest common factor. In the first of the identified cases B is a multiple of b , or b is a multiple of B , and the probability of invalidating a reference symbol can be calculated in a straightforward manner. Since each reference symbols is overlapped by exactly $\lceil \frac{1}{\lambda} \rceil$ reference blocks, it will need to be decoded provided that any of the corresponding reference blocks is required. Hence:

$$\rho(\gamma, \lambda) = 1 - \gamma^{\lceil \frac{1}{\lambda} \rceil}. \quad (2.12)$$

It immediately follows that for $\lambda \in \mathbf{N}$, the reconstruction bound is identical to the previously calculated optimistic bound γ_2 .

For the remaining case, it would be beneficial to express the reconstruction demand in a general convenient form of:

$$\rho(\gamma, \lambda) = 1 - \gamma^{\alpha(\lambda)}, \quad (2.13)$$

with a case dependent function $\alpha(\lambda)$. The corresponding function family for different values of α is shown in Fig. 2.6. The necessary theoretical estimate can be obtained by analyzing the overlap between the reference blocks and symbols. In the following analysis, I consider the case of $hcf(b, B) = 1$. It will be demonstrated that if $B \neq b \neq hcf(b, B)$, the performance slightly improves, but is still well approximated by the typical case formula.

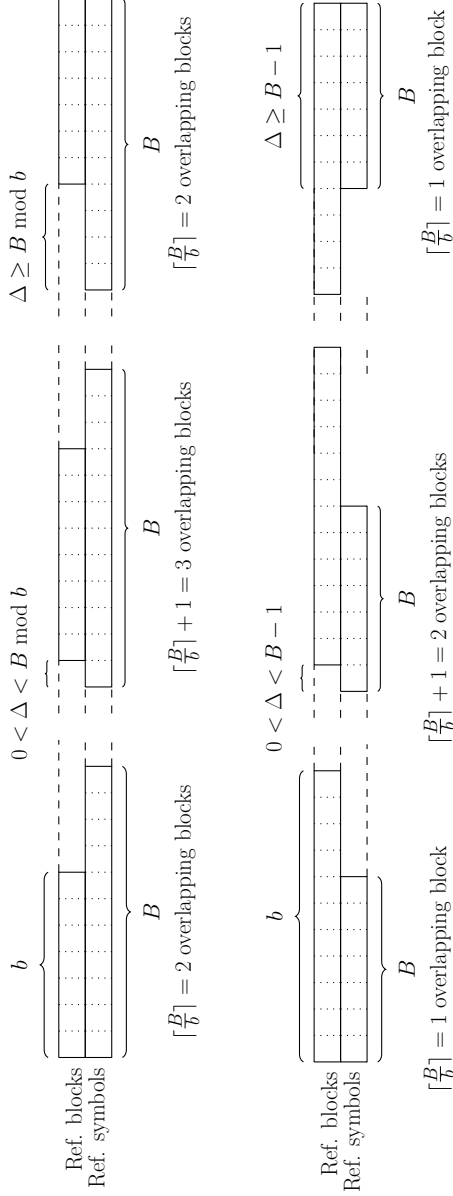


Fig. 2.7: Impact of the misalignment between the reference blocks and symbols on the number of overlapping blocks per single symbol on an example stream division with $b = 7$, $B = 11$, and $b = 11$, $B = 7$.

Proposition 1. *The average number of reference blocks overlapping a reference symbol is bounded by $\frac{1}{\lambda} + 1$.*

Proof. Let $b, B \in \mathbf{N}^+ : \text{hcf}(b, B) = 1$. Division of a bit-stream into b -bit blocks and B -bit symbols is shown in Fig. 2.7, separately for $b < B$, and $b > B$. The misalignment between the starting points of blocks and symbols is denoted as Δ . Depending on the location of the starting point of the next block within a symbol, the number of blocks that overlap a single symbol can assume two values: $\lceil \frac{1}{\lambda} \rceil$ or $\lceil \frac{1}{\lambda} \rceil + 1$. The average number of overlapping reference blocks per reference symbols is a weighted average:

$$(1 - \beta) \lceil \frac{1}{\lambda} \rceil + \beta (\lceil \frac{1}{\lambda} \rceil + 1), \quad (2.14)$$

stemming from the proportion between the two types of symbols.

For any $n \in \mathbf{N}$:

$$Bn \bmod b = 0, 1, \dots, b - 1, \quad (2.15)$$

with a uniform distribution of $\{0, 1, \dots, b - 1\}$. Let me consider the case of $b < B$ first. If $Bn \bmod b = 0$ or $Bn \bmod b \geq B \bmod b$ there is no room for the extra block and the number of overlapping blocks is $\lceil \frac{1}{\lambda} \rceil$. If $0 < Bn \bmod b < B \bmod b$, the additional block will result in a total of $\lceil \frac{1}{\lambda} \rceil + 1$. Thus, the average number of reference blocks overlapping a reference symbol is:

$$\left(1 - \frac{(B \bmod b) - 1}{b}\right) \lceil \frac{1}{\lambda} \rceil + \frac{(B \bmod b) - 1}{b} \left(\lceil \frac{1}{\lambda} \rceil + 1\right). \quad (2.16)$$

By treating the misalignment of $\Delta = 0$ as if it produced the additional overlapping symbol, we obtain an upper bound of:

$$\left(1 - \frac{B \bmod b}{b}\right) \lceil \frac{1}{\lambda} \rceil + \frac{B \bmod b}{b} \left(\lceil \frac{1}{\lambda} \rceil + 1\right). \quad (2.17)$$

After simple algebraic expansion:

$$\left(1 - \frac{B \bmod b}{b}\right) \lceil \frac{1}{\lambda} \rceil + \frac{B \bmod b}{b} \left(\lceil \frac{1}{\lambda} \rceil + 1\right) = \frac{B \bmod b}{b} + \lceil \frac{1}{\lambda} \rceil. \quad (2.18)$$

For $x \in \mathbf{R}$, $x = \lfloor x \rfloor + \{x\}$ where $\{\cdot\}$ denotes the fractional part. $\frac{B \bmod b}{b} = \{\frac{1}{\lambda}\}$ and hence:

$$\frac{B \bmod b}{b} + \lceil \frac{1}{\lambda} \rceil = \{\frac{1}{\lambda}\} + \lfloor \frac{1}{\lambda} \rfloor + 1 = \frac{1}{\lambda} + 1. \quad (2.19)$$

For $b > B$, $\lceil \frac{1}{\lambda} \rceil = 1$, and the number of overlapping reference blocks is always either 1 or 2. The former is true if $\Delta = 0$ or $\Delta \geq B - 1$. The latter, if

$0 < \Delta < B - 1$. Again, for the sake of upper bound calculation, I include the case of $\Delta = 0$ in the latter. Hence, the average number of overlapping reference blocks is:

$$\left(1 - \frac{B}{b}\right) \lceil \frac{1}{\lambda} \rceil + \frac{B}{b} \left(\lceil \frac{1}{\lambda} \rceil + 1\right) \quad (2.20)$$

$$\frac{1}{\lambda} + \lceil \frac{1}{\lambda} \rceil = \frac{1}{\lambda} + 1 \quad (2.21)$$

□

The reconstruction demand is clearly bounded by two weak bounds:

$$1 - \gamma^{\lceil \frac{1}{\lambda} \rceil} \leq \rho(\gamma, \lambda) \leq 1 - \gamma^{\lceil \frac{1}{\lambda} \rceil + 1}. \quad (2.22)$$

It peaks towards the lower bound whenever $hcf(b, B) > 1$ (equals when $\lambda \in \mathbf{N}^+$ or $\frac{1}{\lambda} \in \mathbf{N}^+$). It is possible to find a stronger upper bound. In the following analysis, I show that $1 - \gamma^{\frac{1}{\lambda} + 1}$ is not only an upper bound on the reconstruction demand, but also serves as its good approximation.

Proposition 2. *Given the fraction of necessary reference blocks $\tilde{\gamma}$, the expected fraction of reference symbols that need to be decoded is approximately $1 - \gamma^{\frac{1}{\lambda} + 1}$.*

Proof. In the proof of Proposition 1, I have shown that for $hcf(b, B) = 1$ the proportion between the reference symbols overlapped by $\lceil \frac{1}{\lambda} \rceil + 1$ and by $\lceil \frac{1}{\lambda} \rceil$ reference blocks is approximately $\{\lambda^{-1}\} : 1 - \{\lambda^{-1}\}$. Thus, disregarding the influence of neighboring reference symbols, the expected value of the fraction of authentic reference symbols:

$$1 - \rho(\gamma, \lambda) \approx \{\lambda^{-1}\} \gamma^{\lceil \lambda^{-1} \rceil + 1} + (1 - \{\lambda^{-1}\}) \gamma^{\lceil \lambda^{-1} \rceil}. \quad (2.23)$$

I now estimate the right hand side as follows:

$$\{\lambda^{-1}\} \gamma^{\lceil \lambda^{-1} \rceil + 1} + (1 - \{\lambda^{-1}\}) \gamma^{\lceil \lambda^{-1} \rceil} \approx \gamma^{\frac{1}{\lambda} + 1}, \quad (2.24a)$$

$$\{\lambda^{-1}\} \gamma \gamma^{\lceil \lambda^{-1} \rceil} + \gamma^{\lceil \lambda^{-1} \rceil} - \{\lambda^{-1}\} \gamma^{\lceil \lambda^{-1} \rceil} \approx \gamma^{\{\lambda^{-1}\} + \lceil \lambda^{-1} \rceil}, \quad (2.24b)$$

$$\{\lambda^{-1}\} \gamma^{\lceil \lambda^{-1} \rceil} (\gamma - 1) + \gamma^{\lceil \lambda^{-1} \rceil} \approx \gamma^{\{\lambda^{-1}\}} \gamma^{\lceil \lambda^{-1} \rceil}, \quad (2.24c)$$

$$\{\lambda^{-1}\} (\gamma - 1) + 1 \approx \gamma^{\{\lambda^{-1}\}}. \quad (2.24d)$$

Since the only dependency on λ is on $\{\lambda^{-1}\}$, the approximation will be fulfilled for all $\lambda^{-1} \in \mathbf{R}$ if it is fulfilled for $\lambda^{-1} \in [0, 1]$. Hence:

$$\lambda^{-1} (\gamma - 1) \approx \gamma^{\lambda^{-1}} - 1. \quad (2.25)$$

This approximation stems from the Taylor series expansion of the right hand side around $\gamma = 1$:

$$\lambda^{-1}(\gamma - 1) - \frac{(\gamma - 1)^2(\lambda - 1)}{2\lambda^2} + O((\gamma - 1)^3). \quad (2.26)$$

□

Proposition 3. *The behavior of the reconstruction demand is well represented by the following function:*

$$\rho(\gamma, \lambda) = \begin{cases} 1 - \gamma^{\lceil \frac{1}{\lambda} \rceil}, & \text{if } \frac{1}{\lambda} \in \mathbf{N} \text{ or } \lambda \in \mathbf{N}, \\ 1 - \gamma^{\frac{1}{\lambda} + 1}, & \text{otherwise.} \end{cases} \quad (2.27)$$

For the former condition, the function is an exact formula for the reconstruction demand. In the remaining case, it serves as a close pessimistic approximation. By substituting (2.27) to (2.9a) I obtain the definition of the last of the considered reconstruction success bounds, the γ_3 bound:

$$\begin{cases} \gamma \geq \lambda (1 - \gamma^{\lceil \frac{1}{\lambda} \rceil}), & \text{if } \frac{1}{\lambda} \in \mathbf{N} \text{ or } \lambda \in \mathbf{N}, \\ \gamma \geq \lambda (1 - \gamma^{\frac{1}{\lambda} + 1}), & \text{otherwise.} \end{cases} \quad (2.28)$$

It is not possible to analytically derive a formula for γ_3 and (2.28) needs to be solved numerically. Fig. 2.8 shows all of the defined reconstruction success bounds γ_1 , γ_2 , and γ_3 . For reference, the figure also shows a tampering rate bound derived using a more accurate form of $\rho(\gamma, \lambda)$ from (2.23).

$$\gamma \geq \lambda (1 - \{\lambda^{-1}\} \gamma^{\lceil \lambda^{-1} \rceil + 1} - (1 - \{\lambda^{-1}\}) \gamma^{\lceil \lambda^{-1} \rceil}). \quad (2.29)$$

This bound is denoted as the γ'_3 bound.

2.2 Experimental Evaluation

Experimental validation of the proposed content reconstruction model is divided into two main parts. Firstly, I assess the accuracy of the assumed reconstruction demand estimate (2.27). Secondly, I evaluate the reconstruction success bound via Monte Carlo simulations, and compare the obtained experimental bound with the theoretical results. The experiments are performed on 256×256 px images using a reference self-embedding system.

2.2.1 Reference Self-Embedding Scheme

Operation of the reference self-embedding scheme follows the general algorithm and principles described in Section 2.1. The image is divided into non-overlapping

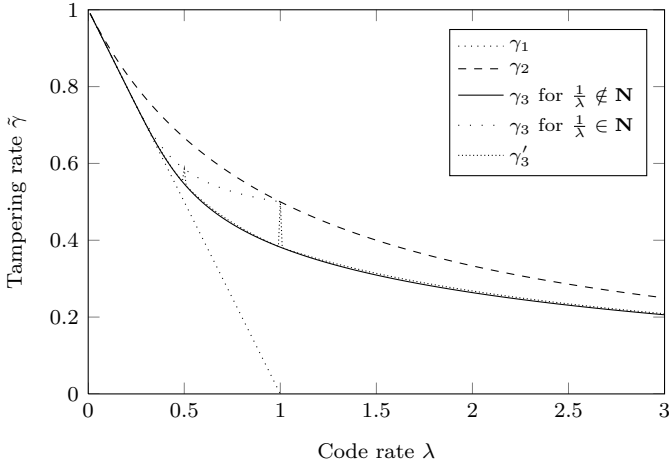


Fig. 2.8: Derived reconstruction success bounds.

8×8 px blocks, which serve as authentication and reconstruction units. Watermark embedding is performed by means of bit substitution in L least significant bit-planes (LSBS). Such embedding strategy is used by most of alternative schemes. Hence, a benchmark configuration with $L = 3$ facilitates fair comparison of the reconstruction performance. The remaining $8 - L$ bit-planes are considered as visually important, and are transformed into DCT domain for generating the reconstruction reference. Such construction ensures that the embedded watermark does not interfere with the reconstruction reference generation basis.

The resulting $64L$ bits of watermark capacity are divided into two parts. 32 bits are used for embedding the H_i hashes obtained by shortening the MD5 hashes by exclusive disjunction on neighboring bit pairs. The remaining $B = 64L - 32$ bits are used for embedding the reconstruction reference. Variations of λ are obtained by controlling b .

The reconstruction reference generation function performs quantization of DCT coefficients. The quantization might be performed with various code-books, and with various precision. This issue is not addressed at this point, since the reconstruction fidelity is not considered in this experiment. A detailed procedure, and a comprehensive quality assessment will follow in the next chapters.

During content restoration, only $8 - L$ most significant bit-planes can be recovered. The remaining L bit-planes are zeroed, with the exception of the L -th least significant plane which is set to 1. The process is performed also for authentic image blocks. This prevents visible noise boundaries between the authentic, and the restored image blocks.

2.2.2 Validation of the Reconstruction Demand Estimate

Validation of the assumed reconstruction demand estimate is performed by repeating the full *content protection* \rightarrow *tampering* \rightarrow *reconstruction* cycle. The scheme is configured with $L = 2$, which corresponds to $B = 96$ bit symbols. The considered range of λ is $(0.2, 3)$, i.e., $b = 19, 20, \dots, 288$. The tampering rates $\tilde{\gamma}$ are randomly drawn from a uniform distribution $U(0, 1)$.

The experiment is repeated 250 times for each of the considered values of b . The decoder records the observed reconstruction demand values ρ . For each 250-point result set, a fit to the $\phi(\gamma)$ function is calculated:

$$\psi_{\alpha}(\gamma) = 1 - \gamma^{\alpha} \quad (2.30)$$

The shape parameter α is estimated by solving a non-linear least squares problem. Fig. 2.9 shows the obtained estimates $\hat{\alpha}$ along with the corresponding theoretical results. Just as expected, the estimates are well approximated by $\frac{1}{\lambda} + 1$ with occasional peaks towards the lower bound of $\lceil \frac{1}{\lambda} \rceil$ whenever $hcf(b, B) > 1$.

Fig. 2.10 shows scatter-plots of $\rho \times \tilde{\gamma}$ for $\lambda = 0.5, 1, 1.5, 2$. The plots show the obtained samples, and the relevant function fits. Success-failure distinction is provided by different marks of the samples. Vertical helper lines represent the appropriate success bounds, which can be intuitively seen as an intersection of the reconstruction demand, and $\lambda^{-1}(1 - \tilde{\gamma})$, i.e., relative fraction of available embedding symbols.

This experiment clearly demonstrates that the derived theoretical estimates closely approximate the real behavior of the reconstruction demand. For $b, B : 1 < hcf(b, B) < b$, the assumed estimate is potentially least accurate. However, practically achievable bounds are better than their theoretical estimates. An example case of $\lambda = 1.5$, where the difference between the theoretical value of the shape parameter α and its estimate $\hat{\alpha}$ is well visible, is shown in Fig. 2.10c. The theoretically expected reconstruction success bound is $\hat{\gamma} = 0.309$, while the practically achievable bound is $\hat{\gamma} = 0.359$.

The obtained results also show the validity of the theoretical reconstruction success bounds. The empirical bound between the reconstruction success and failure coincides with the γ_3 bound. For $\lambda \in \mathbf{N}$, the reconstruction success bound reaches the optimistic bound of γ_2 from (2.11b). The validity of the derived bounds will be better visible in the next experiment.

2.2.3 Validation of the Reconstruction Success Bounds

The purpose of this experiment is to validate the introduced bounds: γ_1, γ_2 , and γ_3 . In each iteration, the image is encoded with a random $\lambda \leftarrow U(0.2, 3)$ setting, a random fraction $\tilde{\gamma} \leftarrow U(0, 1)$ of the available image blocks is tampered, and an

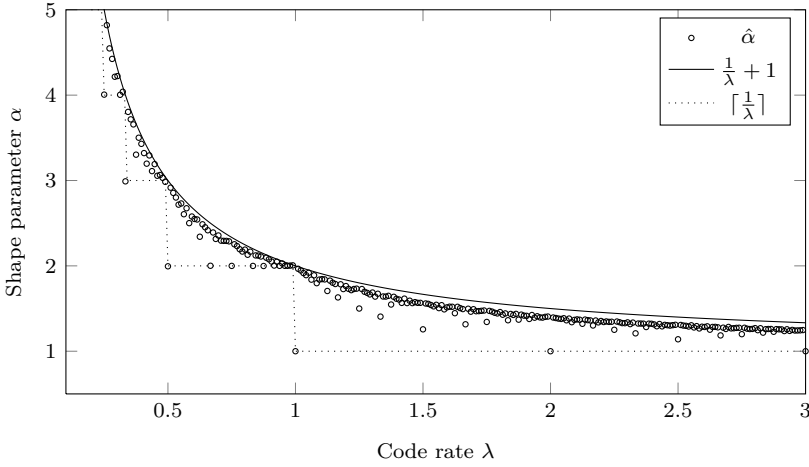


Fig. 2.9: Theoretical vs. empirical estimates of the shape parameter α .

attempt to perform content reconstruction is made. The experiment is replicated 5,000 times, with different seeds for the pseudo-random number generator.

The results are shown in Fig. 2.11 and 2.12. Successes and failures are marked with circles and crosses, respectively. Each plot shows the applicable theoretical bound with a solid line. The γ_1 bound is shown in Fig. 2.11a. It corresponds to decoding with a general erasure formulation of the reconstruction problem, i.e., without reusing the remaining authentic image content. In this case, the success bound does not depend on the tampering pattern, and the experiment was performed with randomly tampered blocks. Evaluation of the self-recovery model is performed for both continuous and random tampering. In general, the former case is bounded by γ_2 (Fig. 2.11b), while the latter by γ_3 (Fig. 2.12a).

The obtained experimental results are in good correspondence with the theoretically derived bounds. Certain reference rates are preferable, as they guarantee alignment between the reference blocks and symbols, and allow for reconstruction near the optimistic success bounds γ_2 . Fig. 2.12b shows evaluation results for configurations with $\lambda \leftarrow \lceil U(0.2, 3) \rceil$. The fractional part serves only the purpose of presentation clarity.

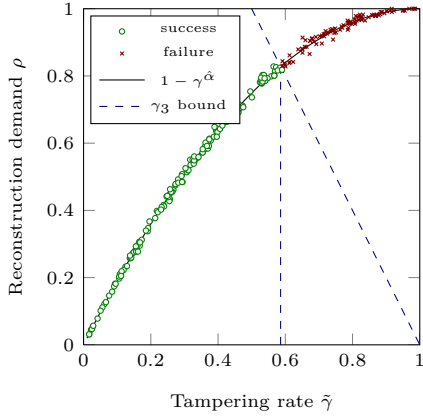
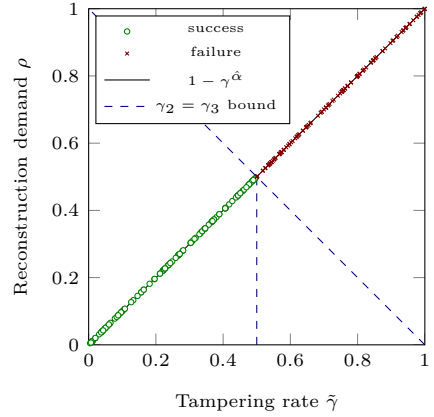
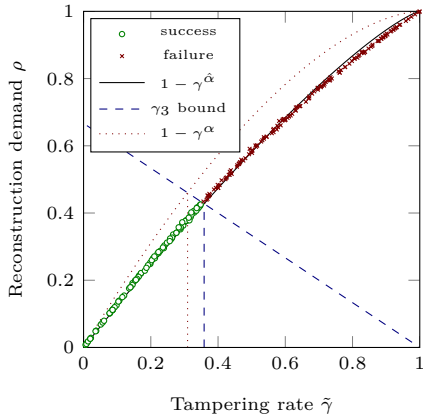
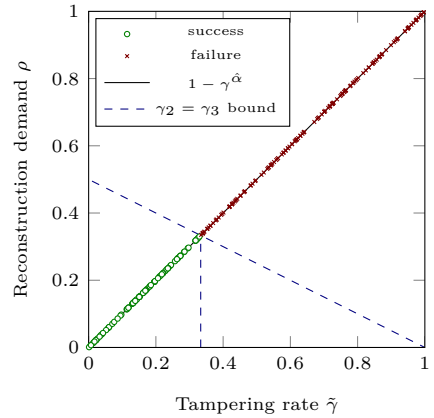
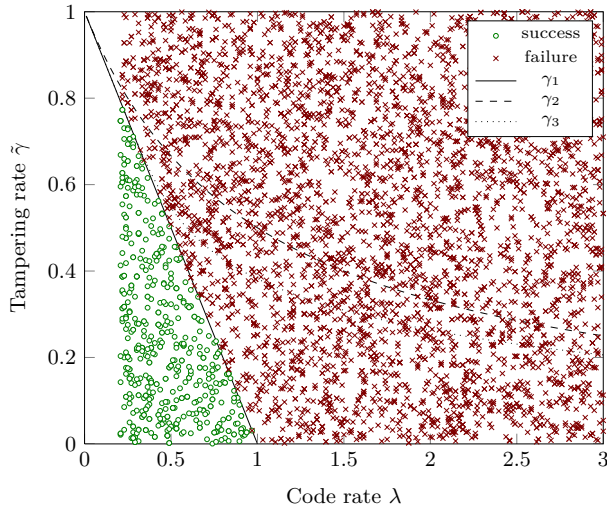
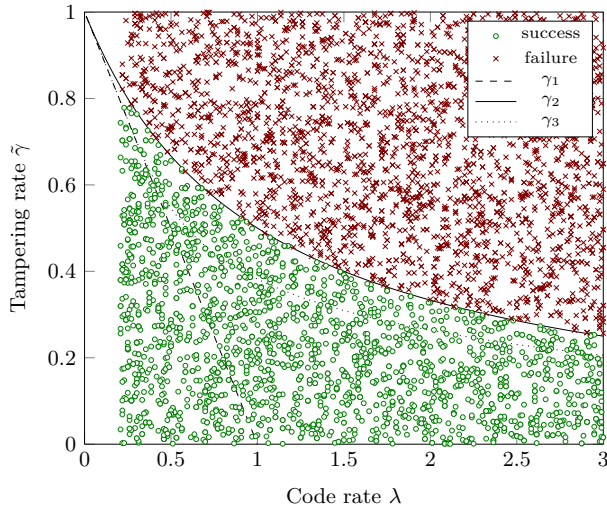
(a) $B = 48$ ($\lambda = 0.5$), $\alpha = 2.00$, $\hat{\alpha} = 1.99$ (b) $B = 96$ ($\lambda = 1$), $\alpha = 1.00$, $\hat{\alpha} = 1.00$ (c) $B = 144$ ($\lambda = 1.5$), $\alpha = 1.67$, $\hat{\alpha} = 1.26$ (d) $B = 192$ ($\lambda = 2$), $\alpha = 1.00$, $\hat{\alpha} = 1.00$

Fig. 2.10: Experimental and theoretical dependency between the reconstruction demand and the tampering rate for selected reference rates; graphical interpretation of the γ_3 bound is depicted with a dashed line.

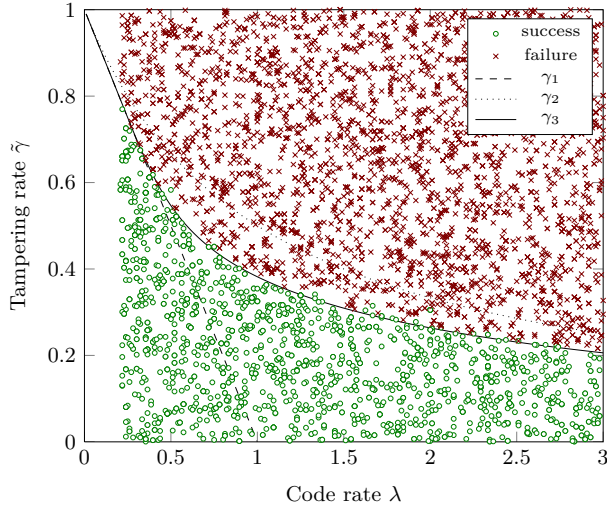


(a) General erasure



(b) Self-recovery, continuous tampering

Fig. 2.11: Experimental evaluation of the reconstruction success bounds γ_1 and γ_2 using Monte Carlo simulations. Each sample represents a single reconstruction attempt. The applicable theoretical bound between the successes and failures is shown as a solid line.



(a) Self-recovery, random tampering

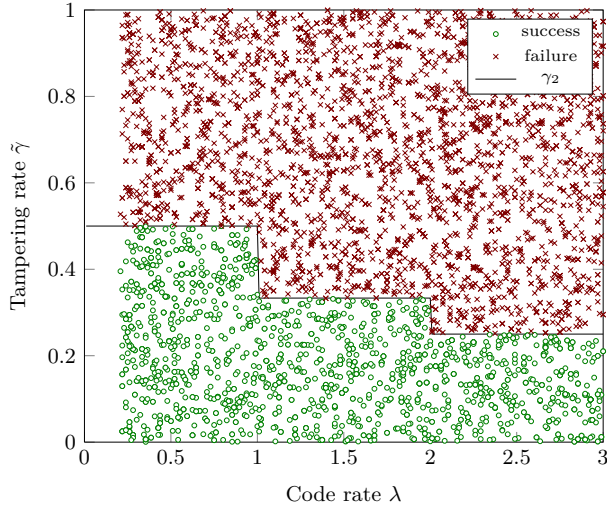
(b) Self-recovery, $\lambda \in \mathbb{N}^+$

Fig. 2.12: Experimental evaluation of the reconstruction success bounds γ_3 and γ_2 using Monte Carlo simulations. Each sample represents a single reconstruction attempt. The applicable theoretical bound between the successes and failures is shown as a solid line.

2.3 Conclusions and Practical Design Guidelines

The presented analysis leads to clear guidelines for the construction of efficient content reconstruction systems:

1. Fountain codes can be used to generate a reference stream of arbitrary length while uniformly spreading the reference information over the whole image;
2. During distribution of the reference information over the image, each content fragment should uniformly contribute to the whole embedded watermark;
3. Exploitation of the authentic image content is essential to achieve optimal performance;
4. The watermark embedding function f should not interfere with the reference generation function g_b ;
5. Random access to the reference stream is necessary to determine its actually needed fragments;
6. The fraction of the necessary reference stream fragments should be linearly proportional to the tampering rate, i.e., $\rho = \tilde{\gamma}$. This can be achieved by perfect alignment between the reference blocks and the reference symbols.

Perfect alignment between the reference blocks and symbols could also be achieved by embedding multiple shorter symbols in a single image block. Asymptotically, 1-bit symbols could be considered, and RLF would produce $N(64L - 32)$ watermark bits from bN reference bits. Then, the applicable success bound would always be the optimistic γ_2 . However, due to high computational complexity of the RLF decoding, such an approach is not feasible. The typical solution of using a sparse generator matrix and belief propagation for the decoding is not applicable for the self-recovery problem, as low-degree symbols are likely to be quickly reduced to null useless symbols in the process of eliminating the knowledge of authentic image fragments. This issue could potentially be addressed by designing a dedicated degree distribution, but it is a separate research problem. With the use of M-ary symbols, the proposed approach can be efficiently implemented in practice, and with proper choice of (b, B) , the optimistic success bounds can still be reached.

By reusing the authentic image content, it is possible to make the reference stream forget about the authentic fragments. Hence, the watermark capacity is not wasted, and it is possible to embed reference streams longer than the available capacity. Compared to a situation when the remaining authentic content is not exploited, the improvement introduced by the adoption of the proposed

reconstruction model reaches 50% of the image area. Table 2.1 collects the theoretical success bounds for selected values of λ . With proper scheme design, the optimistic bound γ_2 is achievable. When, the scheme fails to guarantee alignment between the reference blocks and symbols, the γ_3 bound is applicable.

Table 2.1: Reconstruction success bounds for selected reference rates.

Bound	$\lambda = 3$	$\lambda = 2$	$\lambda = 1$	$\lambda = \frac{3}{4}$	$\lambda = \frac{1}{2}$	$\lambda = \frac{1}{4}$
γ_1	0	0	0	0.25	0.5	0.75
γ_2	0.25	0.33	0.50	0.57	0.66	0.80
γ_3	0.20	0.26	0.38	0.44	0.59	0.75

This chapter deals with adaptive self-embedding, where the reconstruction quality is controlled on a per-block basis, i.e., reference rate $b = b(i)$. It is assumed that the system uses P reconstruction profiles, each characterized by its own coefficient precision and quantization strategy. The mapping between the image blocks and the profiles is referred to as a *quality descriptor* $q(i)$:

$$q : \{1, \dots, N\} \rightarrow \{1, \dots, P\}. \quad (3.1)$$

In general, there might exist multiple reconstruction profiles with the same reference payload, i.e., $b(i) = b(j) : i \neq j$, but optimized for different block content. Hence, the quality descriptor implicitly determines the mapping $b(i)$, or equivalently:

$$\lambda : \{1, \dots, N\} \rightarrow \{\lambda_1, \dots, \lambda_{S'}\}, \quad (3.2)$$

which is fundamental in the analysis of the success bounds in adaptive self-embedding systems.

The main focus of this chapter is on the impact of multiple reconstruction profiles on the achievable success bounds. I will show that introduction of content adaptivity, even with profiles of lower restoration fidelity, does not necessarily lead to improvement in terms of achievable tampering rates. Based on both theoretical, and experimental analysis, I present an algorithm for automatic design of quality descriptors. The design objectives include guarantees on the reconstruction fidelity of selected image fragments, and the achievable tampering rates.

3.1 Reconstruction Success Bound Analysis

Incorporation of multiple reconstruction profiles with various reference rates impacts the achievable success bounds. Depending on the observed tampering pattern, different success bounds arise. In Chapter 2, I have shown that certain reference rates are discouraged, due to an inherent performance penalty. Specifically, optimal performance is achieved when $\lambda \in \mathbf{N}$. Hence, the analysis focuses on the case of $\lambda(i) \in \mathbf{N}^+$ for $i \in \{1, \dots, N\}$.

Analogously to uniform-quality reconstruction, success bounds for adaptive systems can be derived by analyzing the behavior of the reconstruction demand. Assuming a zero overhead of the fountain code, the restoration is successful, if the number of necessary reference symbols is lower than the number of authentic embedding symbols. The success bound has an intuitive graphical interpretation as an intersection of the reconstruction demand, and a relative fraction of available embedding symbols. Fig. 3.1 shows a graphical representation of the described bound for two uniform-quality configurations: $\lambda = 0.5$ and 2. The maximal tampering rates $\tilde{\gamma}_{\max}$, correspond to the success bounds derived in Chapter 2.

In adaptive self-embedding systems, the behavior of the reconstruction demand ρ is highly dependent on the observed tampering pattern. Let:

$$\Lambda = [\lambda_1, \dots, \lambda_S] : \lambda_1 > \dots > \lambda_S \quad (3.3)$$

be a vector of unique reference rates used in a given quality descriptor. The occurrence frequency for each reference rate can be represented by a weight vector \mathbf{w} with $S - 1$ degrees of freedom:

$$\mathbf{w} = [w_1, w_2, \dots, w_S] : w_s \in (0; 1) \text{ for } s \in \{1, \dots, S\}, \quad (3.4a)$$

$$\sum_{s=1}^S w_s = 1 \Rightarrow w_S = 1 - \sum_{s=1}^{S-1} w_s. \quad (3.4b)$$

which means that $S - 1$ weights can be adjusted independently.

The success criterion (2.9a) can be rewritten as an equation:

$$\rho(\tilde{\gamma}|\Lambda, \mathbf{w}) = \frac{\gamma}{\lambda_{\text{ave}}} = \lambda_{\text{ave}}^{-1}(1 - \tilde{\gamma}). \quad (3.5)$$

where $\rho(\tilde{\gamma}|\Lambda, \mathbf{w})$ denotes the reconstruction demand for a given set of reference rates Λ with weights \mathbf{w} . The right hand side represents the fraction of the available watermark symbols, normalized with respect to the total number of reference symbols. It is always a linear function, with slope corresponding to the weighted average reference rate:

$$\lambda_{\text{ave}} = \sum_{s=1}^S w_s \lambda_s. \quad (3.6)$$

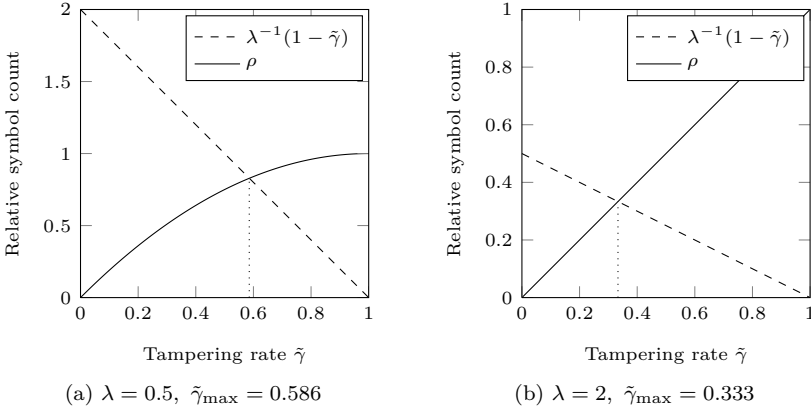


Fig. 3.1: Graphical interpretation of the reconstruction success bound.

3.1.1 Success Bound Types

When analyzing the reconstruction demand in adaptive self-embedding systems, it is necessary to consider different success bounds, applicable for different tampering patterns. In general, two representative cases can be distinguished: pessimistic success bound $\tilde{\gamma}_{\min}$ and average success bound $\tilde{\gamma}_{\text{ave}}$.

Pessimistic Success Bound

The pessimistic bound $\tilde{\gamma}_{\min}$ is the minimum tampering rate that can render the restoration impossible. It is applicable when the reconstruction demand begins with the steepest possible growth. In adaptive schemes, this happens when the blocks are tampered with decreasing order of reference rates $\lambda(i)$. This corresponds to a realistic assumption that the most important, and hence tampering-prone, content is likely to be assigned higher values of λ . Within these implicitly defined content importance levels no particular secondary importance is assumed. During simulations, image blocks within a single importance level are chosen randomly. Such modifications are referred to as the pessimistic tampering.

For pessimistic tampering, the reconstruction demand becomes a piecewise-linear function with slopes dependent on both reference rates and their weights. Let w_s^+ denote the cumulative weight for the s -th reference rate:

$$w_s^+ = \begin{cases} 0, & \text{for } s = 0, \\ \sum_{i=1}^s w_i, & \text{for } s = 1, \dots, S. \end{cases} \quad (3.7)$$

Then, given the reference rates Λ with weights \mathbf{w} , the reconstruction demand is simply a function of the tampering rate only:

$$\rho(\tilde{\gamma}|\Lambda, \mathbf{w}) = \begin{cases} \alpha_1 \tilde{\gamma} + \beta_1, & \text{for } \tilde{\gamma} \in (0, w_1^+], \\ \dots \\ \alpha_s \tilde{\gamma} + \beta_s, & \text{for } \tilde{\gamma} \in (w_{s-1}^+, w_s^+], \\ \dots \\ \alpha_S \tilde{\gamma} + \beta_S, & \text{for } \tilde{\gamma} \in (w_{S-1}^+, 1]. \end{cases} \quad (3.8)$$

where the slopes α_s can be calculated as:

$$\alpha_s = \lambda_s / \sum_{i=1}^S w_i \lambda_i, \quad (3.9)$$

and the y-intercepts β_s for individual components are chosen to ensure function continuity:

$$\beta_s : \bigvee_{s \in \{1, \dots, S\}} \rho(w_s^+|\lambda, \mathbf{w}) = \lim_{\tilde{\gamma} \rightarrow (w_s^+)^+} \rho(\tilde{\gamma}|\lambda, \mathbf{w}). \quad (3.10)$$

For the currently considered scenario, the y-intercepts β_s can be calculated as:

$$\beta_s = \begin{cases} 0, & \text{for } s = 1, \\ (\alpha_{s-1} - \alpha_s)w_{s-1}^+ + \beta_{s-1}, & \text{otherwise.} \end{cases} \quad (3.11)$$

Fig. 3.2 shows the reconstruction demand for a randomly generated quality descriptor with $\Lambda = [3, 2, 1]$ and $\mathbf{w} = [0.1429, 0.2857, 0.5714]$. (a) shows the piecewise-linear character of the resulting function with the slopes of its successive components. The theoretically expected character is verified experimentally by performing reconstruction attempts after pessimistic tampering with increasing modification rates. The obtained experimental results are shown in (b). Individual attempts are marked with circles and crosses for successful and unsuccessful reconstruction, respectively. The dashed lines represent the relative number of available embedding symbols, and the achievable tampering rate bound $\tilde{\gamma}_{\min} = 0.2864$.

It is also possible to distinguish a dual, optimistic reconstruction success bound, which is analogous to the pessimistic bound, with the difference that image blocks with minimal impact on ρ are tampered first. Such tampering has no reasonable justification in prospective forgery scenarios, and is therefore not considered.

Average Success Bound

Oblivious tampering randomly selects image blocks for modification. For each individual realization of the tampering process, the achievable success bound will

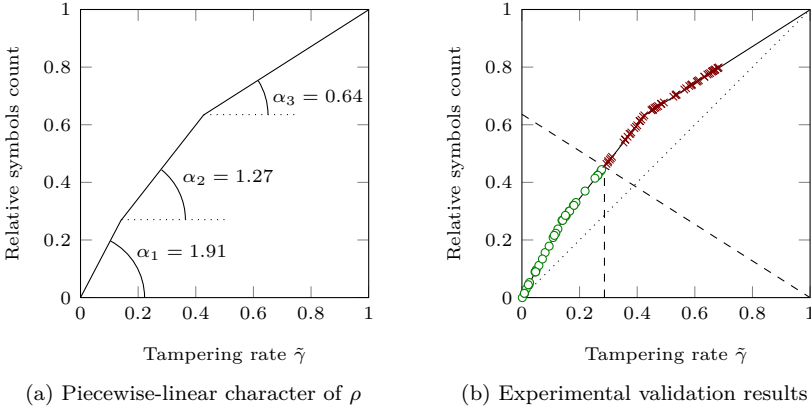


Fig. 3.2: Behavior of the reconstruction demand for pessimistic tampering.

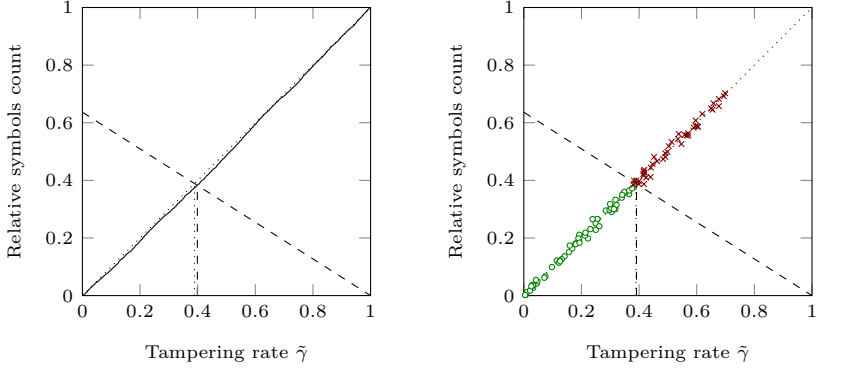
be slightly different, depending on the quality settings of the tampered blocks, and the resulting curve of the reconstruction demand. The average tampering rate represents an estimate of the expected value of the tampering rate bounds.

It can be shown, that for oblivious tampering, the reconstruction demand has stochastic character with oscillations around the expected linear growth. An example realization of the process is shown in Fig. 3.3a with a solid line. The average linear growth is marked with a dotted line. Since the analysis is relative to the total number of reference symbols, the slope of the expected growth is always 1, i.e., $\rho(\tilde{\gamma}|\Lambda, \mathbf{w}) = \tilde{\gamma}$. Hence, the average success bound can be calculated as:

$$\tilde{\gamma}_{\text{ave}} = \frac{1}{1 + \lambda_{\text{ave}}}. \quad (3.12)$$

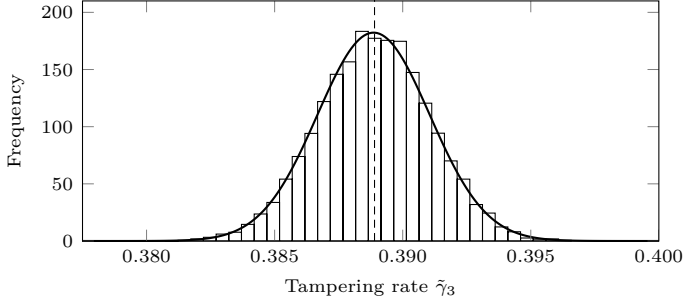
where λ_{ave} is the weighted average reference rate calculated according to (3.6).

Fig. 3.3b shows the results of experimental validation of the described behavior. The illustrated example is the same as in the description of the pessimistic success bound, i.e., $\Lambda = [3, 2, 1]$ and $\mathbf{w} = [0.1429, 0.2857, 0.5714]$. The average success bound is $\tilde{\gamma}_{\text{ave}} = 0.3889$. Since individual realizations oscillate around the linear growth, the actual bound oscillates around $\tilde{\gamma}_{\text{ave}}$. (c) shows a histogram of the bounds obtained from 10,000 simulated realizations of the process. The sample mean is 0.3888.



(a) Linear growth of the reconstruction demand; example realization

(b) Experimental validation of the expected linear growth



(c) Histogram of individual success bounds for 10,000 random realizations

Fig. 3.3: Graphical interpretation of the average tampering rate bound; example case of $\Lambda = [3, 2, 1]$ and $\mathbf{w} = [0.1429, 0.2857, 0.5714]$.

3.1.2 Impact of Multiple Reconstruction Profiles

In this section, I investigate the impact of the introduction of multiple reconstruction profiles on the achievable success bounds. I will show that even introduction of profiles with lower restoration fidelity, does not necessarily improve the supported tampering rates.

For a configuration with two reconstruction profiles, the corresponding weight vector \mathbf{w} has one degree of freedom:

$$\mathbf{w} = [w, 1 - w], \quad (3.13)$$

and the parameter $w \in [0, 1]$ alone is sufficient for controlling the proportion

between the high-quality and low-quality blocks. The reference rates λ_1 and λ_2 control the asymptotic tampering rates for $w \rightarrow 0$, and $w \rightarrow 1$. It can be shown that there is a threshold value of w , above which there is no improvement in terms of the pessimistic tampering rate $\tilde{\gamma}_{\min}$.

Proposition 4. *For a two-profile configuration of an adaptive self-embedding scheme with reference rates $\lambda_1 > \lambda_2$, the pessimistic tampering rate bound for weights $w \in (\frac{1}{1+\lambda_1}, 1]$ is constant and dependent only on the higher reference rate λ_1 .*

Proof. A two-profile configuration is described by reference rates $\Lambda = [\lambda_1, \lambda_2]$, and weights $\mathbf{w} = [w, 1 - w]$. From (3.8)-(3.11), for any $w \in [0, 1]$ the reconstruction demand $\rho(\tilde{\gamma}|\Lambda, \mathbf{w})$ can be rewritten for a pessimistic tampering pattern as:

$$\rho(\tilde{\gamma}|\Lambda, \mathbf{w}) = \begin{cases} \frac{\lambda_1}{\lambda_{\text{ave}}} \tilde{\gamma}, & \text{for } \tilde{\gamma} \in [0, w], \\ \frac{\lambda_2}{\lambda_{\text{ave}}} \tilde{\gamma} + \left(\frac{\lambda_1}{\lambda_{\text{ave}}} - \frac{\lambda_2}{\lambda_{\text{ave}}} \right) w, & \text{for } \tilde{\gamma} \in (w, 1]. \end{cases} \quad (3.14)$$

For the first component, the solution to (3.5) is dependent only on the higher reference rate λ_1 :

$$\frac{\lambda_1}{\lambda_{\text{ave}}} \tilde{\gamma}_{\min} = \frac{1}{\lambda_{\text{ave}}} (1 - \tilde{\gamma}_{\min}), \quad (3.15a)$$

$$\tilde{\gamma}_{\min} = \frac{1}{1 + \lambda_1}. \quad (3.15b)$$

The first component determines the reconstruction success bound as long as the reconstruction demand for $\tilde{\gamma} = w$ is greater than the relative number of correctly extracted embedding symbols:

$$\rho(w|\Lambda, \mathbf{w}) > (1 - w)/\lambda_{\text{ave}}, \quad (3.16a)$$

$$w > \frac{1}{1 + \lambda_1} = w_{\text{th}}. \quad (3.16b)$$

□

It is also possible to derive an analytical expression for the reconstruction demand for the remaining range $w \in [0, \frac{1}{1+\lambda_1}]$:

$$\frac{\lambda_2}{\lambda_{\text{ave}}} \tilde{\gamma}_{\min} + \frac{\lambda_1 - \lambda_2}{\lambda_{\text{ave}}} w = \frac{1}{\lambda_{\text{ave}}} (1 - \tilde{\gamma}_{\min}), \quad (3.17a)$$

$$\tilde{\gamma}_{\min} = \frac{1 + (\lambda_2 - \lambda_1)w}{\lambda_2 + 1}. \quad (3.17b)$$

where the behavior is controlled by both reference rates λ_1 , and λ_2 .

Fig. 3.4ab show both the pessimistic, and the average success bounds for $\Lambda = [2, 1]$, $\Lambda = [3, 1]$, and $\Lambda = [3, 2, 1]$. It can be observed, that the weight w needs to drop below the threshold w_{th} in order to obtain an improvement of the pessimistic success bound. Above the threshold, the success bound is determined by the highest reference rate alone. The average success bound increases systematically.

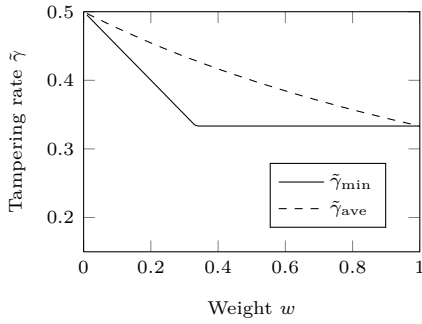
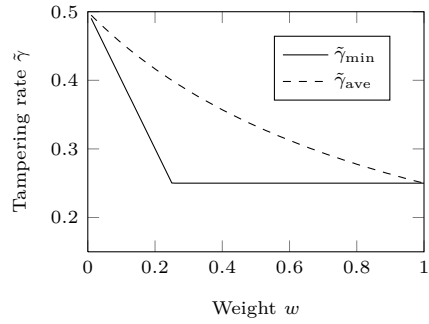
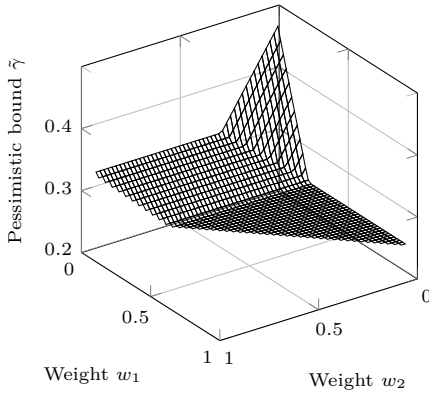
Analogous derivations could be worked out for a three-profile configuration, where there are two degrees of freedom, i.e., for $\mathbf{w} = [w_1, w_2, 1 - w_1 - w_2]$. This issue is out of scope of this analysis. A graphical representation of both the pessimistic, and the average success bounds for a configuration with $\Lambda = [3, 2, 1]$ is shown in Fig. 3.4cd.

Impact of the Null Profile

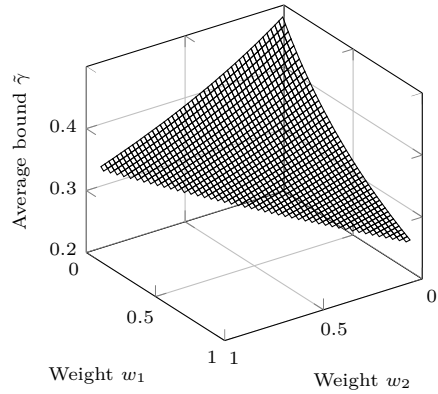
A *null profile* is characterized by a zero-length reference bit-stream. Depending on the scheme design, such blocks may be skipped or restored with a different method, e.g., prediction or inpainting. The impact of the null profile becomes visible only if the tampering affects image blocks marked as such. The number of available embedding symbols is not affected, and the right hand side of (3.5) remains unchanged. The left hand side, i.e., the reconstruction demand, takes into account only the remaining blocks, and is not increased when null-profile blocks are tampered.

The proposed self-recovery model is capable of removing the dependencies on the authentic fragments from the reconstruction reference, regardless of their profile assignment. Hence, in general, there is no improvement in terms of the reconstruction success bounds. The only improvement is in terms of computation time, as there is no need to include and eliminate the information which is known to be unnecessary. Such an approach is particularly useful for applications like reversible privacy protection, where by definition only selected RoIs can be recovered.

Additional care must be taken, when implementing prospective optimizations in the described reconstruction model, e.g., by using a dedicated degree distribution in the fountain coding process. Then, the decisions about the inclusion of individual image fragments in a given embedding symbol are no longer independent, and it becomes justified to deny the reconstruction capability for the irrelevant image content.

(a) Two profiles, $\lambda_1 = 2$, $\lambda_2 = 1$ (b) Two profiles, $\lambda_1 = 3$, $\lambda_2 = 1$ 

(c) Pessimistic bound for three profiles



(d) Average bound for three profiles

Fig. 3.4: Impact of multiple reconstruction profiles on the pessimistic, and the average success bounds for configurations with two, and three reconstruction profiles.

3.2 Automatic Design of Quality Descriptors

There are two main reasons which motivate the use of quality descriptors. Firstly, they allow for further improvement of the achievable restoration quality. A set of reconstruction profiles with the same principal rate can be designed with the use of dedicated allocation and quantization strategies, based on the actual block content. In this study I differentiate the blocks based on their texture level. Since all of the profiles use the same reference rate, the success bounds are not affected, and are identical to the non-adaptive variant of the algorithm.

Secondly, the quality descriptor allows for enforcing the desired reconstruction quality for selected image blocks. The quality of the remaining background content can then be adjusted to meet the prospective requirements towards the tampering rate. Hence, it becomes possible to provide guarantees for both local reconstruction quality, and the achievable success bounds.

3.2.1 Descriptor Design Procedure

This section presents a procedure for generation of quality descriptors. In general, the complete process involves two phases: *profile mapping*, and *descriptor design*. The former chooses the profiles in order to maximize the restoration fidelity. The latter optimizes the descriptor to satisfy both reconstruction quality, and tampering rate requirements.

Definition of Requirements

The possible set of requirements that can be specified for the quality descriptor design procedure includes an *importance map*, a requested target tampering rate, success bound type (i.e., pessimistic or average), a set of *guard profiles* \mathbf{p}_g , and other minor, implementation-specific parameters.

The importance map controls how individual image blocks are handled in the quality adaptation process. In this study, the importance map $\phi(i) : i \in \{1, \dots, N\}$ can assume two possible values:

$$\phi(i) = \begin{cases} 1, & \text{reconstruct block with maximal possible quality,} \\ 0, & \text{allow for quality degradation to design objectives.} \end{cases} \quad (3.18)$$

Optimal profile assignment for all image blocks stems from the profile mapping step. In the successive iterations of the following descriptor design phase, image blocks $\{I_i : \phi(i) = 0\}$ can be degraded to meet the given design objective, i.e., the target tampering rate $\tilde{\gamma}_{\text{target}}$. The blocks for degradation are selected based on their current impact on the overall reconstruction quality.

In order to prevent excessive degradation, the design procedure uses a set of guard profiles \mathbf{p}_g , below which an individual block cannot be degraded. This concept is particularly useful if the given set of reconstruction profiles distinguishes multiple block texture levels. It prevents incorrect assignment to a wrong block type.

Profile Mapping Procedure

The goal of the profile mapping procedure is twofold. Firstly, it provides a quality descriptor q , which optimizes the reconstruction quality among the available fidelity profiles:

$$q(i) : \bigvee_{i=\{1,\dots,N\}} q(i) = \underset{p \in \{1,\dots,P\}}{\operatorname{argmin}} \operatorname{MSE}(I_i, I_i^{(\operatorname{ref}[p])}). \quad (3.19)$$

where $I_i^{(\operatorname{ref}[p])}$ denotes the reconstruction result of the i -th image block with the use of the p -th reconstruction profile. Secondly, it yields a distortion map $d(i, k)$:

$$d(i, p) : \{1, \dots, N\} \times \{1, \dots, P\} \rightarrow \operatorname{MSE}(I_i, I_i^{(\operatorname{ref}[p])}) \in \mathbf{R}^+, \quad (3.20)$$

needed in the following descriptor design phase. For the sake of computation efficiency, the distortion map may include only the allowed profiles for each individual image block.

Descriptor Design Procedure

The descriptor design procedure relies on the initial descriptor, and the distortion mapping obtained from the preceding profile mapping phase. The general idea is to perform iterative degradation of successive image blocks. Image blocks for degradation are selected in the order of increasing distortion impact D calculated using the distortion map $d(i, p)$ obtained according to (3.20):

$$D(i) \leftarrow d(i, q(i) - 1) - d(i, q(i)), \quad (3.21)$$

while respecting the content importance provided by ϕ , and the guard levels \mathbf{G} .

In order to speed up the design, the degradation is performed in batches of Δ_s image blocks. By increasing Δ_s it is possible to speed up the design, at the cost of obtaining a slightly better tampering rate than actually required. The complete procedure is presented in Algorithm 2.

3.2.2 Experimental Evaluation of the Design Procedure

The goal of this experiment is to validate that the quality descriptors designed by the presented algorithm indeed allows for reaching the specified target tampering

Algorithm 2 Quality descriptor design procedure

Require: $\phi, \tilde{\gamma}_{\text{target}}, \Delta_s, \mathbf{p}_g$ // User requirements

Require: q, d // Results of profile mapping

 $\Lambda, \mathbf{w} \leftarrow$ update from q
 $\tilde{\gamma} \leftarrow \tilde{\gamma} : \rho(\tilde{\gamma} | \Lambda, \mathbf{w}) = \frac{1-\tilde{\gamma}}{\lambda_{\text{ave}}}$

// Generate distortion impact map

 $D(i) \leftarrow d(i, q(i) - 1) - d(i, q(i))$ for $i \in \{1, \dots, N\}$

// Mark blocks which cannot be degraded due to importance map

 $D(i) \leftarrow \infty$ for $i : \phi(i) = 1$

// ... or due to reaching a guard level

 $D(i) \leftarrow \infty$ for $i : q(i) \in \mathbf{G}$
while $\tilde{\gamma} < \tilde{\gamma}_{\text{target}}$ **do**

 for $t \in \{1, \dots, \Delta_s\}$ **do**

 $i^* = \underset{i \in \{1, \dots, N\}}{\text{argmin}} \ D(i)$

 if $D(i^*) = \infty$ **then**

 return Design failed, constraint not met

 end if

 $q(i^*) \leftarrow q(i) - 1$

 if $q(i^*) \in \mathbf{G}$ **then**

 $D(i^*) \leftarrow \infty$

 else

 $D(i^*) \leftarrow d(i^*, q(i^*) - 1) - d(i^*, q(i^*))$

 end if

 end for
 $\Lambda, \mathbf{w} \leftarrow$ update from q
 $\tilde{\gamma} \leftarrow \tilde{\gamma} : \rho(\tilde{\gamma} | \Lambda, \mathbf{w}) = \frac{1-\tilde{\gamma}}{\lambda_{\text{ave}}}$
end while
return q

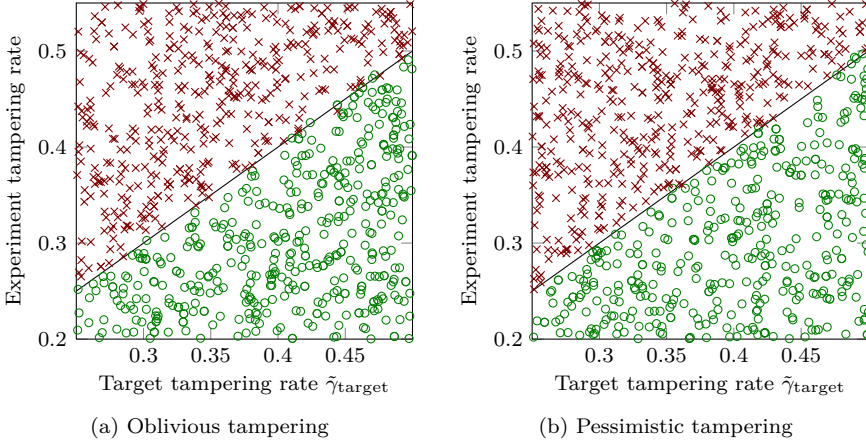


Fig. 3.5: Experimental validation of the descriptor design objective for 1,000 replications of the reconstruction process.

rates. For this purpose, I consider four reconstruction profiles with $\lambda_s = 1, 2, 3, 4$. Then, Algorithm 2 is used to calculate the descriptor for a randomly selected target tampering rate, drawn uniformly from $(0.25, 0.5)$. The self-embedding encoder produces a protected image with the reconstruction quality controlled by the obtained descriptor. The protected image is then randomly tampered, either in an oblivious or in a pessimistic manner. The tampering rate is drawn uniformly from $(0.2, 0.6)$. Finally, a reconstruction attempt is made.

It is expected that successful reconstruction should be possible when the tampering rate is lower than the target rate $\tilde{\gamma}_{target}$, which was used as the design objective. Fig. 3.5a and Fig. 3.5b show the obtained results for the oblivious, and the pessimistic tampering. Successful and unsuccessful reconstructions, marked with green circles and red crosses, respectively, are nearly perfectly separated by the expected tampering rate equality line.

Fig. 3.6 shows how the achievable tampering rates change with successive iterations of the design procedure. The average tampering rate behaves identically, regardless of the image content. Consistently with theoretical analysis, the pessimistic tampering rate starts to increase after a certain delay, and its growth is content-dependent. Two example graphs are shown in Fig. 3.6.

Example descriptors for image *4949.png* from the *bows* data-set are shown in Fig. 3.7. The figure presents successive quality descriptors, designed for increasing target tampering rates. The calculation was performed for oblivious tampering. (b-h) show the descriptors obtained without any importance mapping, i.e., when

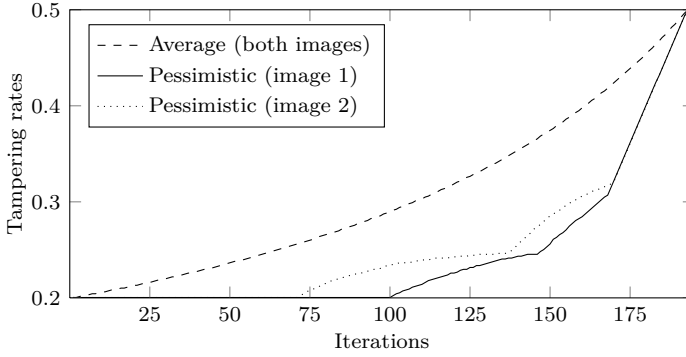


Fig. 3.6: Behavior of the achievable tampering rates for successive iterations of the design procedure on two different images.

any block can be degraded. (h-k) show the descriptors obtained with an example importance map (g) which ensures that the fish is always restored with the highest available quality. Given this requirement, it was impossible to obtain the target tampering rate $\tilde{\gamma} = 0.45$.

3.3 Conclusions

This chapter addressed the impact of incorporating multiple reconstruction profiles on the restoration conditions. It was demonstrated that the notion of the reconstruction demand, originally introduced in Chapter 2, can be used to accurately model the behavior of adaptive self-embedding schemes. The growth of the reconstruction demand is highly dependent on the observed tampering, which leads to fluctuations of the supported tampering rates along with the distribution of the modified content. Reconstruction profiles with high quality have dominant influence, and even adoption of lower fidelity profiles does not necessarily lead to improvement of the restoration conditions, despite shortening the reconstruction reference.

The derived theoretical dependencies can be exploited during calculation of the mapping between the reconstruction profiles and image blocks. The presented procedure takes into account the constraints on the local reconstruction quality, and the desired tampering rates. Hence, it is possible to provide guarantees for certain aspects of the reconstruction process, which is of high importance in certain applications. This functionality will be exploited for constructing an adaptive self-embedding scheme in Chapter 5.

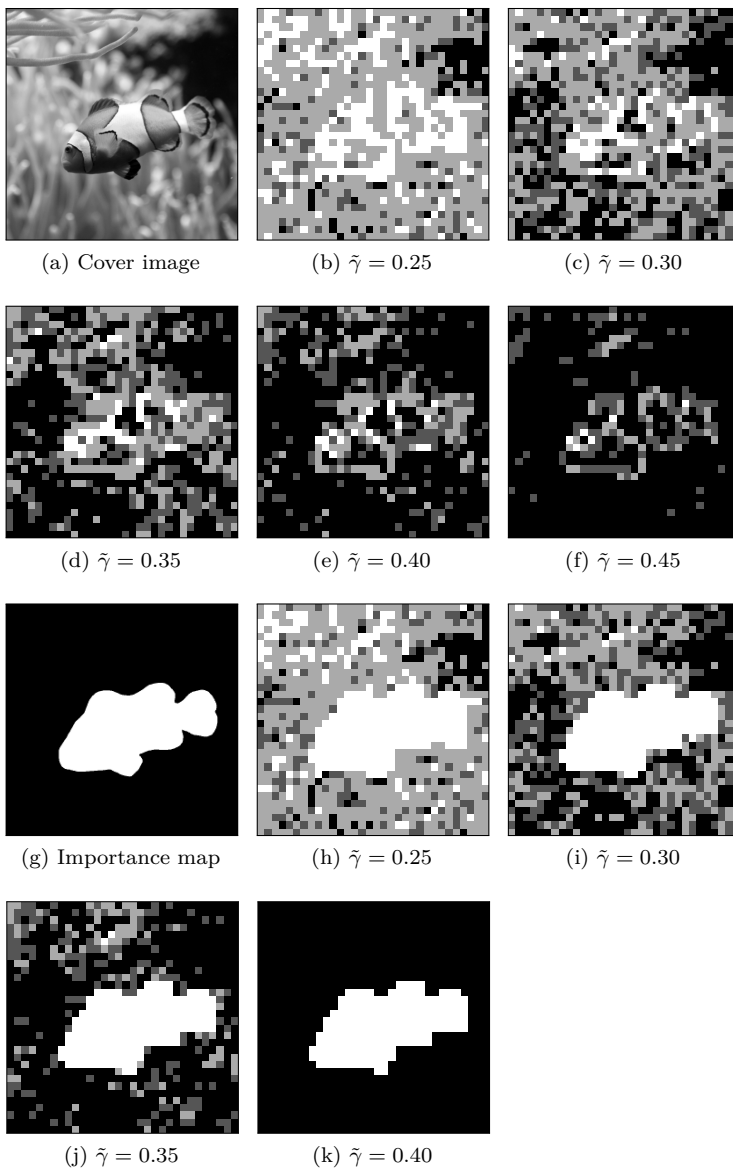


Fig. 3.7: Successive descriptors during the design procedure, both without (b-f) and with an importance map (h-k).

The discussed content reconstruction approach uses DCT coefficients of the principal image content to obtain its reference information. Hence, the reconstruction quality can be improved by optimizing the coefficient quantization procedure, and the precision of their representation. The latter is a typical resource allocation problem (Section 4.1). The goal is to determine the number of bits for the each of the coefficients, given a certain budget of the total number of bits. A dedicated algorithm for efficiently solving the problem is used, and the obtained results are validated against a general solver for nonlinear integer programming problems.

The quantization procedure can be optimized by replacing the commonly used uniform quantizer with a different code-book, better suited to the data. In this study, I consider a family of Lloyd-Max code-books [49], designed individually for groups of coefficients with similar distributions. The considered configuration is described in detail in Section 4.2.1, along with an exhaustive experimental evaluation of the achievable fidelity improvement.

The reconstruction quality can also be optimized by introducing content adaptivity. Its impact on the reconstruction success bounds was discussed in Chapter 3. The achievable improvement of reconstruction quality is assessed in Section 4.2.2. For this purpose, I design dedicated reconstruction profiles for blocks with low, medium, and high amounts of texture.

All of the described techniques are combined together in a fully functional self-embedding scheme in Chapter 5. Experimental evaluation is performed on popular image test sets: *sipi*, *ucid*, and *bows*. A sub-set of representative images from each of the sets serves as a training set for quantization code-book design, and distortion modeling. The impact of the discussed quality optimization techniques is illustrated on example test images (Fig. 4.1). For detailed information about the test sets, please refer to Appendix C.

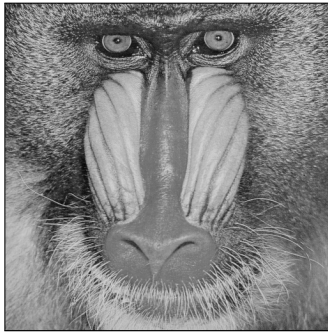
(a) *sipi* / baboon(b) *sipi* / lena(c) *sipi* / peppers(d) *bows* / 3879

Fig. 4.1: Sample images from the *sipi*, and the *bows* data-sets; printed at 300 dpi.

4.1 Reference Payload Allocation

The proposed self-embedding framework requires random access to the reference stream (Chapter 2). For this purpose, each block is represented by a previously known number of bits. The reconstruction reference generation procedure encodes a predefined set of spectrum coefficients, each represented with a given precision. The objective is to select the precision which maximizes the expected reconstruction quality within a given budget of available reference bits.

A formalized procedure for finding the optimal coefficient precisions is of high importance, as it allows content reconstruction schemes to be quickly adapted to various requirements. Such situation is particularly relevant for adaptive self-embedding, where multiple reconstruction profiles are used. Any variations in the desired number of reference bits, the characteristics of the content, or the quantization code-books can be efficiently, and reliably handled.

4.1.1 Formal Problem Statement

The problem is to find the optimal allocation of b bits to encode the coefficients of a 8×8 DCT spectrum, represented by an allocation matrix \mathbf{V} :

$$\mathbf{V} = \begin{bmatrix} v_{1,1} & v_{1,2} & v_{1,3} & v_{1,4} & v_{1,5} & v_{1,6} & v_{1,7} & v_{1,8} \\ v_{2,1} & v_{2,2} & v_{2,3} & v_{2,4} & v_{2,5} & v_{2,6} & v_{2,7} & v_{2,8} \\ v_{3,1} & v_{3,2} & v_{3,3} & v_{3,4} & v_{3,5} & v_{3,6} & v_{3,7} & v_{3,8} \\ v_{4,1} & v_{4,2} & v_{4,3} & v_{4,4} & v_{4,5} & v_{4,6} & v_{4,7} & v_{4,8} \\ v_{5,1} & v_{5,2} & v_{5,3} & v_{5,4} & v_{5,5} & v_{5,6} & v_{5,7} & v_{5,8} \\ v_{6,1} & v_{6,2} & v_{6,3} & v_{6,4} & v_{6,5} & v_{6,6} & v_{6,7} & v_{6,8} \\ v_{7,1} & v_{7,2} & v_{7,3} & v_{7,4} & v_{7,5} & v_{7,6} & v_{7,7} & v_{7,8} \\ v_{8,1} & v_{8,2} & v_{8,3} & v_{8,4} & v_{8,5} & v_{8,6} & v_{8,7} & v_{8,8} \end{bmatrix},$$

with elements $v_{i,j} \in \{0, 1, \dots, 8\}$ for $i, j \in \{1, \dots, 8\}$. The optimization objective is to minimize the mean squared error (MSE) on training images. Let $\{I_t : t = 1, \dots, N_t\}$ be a set of image blocks extracted from the training set. Hence, the objective can be written as:

$$\mathbf{V} = \underset{\mathbf{V}}{\operatorname{argmin}} \frac{1}{N_t} \sum_{t=1}^{N_t} \operatorname{MSE}(I_t, I_t^{(\operatorname{ref}|\mathbf{V})}),$$

where $I_t^{(\operatorname{ref}|\mathbf{V})}$ is the restored image block for a given allocation matrix \mathbf{V} .

Due to similar distribution of certain coefficients, it is possible to address the optimization problem in a simplified version. For this purpose, I consider groups of coefficients, defined by a constant sum of coefficients' coordinates, i.e., $i + j = \text{const}$. Coefficients within such groups are characterized by similar

distribution. Commonly used allocation matrices are highly symmetrical, and elements $\{v_{i,j} : i + j = k\}$ map to the same values for each k [33, 73]. Hence, the allocation can be represented by a vector \mathbf{v} :

$$\mathbf{v} = [v_1, v_2, \dots, v_{15}]. \quad (4.1)$$

which begins with a continuous block of non-zero elements, followed by zeroes. It is convenient to represent it using a shorter vector \mathbf{v}^+ , consisting of non-zero elements of \mathbf{v} . After appending a vector of zeros $\bar{\mathbf{0}}$, it gives the original vector \mathbf{v} :

$$\mathbf{v} = [\mathbf{v}^+ \ \bar{\mathbf{0}}]. \quad (4.2)$$

The allocation vector maps to the allocation matrix as follows:

$$\mathbf{V} = \begin{bmatrix} v_1 & v_2 & v_3 & v_4 & v_5 & v_6 & v_7 & v_8 \\ v_2 & v_3 & v_4 & v_5 & v_6 & v_7 & v_8 & v_9 \\ v_3 & v_4 & v_5 & v_6 & v_7 & v_8 & v_9 & v_{10} \\ v_4 & v_5 & v_6 & v_7 & v_8 & v_9 & v_{10} & v_{11} \\ v_5 & v_6 & v_7 & v_8 & v_9 & v_{10} & v_{11} & v_{12} \\ v_6 & v_7 & v_8 & v_9 & v_{10} & v_{11} & v_{12} & v_{13} \\ v_7 & v_8 & v_9 & v_{10} & v_{11} & v_{12} & v_{13} & v_{14} \\ v_8 & v_9 & v_{10} & v_{11} & v_{12} & v_{13} & v_{14} & v_{15} \end{bmatrix}. \quad (4.3)$$

The number of occurrences of the components of \mathbf{v} can be represented by a helper column vector \mathbf{a} :

$$\mathbf{a} = [1, 2, 3, 4, 5, 6, 7, 8, 7, 6, 5, 4, 3, 2, 1]^T. \quad (4.4)$$

The resource allocation can be formulated in terms of integer nonlinear programming (INLP):

$$\min \theta(\mathbf{v}), \quad (4.5a)$$

$$\text{s.t. } \forall_{k \in \{1, \dots, 15\}} v_k \geq 0, \quad (4.5b)$$

$$\forall_{k \in \{1, \dots, 15\}} v_k \leq 8, \quad (4.5c)$$

$$\mathbf{v} \in \mathbf{Z}^{15}, \quad (4.5d)$$

$$\mathbf{v} \cdot \mathbf{a} = b. \quad (4.5e)$$

where $\theta(\cdot)$ is a non-linear objective cost function, which represents the average distortion for a specific allocation vector \mathbf{v} , or allocation matrix \mathbf{V} . For each component v_k , corresponding to a coefficient group k , a distortion model is built, for quick assessment of the expected distortion. Conditions (4.5b)-(4.5c) ensure that the precision is in the range 0-8 bits, and (4.5b) constrains the solution to integers only. The payload budget is guaranteed by (4.5e).

Due to the properties of the objective function, the optimization problem can be solved more efficiently with a dedicated algorithm, described in detail in Section 4.1.3. A general solver for INLP problems is also used to validate the obtained results. Due to high computational complexity, the INLP-solver operates on the simplified, 15-element version of the optimization problem. The proposed algorithm solves a full, 64-element version of the problem.

4.1.2 Modeling the Objective Distortion

The objective cost function $\theta(\mathbf{v})$ represents the average distortion for a single image block given a specific allocation vector \mathbf{v} . The DCT transformation is orthonormal, and it is possible to calculate the MSE directly in the DCT domain. The MSE can be represented by a distortion vector $\mathbf{d}(\mathbf{v}) = [d_1(v_1), \dots, d_{15}(v_{15})]$, and calculated as:

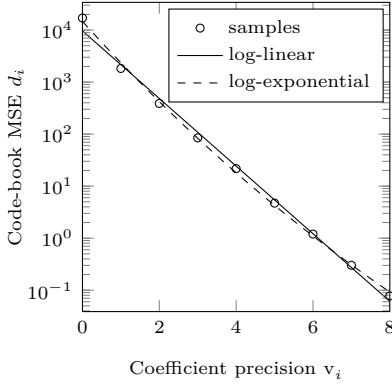
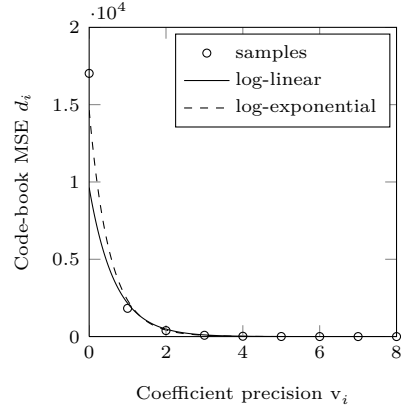
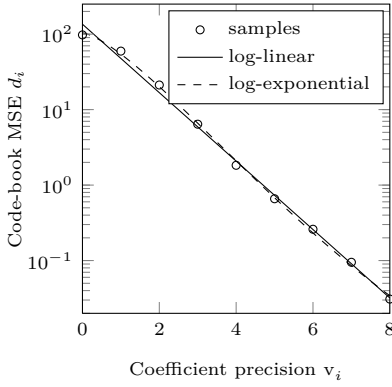
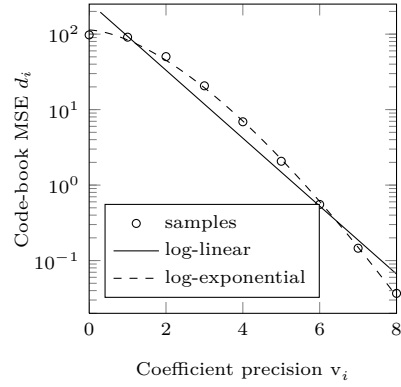
$$\theta(\mathbf{v}) = \mathbf{d}(\mathbf{v}) \cdot \mathbf{a} / 64. \quad (4.6)$$

Each component $d_k(v_k)$ represents the MSE for the coefficients within group k . The distortion changes with the precision of the coefficients v_k . While $v_k \in \{0, 1, \dots, 8\}$, it is actually necessary to allow for continuous values from the range $[0, 8]$. Existing INLP solvers are based on traditional solvers for general NLP, and arbitrary floating point values are used during intermediate calculations. As a result, I consider two *distortion models* to describe this dependency: a log-exponential, and a log-linear model. Individual components d_k of the vector are represented by:

$$\ln d_k(v_k) = \begin{cases} \vartheta_{1,k}^{(e)} e^{-\vartheta_{2,k}^{(e)} v_k^{\vartheta_{3,k}^{(e)}}} + \vartheta_{4,k}^{(e)}, & \text{for the log-exponential model,} \\ \vartheta_{1,k}^{(l)} + v_k \vartheta_{2,k}^{(l)}, & \text{for the log-linear model.} \end{cases} \quad (4.7)$$

where $\vartheta_{l,k}^{(e)}$ and $\vartheta_{l,k}^{(l)}$ are the parameters of the models, obtained by fitting to average distortions observed on training images, in logarithmic scale. The considered quantization code-books have $1, 2^1, \dots, 2^8$ values; the 1-value quantizer always assumes the value 0. Fig. 4.2 shows example distortion models for components v_1 and v_2 . The use of the logarithmic scale is motivated by insufficient accuracy of linear-scale models. For low-frequency coefficient groups, e.g., $k = 1, 2, 3$, the low-precision distortions become disproportionately large and it is not possible to obtain a reliable fit for higher precisions.

The presented distortion models are used also for the full 64-element optimization performed by the dedicated solver. For each individual coefficient, its distortion is modeled in the same way, as for the other coefficients from its group. The solver, however, is not constrained to produce symmetrical allocation matrices. Such an approach facilitates easier verification of the proposed solver.

(a) Component v_1 , uniform code-book(b) Component v_1 , uniform code-book(c) Component v_2 , Lloyd-Max code-book(d) Component v_2 , uniform code-bookFig. 4.2: Example fitted distortion model for components v_1 , and v_2 .

Distortion Model Training

The distortion models, used for construction of the objective function, are obtained during a training phase, along with the corresponding quantization code-books. The training phase consists of the following steps:

1. Select a training set of gray-scale images, which best represent the prospective content; the set should cover the whole space of possible image blocks, i.e., dark, bright, highly-textured, flat, etc.
2. Discard L least significant bit-planes.
3. Calculate a block-wise DCT, and collect individual statistics of the transform coefficients for each of the v_1, \dots, v_{15} components.
4. For each component v_k , design quantization code-books with 1, 2, 4, \dots , 256 values, and record the MSE.
5. Calculate the log-linear and log-exponential distortion models by fitting (4.7) to the obtained MSE for each of the v_1, \dots, v_{15} components.

4.1.3 Solving the Reference Payload Allocation Problem

The character of the objective function, makes it possible to simplify the algorithm for solving the optimization problem (4.5). The impact of each individual component on the final objective distortion is strictly monotonic. Hence, when considering promotion of an individual component, it is certain that the highest achievable improvement for the component at hand is directly available. The process begins from a zeroed allocation vector. It then evaluates the prospective improvement for each of the components, which would be obtained by increasing its precision by 1. The component having the greatest possible impact on the final objective function is then selected, and promoted to a higher precision. Then, its prospective impact is updated for the next iteration of the procedure. The promotion loop follows until all of the bits are successfully allocated.

Such an algorithm is less computationally demanding than using a general INLP solver. It becomes feasible to perform the procedure for a full 64-component version of the problem. Moreover, it can be easily extended to take into account the minimal allowed coefficient precision v_{\min} . This allows for instance for automatically preventing the use of undesirable 1-bit code-books. In order to take it into account, the algorithm starts by evaluating a normalized impact of promoting the coefficients from 0 to v_{\min} . After normalization, the impact is expressed by means of improvement per single promotion unit. When chosen, the promotion is performed with v_{\min} bits at once. The complete procedure is presented in Algorithm 3.

Analogous extension of the optimization problem is straightforward also for INLP. From a formal problem statement perspective, the issue can be resolved by introducing stub variables, and adding a few new constraints, e.g., for a minimum precision of 2 bits the problem becomes:

$$\begin{aligned}
 & \min \theta(\mathbf{v}), \\
 \text{s.t. } & \forall_{k \in \{1, \dots, 15\}} v_i \geq 0 \\
 & \forall_{k \in \{1, \dots, 15\}} v_k \leq 8 \\
 & \forall_{k \in \{1, \dots, 15\}} v_k + 8u_k \leq 8. \\
 & \forall_{k \in \{1, \dots, 15\}} v_k + 2u_k \geq 2. \\
 & \forall_{k \in \{1, \dots, 15\}} u_k \in \{0, 1\} \\
 & \mathbf{v} \in \mathbf{Z}^{15}, \\
 & \mathbf{v} \cdot \mathbf{a} = b.
 \end{aligned} \tag{4.8}$$

Unfortunately, the problem becomes excessively complex for existing mixed integer nonlinear programming (MINLP) solvers, which cannot efficiently handle more than a dozen of integer variables. Hence, in this study I consider the original formulation from (4.5).

The INLP-based optimization is used for validation of the results of the proposed optimization procedure. The utilized solver is a branch and bound solver [60] of mixed-integer nonlinear programming problems for Matlab [41]. In contrast to the proposed algorithm, the generic solver requires a valid starting point. To provide further verification of the convergence, I consider two carefully chosen starting points, stemming from the use of both distortion models.

The log-exponential distortion model is more accurate than the log-linear model. At the same time, it is susceptible to the presence of local extrema of the objective function. Depending on the starting point, the solver might yield slightly different results. I have experimentally established that the best results are typically obtained with a uniform starting point, generated according to Algorithm 4.

The simpler log-linear model is not susceptible to the presence of local extrema and for all of the considered starting points, the solver yielded the same results. The resulting solution can then be used as a starting point for optimization with the log-exponential distortion model, in order to confirm the convergence of the uniformly-initialized optimization. In a few cases, the results have been noticed to differ, yet with no statistically significant differences in the objective function.

The final procedure for generating the allocation vector using the generic solver follows as:

1. Solve (4.5) with the log-linear distortion model starting from a uniform \mathbf{v}_0 . Once successful, the procedure yields an allocation vector \mathbf{v}_{lin} .

Algorithm 3 Allocation matrix calculation procedure

Require: reference payload b , distortion model $\theta(\cdot)$, minimum precision v_{\min}

```

v  $\leftarrow$  0 // Current allocation vector
d  $\leftarrow$  0 // Prospective promotion impact
// Calculate prospective promotion impact
for  $i = 1 \rightarrow 64$  do
  v(t)  $\leftarrow$  v // Temporary variable
   $v_i^{(t)} \leftarrow v_{\min}$ 
   $d_i \leftarrow (\theta(\mathbf{v}^{(t)}) - \theta(\mathbf{v})) / v_{\min}$ 
end for
while  $\sum_{i=1}^{64} v_i < b$  do
   $j = \underset{i}{\operatorname{argmax}} d_i$ 
  // If budget would be exceeded, skip and discard coefficient
  if  $v_j = 0$  and  $\sum_{i=1}^{64} v_i \geq b - v_{\min} + 1$  then
     $d_j = 0$ 
    continue
  end if
  // If no further improvement possible, report failure
  if  $d_j = 0$  then
    return Allocation failed
  end if
  if  $v_j = 0$  then
     $v_j \leftarrow v_j + v_{\min}$ 
  else
     $v_j \leftarrow v_j + 1$ 
  end if
  // If reached max. precision, discard the coefficient
  // from further promotions, otherwise update its impact
  if  $v_j = 8$  then
     $d_j \leftarrow 0$ 
  else
    v(t)  $\leftarrow$  v
     $v_j^{(t)} \leftarrow v_j^{(t)} + 1$ 
     $d_j \leftarrow \theta(\mathbf{v}^{(t)}) - \theta(\mathbf{v})$ 
  end if
end while
return v

```

Algorithm 4 Uniform starting point generation

Require: b
 $\mathbf{v} \leftarrow \mathbf{0}$
 $i \leftarrow 0$
while $\mathbf{v}\mathbf{a} \leq b$ **do**
 $i \leftarrow i \bmod 15 + 1$
 $v_i \leftarrow v_i + 1$
end while
 $v_i \leftarrow v_i - 1$
while $\mathbf{v}\mathbf{a} - b < 0$ **do**
 $i \leftarrow b - \mathbf{v}\mathbf{a}$
 $v_i \leftarrow v_i + 1$
end while
return \mathbf{v}

2. Solve (4.5) with the log-exponential distortion model, starting from a uniform \mathbf{v}_0 . Once successful, the procedure yields an allocation vector \mathbf{v}_{exp} .
3. Solve (4.5) with the log-exponential distortion model, starting from the solution of the linear model \mathbf{v}_{lin} . Once, successful, the procedure yields an allocation vector \mathbf{v}_{ref} .
4. Based on the objective function, choose either \mathbf{v}_{exp} or \mathbf{v}_{ref} .

4.1.4 Optimization Results

The obtained solutions to the reference payload allocation problem (4.5) for selected configurations of the reconstruction procedure are collected in Appendix D in Tables D.1-D.4. The left columns show the results yielded by the proposed optimization procedure, while the right show the corresponding results from the INLP solver. The obtained results are nearly identical. The only differences stem from the additional flexibility of the dedicated algorithm, which operates on a full 64-element version of the problem. It allows for further slight improvement of the objective cost function.

The tables also show intermediate results from the INLP solver. Namely, they present the allocation vectors \mathbf{v}_{exp} , \mathbf{v}_{lin} , and \mathbf{v}_{ref} , obtained with various starting points. In most cases, the pre-initialized execution still converges to \mathbf{v}_{exp} . Conversely, the differences between the solutions are infinitesimal, and can be safely disregarded.

In general, the optimal solution slightly changes both with the training set, and the quantization code-book. The changes are in perfect correspondence with

the expected behavior, and will be discussed in detail in the successive parts of this chapter. Section 4.2.1 investigates how the optimal allocation changes with the adopted quantization code-book. The influence of the amount of texture is elaborated in Section 4.2.2.

4.2 Reconstruction Fidelity Improvement Techniques

This section presents an extensive assessment of the reconstruction fidelity improvement that can be achieved with two simple techniques. Firstly, I discuss the impact of replacing the uniform quantizer with a tailored Lloyd-Max quantizer. Secondly, I introduce multiple reconstruction profiles, dedicated to image blocks with various content. For all of the performed experiments, the necessary allocation matrices were obtained with the optimization algorithm presented in Section 4.1. The section ends with example reference images, obtained with all of the discussed techniques.

4.2.1 Quantization Strategy Impact

Distortion of the reference image results from a confined number of encoded spectrum coefficients, and limited precision of their representation. A proper choice of the quantization strategy allows for significant improvement of the restoration fidelity, compared to a standard, uniform quantizer. In this study, I consider the Lloyd-Max code-book [49], which is optimal in the sense of the MSE, i.e., it minimizes the expected distortion with respect to the L^2 metric. Assessment of the achievable reconstruction quality improvement is the purpose of this experiment.

The designed scheme uses two types of quantization code-books: uniform and Lloyd-Max. Component v_1 , which corresponds to the DC coefficient, is always represented with a uniform code-book. Quantization of the remaining AC coefficients is performed according to the configuration of the system. Fig. 4.3a shows a histogram of the AC coefficients corresponding to the v_2 component, and a MLE-fitted generalized Gaussian distribution (GGD) with mean 0.00, scale 0.35, and shape parameter 0.40. A corresponding Lloyd-Max code-book is shown in Fig. 4.3b.

Table 4.1 presents the measured distortion for all of the considered data-sets for $b = 40, 80, 160$, and 240 bits per block. $L = 2$ least significant bit-planes were discarded during pre-processing. The reported ranges correspond to typical scores, obtained by eliminating the outliers from the empirical distribution, which fall outside the range:

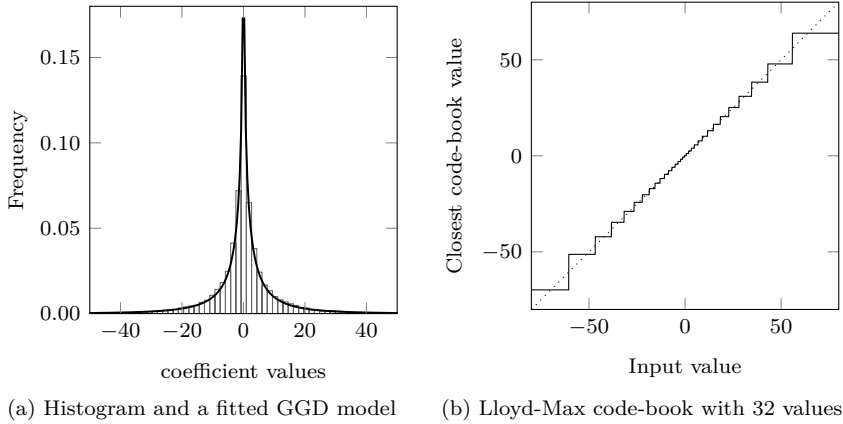


Fig. 4.3: Distribution of the coefficients and a corresponding Lloyd-Max quantization code-book for component v_2 .

$$[q_1 - c_q(q_3 - q_1) ; q_3 + c_q(q_3 - q_1)], \quad (4.9)$$

where q_1 and q_3 denote the 25th and the 75th percentiles, respectively, and c_q is a scaling factor, chosen empirically in order to discard approximately 1% of the samples as outliers.

The differences in the achievable scores ΔPSNR are collected in Table 4.2. ΔPSNR is calculated as:

$$\Delta\text{PSNR} = \text{PSNR}(I, I^{(\text{ref},e)}) - \text{PSNR}(I, I^{(\text{ref},o)}) \quad (4.10)$$

where $I^{(\text{ref},e)}$ and $I^{(\text{ref},o)}$ represent the prospectively improved, and the original reference images, respectively. In this experiment, the former corresponds to the Lloyd-Max, and the latter to uniform quantization.

In general, it can be observed that quantization with the Lloyd-Max code-book indeed leads to better restoration fidelity. For reference payload $b = 40$ bits per block, the average improvement in terms of PSNR is 0.6 dB, and increases along with b to reach 1.6-2.0 dB for $b = 240$ bpb.

Despite the overall positive impact, closer examination of the obtained scores shows that in some cases the reconstruction fidelity actually deteriorates. Fig. 4.4 shows the histogram of PSNR differences ΔPSNR obtained on 10,000 images in the *bows* test-set. The distribution is asymmetric with a small portion of negative values. The exact amount of negative scores depends both on the content of the images, and on the reference rate. It tends to be higher for lower rates.

Table 4.1: Reconstruction quality for the uniform, and the Lloyd-Max code-books. The reported distortions correspond to the case of $L = 2$, and are expressed in PSNR [dB].

Image / Set	Uniform code-book			Lloyd-Max code-book		
	Mean	Median	Range	Mean	Median	Range
Reference payload $b = 40$ bpb						
baboon	22.4	-	-	23.2	-	-
lena	30.5	-	-	31.4	-	-
peppers	30.1	-	-	30.6	-	-
<i>sipi</i> test-set	28.4	27.9	24.2 - 33.1	29.0	28.7	23.2 - 33.2
<i>ucid</i> test-set	26.7	26.5	16.5 - 37.0	27.3	27.2	17.1 - 37.1
<i>bows</i> test-set	29.8	29.6	20.2 - 39.4	30.5	30.4	22.8 - 38.3
Reference payload $b = 80$ bpb						
baboon	24.7	-	-	25.7	-	-
lena	33.0	-	-	34.5	-	-
peppers	32.4	-	-	33.5	-	-
<i>sipi</i> test-set	30.8	30.6	24.7 - 35.0	31.9	31.8	25.7 - 36.4
<i>ucid</i> test-set	29.5	29.4	18.7 - 40.2	30.5	30.4	18.7 - 41.9
<i>bows</i> test-set	32.2	32.1	24.0 - 40.6	33.9	33.8	25.4 - 42.4
Reference payload $b = 160$ bpb						
baboon	28.5	-	-	30.1	-	-
lena	36.9	-	-	38.8	-	-
peppers	35.6	-	-	37.2	-	-
<i>sipi</i> test-set	34.6	34.8	29.3 - 38.7	36.2	37.2	30.1 - 40.4
<i>ucid</i> test-set	33.9	33.9	22.2 - 43.9	35.2	35.4	21.2 - 43.9
<i>bows</i> test-set	36.7	36.7	29.1 - 44.4	38.6	38.8	31.6 - 45.1
Reference payload $b = 240$ bpb						
baboon	33.8	-	-	34.8	-	-
lena	38.9	-	-	42.2	-	-
peppers	37.3	-	-	40.7	-	-
<i>sipi</i> test-set	37.4	37.4	33.8 - 39.9	39.9	41.0	23.8 - 43.0
<i>ucid</i> test-set	37.6	37.8	24.1 - 45.2	39.2	39.9	23.8 - 44.5
<i>bows</i> test-set	39.8	39.8	34.5 - 45.1	41.9	42.2	36.8 - 46.5

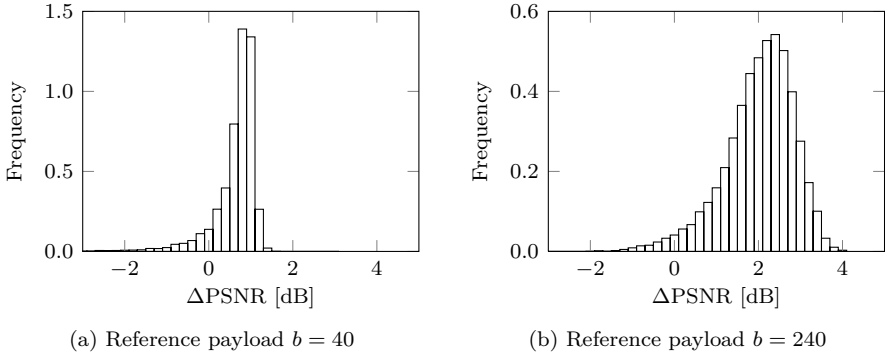


Fig. 4.4: Reconstruction quality improvement due to adoption of Lloyd-Max quantization; Δ PSNR for 10,000 images from the *bows* data-set.

The highest observed image deterioration rate is 8% for the *bows* test-set and $b = 40$. Further examination shows that most of the problematic images are extremely flat, with large solid-color areas. Fig. 4.5 shows 4 images from the *bows* data-set, which experienced the highest quality loss. Two main causes of this quality loss can be identified. First, the images depict content with statistics different from the samples from the training set. Secondly, by comparing the allocation vectors \mathbf{v}^+ :

$$\mathbf{v}^+ = \begin{cases} [5, 4, 3, 2, 2], & \text{for the Lloyd-Max code-book,} \\ [6, 5, 4, 3], & \text{for the uniform code-book,} \end{cases} \quad (4.11)$$

it can be observed that in case of the uniform code-book, the precision of the DC coefficient is higher. While for textured images such assignment is beneficial, for flat images the differences in the average block intensities become a dominant component of the overall distortion. If necessary, the problem can be easily solved by enforcing a minimum precision of the DC coefficient, at the cost of deteriorating the average reconstruction quality.

The problem is applicable mainly for small reference payloads. The differences in the MSE for higher precisions, e.g., 8 vs. 7, quickly become negligible. In fact, for higher b , the improvement introduced by the Lloyd-Max code-book becomes even more prominent (Table 4.2). Not only does the fidelity gain improve, but also the number of images with quality loss drops. Better fidelity of the code-book allows for encoding more coefficients. It is also possible for the solve to assign higher precision of the DC coefficient to the Lloyd-Max variant, e.g., for reference

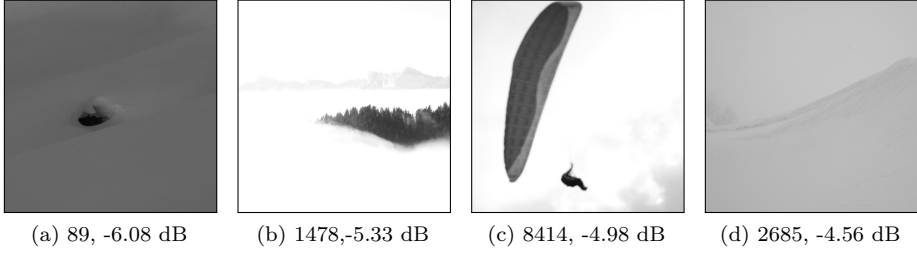


Fig. 4.5: Images from the *bows* data-set with the highest negative impact of replacing the uniform with the Lloyd-Max quantizer.

Table 4.2: Quality improvement due to the adoption of Lloyd-Max quantization.

Parameter		Reference payload [bpb]			
		40	80	160	240
<i>ucid</i> test-set					
Mean Δ PSNR	[dB]	0.56	0.96	1.29	1.62
Median Δ PSNR	[dB]	0.59	0.96	1.36	1.63
Min typical Δ PSNR	[dB]	-0.56	-0.67	-1.04	-0.63
Max typical Δ PSNR	[dB]	1.44	2.37	2.87	3.93
Images with quality loss	[%]	4.86	5.08	5.53	4.78
<i>bows</i> test-set					
Mean Δ PSNR	[dB]	0.67	1.65	1.85	2.04
Median Δ PSNR	[dB]	0.79	1.68	1.97	2.13
Min typical Δ PSNR	[dB]	-1.59	0.24	-0.20	-0.48
Max typical Δ PSNR	[dB]	1.67	3.04	3.15	4.45
Images with quality loss	[%]	8.05	0.00	0.46	1.54

payload $b = 80$:

$$\mathbf{v}^+ = \begin{cases} [7, 4, 4, 3, 3, 2, 2], & \text{for the Lloyd-Max code-book.} \\ [6, 5, 5, 4, 3, 3], & \text{for the uniform code-book.} \end{cases} \quad (4.12)$$

More optimization results are presented in Appendix D. Tables D.1 and D.1 compare the obtained allocation matrices for the uniform and the Lloyd-Max code-books. Example reference images for both code-books are collected in Section 4.2.3.

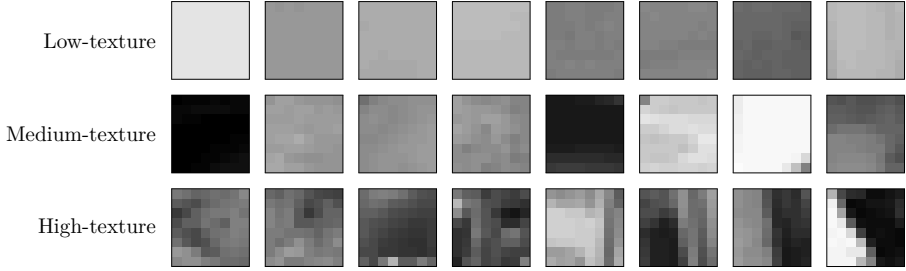


Fig. 4.6: Sample image blocks from the considered block classes; the samples are collected from the *bows* training-set, and span the whole range of values of σ .

4.2.2 Content Adaptivity Impact

Incorporation of multiple reconstruction profiles allows not only for differentiation between the desired quality levels, but also for their optimization to various content characteristics. For a given b , both the allocation matrix, and the quantization code-books could potentially differ for image blocks with distinct texture. The purpose of this experiment is to assess the prospective improvement of the restoration fidelity, both in terms of the introduced distortion, and the subjective impact perceived by a human observer. To that end, I consider 3 block classes: low, medium, and high-texture blocks, classified based on their standard deviation σ :

$$\begin{cases} \text{low-texture block,} & \text{if } 2^L \sigma_i \leq 5, \\ \text{medium-texture block,} & \text{if } 5 < 2^L \sigma_i \leq 20, \\ \text{high-texture block,} & \text{if } 2^L \sigma_i > 20. \end{cases} \quad (4.13)$$

where σ_i denotes the standard deviation for the i -th image block.

The thresholds were selected by visual inspection of the resulting block sets. Fig. 4.6 shows samples image blocks from the considered classes. The blocks are ordered by increasing σ , and span the whole range of possible values. The optimization process is identical to the single-profile case, and it is simply repeated separately for each of the classes.

The fundamental concept of the discussed technique is to provide additional reconstruction profiles, adapted to various block characteristics. During reconstruction reference generation, the encoder selects the reconstruction profile, which yields minimal distortion with respect to the original image blocks. Table 4.3 collects the improvement statistics from the *ucid* and *bows* data-sets. A histogram of PSNR differences ΔPSNR for the reference payloads $b = 40$ and $b = 160$ bpb is shown in Fig. 4.7. The achievable improvement averages between 0.5 dB and 1.2 dB. However, for some of the images, quality loss can actually been observed.

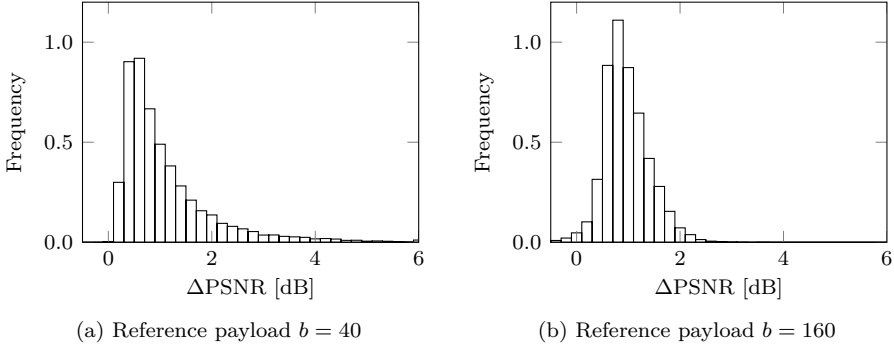


Fig. 4.7: Histogram of ΔPSNR between the single-profile, and the 3-profile configurations; quantization with Lloyd-Max code-books; obtained on 10,000 images from the *bows* data-set.

Most of the deteriorated images occurred for one particular configuration of the process, i.e., for $b = 160$ bpb, where approximately 1% of the images experienced fidelity loss. In the remaining cases quality loss was incidental, or did not occur at all.

Tables 4.4 and 4.5 collect some of the obtained reference allocation matrices for all of the considered block classes. For low-texture blocks, the solver assigns higher precision to the DC coefficient, which effectively reduces the blocking artifacts. At the same time, low-frequency coefficients can be represented with less precision, and as a result more coefficients can actually be encoded. Such behavior is intuitively clear, and consistent with the theoretically expected one. However, the obtained results clearly show that, in general, some image blocks are better represented with a globally determined coefficient precision. Hence, in practice it might be beneficial to include a general reconstruction profile, in addition to the dedicated ones.

Table 4.3: Reconstruction quality improvement due to adoption of three reconstruction profiles for low, medium, and high texture blocks.

Parameter		Reference payload [bpb]			
		40	80	160	240
<i>ucid</i> data-set					
Mean Δ PSNR	[dB]	0.53	0.54	0.67	0.99
Median Δ PSNR	[dB]	0.42	0.49	0.63	0.98
Min typical Δ PSNR	[dB]	0.01	-0.13	-0.42	0.30
Max typical Δ PSNR	[dB]	2.24	1.34	1.86	1.67
Images with quality loss	[%]	0.00	0.15	1.20	0.00
<i>bows</i> data-set					
Mean Δ PSNR	[dB]	1.05	0.82	0.95	1.19
Median Δ PSNR	[dB]	0.80	0.71	0.90	1.18
Min typical Δ PSNR	[dB]	0.02	-0.45	-0.33	0.47
Max typical Δ PSNR	[dB]	4.58	2.40	2.22	1.90
Images with quality loss	[%]	0.00	0.08	0.82	0.00

4.2.3 Example Reference Images

This section presents example reference images, obtained with the use of the described fidelity optimization techniques. Each image fragment is reconstructed exactly as it appears in the reference image. The images were generated with reference payload $b = 40$ bpb, and with $L = 2$ discarded least significant bit-planes. Allocation of the coefficient precision was performed according to Section 4.1.3. In order better visualize the restoration artifacts, and clearly present the differences between the images, only a clipped fragment is shown, and the print resolution is set to 72 dpi.

For each test image, 3 reference images are shown. The first one is obtained using a uniform quantizer, and a single general reconstruction profile. The remaining two are obtained using the Lloyd-Max quantizer, and both one general, and three dedicated reconstruction profiles. Compared to the uniformly quantized version, the Lloyd-Max code-book produces stronger edges, and allows for restoration of higher amount of details. Introduction of multiple reconstruction profile further improves the quality, mainly by reducing blocking artifacts.



(a) baboon, original



(b) baboon, uniform quantizer, 22.44 dB

Fig. 4.8: Reference images for the *baboon* image from the *sipi* test set (*cont.*).



(c) baboon, Lloyd-Max quantizer, 1 general profile 23.23 dB



(d) baboon, Lloyd-Max quantizer, 3 dedicated profiles, 23.41 dB

Fig. 4.8: Reference images for the *baboon* image from the *sipi* test set.



(a) lena, original



(b) lena, uniform quantizer, 30.54 dB

Fig. 4.9: Reference images for the *lena* image from the *sipi* test set (*cont.*).

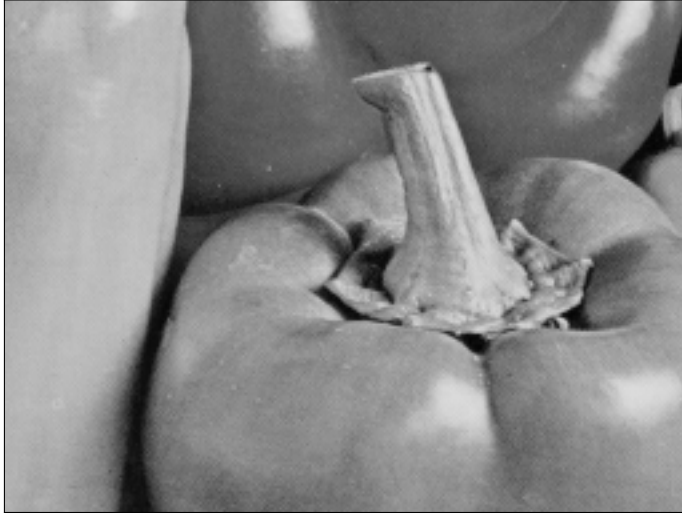


(c) lena, Lloyd-Max quantizer, 1 general profile, 31.40 dB



(d) lena, Lloyd-Max quantizer, 3 dedicated profiles, 32.24 dB

Fig. 4.9: Reference images for the *lena* image from the *sipi* test set.



(a) peppers, original



(b) peppers, uniform quantizer, 30.11 dB

Fig. 4.10: Reference images for the *peppers* image from the *sipi* test set (*cont.*).



(c) peppers, Lloyd-Max quantizer, 1 general profile, 30.63 dB



(d) peppers, Lloyd-Max quantizer, 3 dedicated profiles, 31.43 dB

Fig. 4.10: Reference images for the *peppers* image from the *sipi* test set.



(a) 3879, original



(b) 3879, uniform quantizer, 27.256 dB

Fig. 4.11: Reference images for the 3879 image from the *bows* test set (*cont.*).



(c) 3879, Lloyd-Max quantizer, 1 general profile, 28.52 dB



(d) 3879, Lloyd-Max quantizer, 3 dedicated profiles, 29.05 dB

Fig. 4.11: Reference images for the 3879 image from the *bows* test set.

This chapter describes the construction of practical self-embedding schemes, based on the proposed content reconstruction model. The schemes are intended for the reconstruction of maliciously tampered content of gray-scale loss-less digital images. Efficient handling of lossy-compressed color content is out of scope of this study. A brief summary of the necessary modifications is described in Appendix E.

I consider two classes of self-embedding schemes: traditional, with uniformly distributed reconstruction fidelity; and adaptive, which allows for selection of the best reconstruction settings from a set of predefined reconstruction profiles. The analysis presented in this chapter focuses on extensive evaluation of the reconstruction performance, and comparison with state-of-the-art alternative schemes. The comparison is performed using a reference scheme, introduced in Section 2.2.1. Just as most of alternatives, the reference scheme uses the same fidelity for the whole image area, which facilitates fair comparison of the reconstruction approach.

The second aspect of the presented evaluation is related to the reconstruction efficiency, i.e., the dependency between the reconstruction quality, and the achievable tampering rates. The issue is first assessed for the traditional, reference scheme, and then for the adaptive schemes, which have more flexibility in controlling the trade-off at hand.

5.1 Adaptive Self-Embedding Scheme

Operation of the adaptive self-embedding scheme is analogous to the reference scheme, but uses the quality descriptor to control the reference generation, and restoration procedures. The quality descriptor is designed in the encoder, and then communicated to the decoder through the same channel as the reconstruction reference.

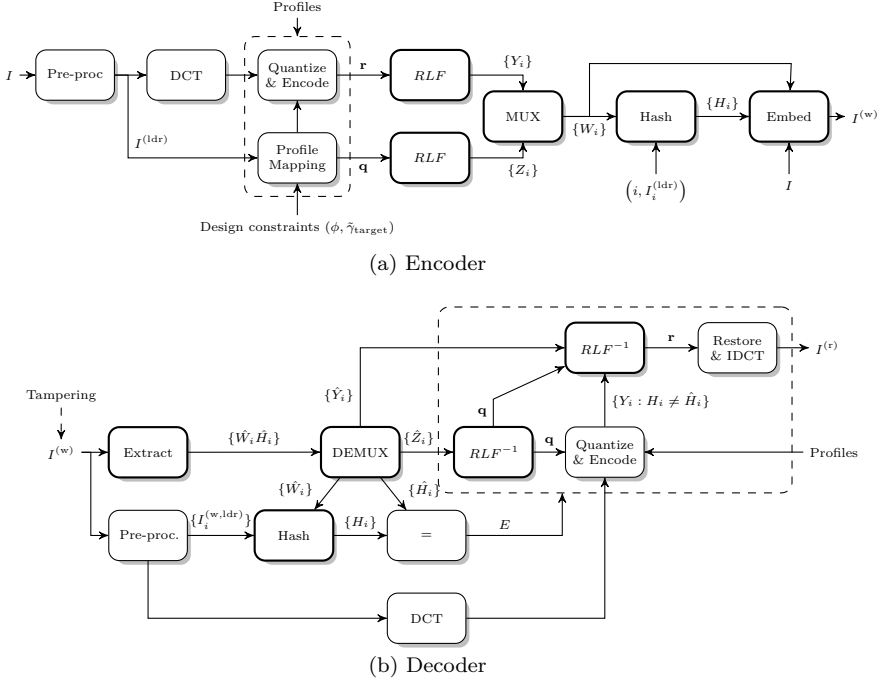


Fig. 5.1: Operation of the adaptive self-embedding scheme; I - cover image, $I^{(w)}$ - protected, watermarked image, $I^{(r)}$ - recovered, authenticated image, E - tampering map, \mathbf{q} - bit-stream of the quality descriptor; Operations directly controlled by the security context κ are marked by thick borders.

Two variants of the scheme are considered:

- *content-adaptive*, with 4 reconstruction profiles, and constant reference rates, i.e., $S = 4$, and $b = \text{const}$;
- *descriptor-adaptive*, with 8 reconstruction profiles, and variable reference rates, i.e., $S = 8$, and $b = b(i)$.

The distinction is motivated by the impact of the quality descriptor on the reconstruction success bounds. In the content-adaptive scheme, all profiles are characterized by the same reference rate. Hence, the reconstruction demand is not affected by any changes in the quality descriptor (Chapter 3).

The descriptor-adaptive scheme uses reconstruction profiles with various reference rates. Hence, the reconstruction success bound changes with the profile

assignment. By proper design of the descriptor, it is possible to guarantee a given target tampering rate.

Methods for Image Encoding

Operation of the encoder is illustrated in Fig. 5.1a. The procedure begins with a pre-processing step, which limits the dynamic range of the original image by discarding L least significant bit-planes, i.e., $I^{(\text{ldr})} = \lfloor I/2^L \rfloor$. This step is followed by block-wise computation of the DCT spectrum.

The following reconstruction reference generation is controlled by a quality descriptor, designed to optimize the reconstruction quality, given prospective external requirements. The details of this step are described in Section 3.2. Based on the assigned reconstruction profile, the encoder uses the corresponding allocation matrix and quantization code-books to obtain the variable length reference:

$$\mathbf{r} = \{r_i\} = \{g_{b(i)}(I_i^{(\text{ldr})})\}. \quad (5.1)$$

where $g_b(\cdot)$ is the reconstruction reference generation function, which produces exactly b bits.

Both the quality descriptor, and the reconstruction reference are communicated to the decoder through an erasure channel. Let λ_{qd} denote the rate of the fountain-encoded quality descriptor. The communication model in the former case corresponds to a generic erasure transmission [51], while in the latter to the self-recovery transmission (Chapter 2). To that end, both \mathbf{q} and \mathbf{r} are divided into constant-length descriptor symbols $Q_1, \dots, Q_{N\lambda_{\text{qd}}}$, and reference symbols X_1, \dots, X_K , respectively.

Both $\{X_k\}$, and $\{Q_j\}$ are then encoded with RLF to produce embedding symbols Y_1, \dots, Y_N , and Z_1, \dots, Z_N , which are then combined by a multiplexer (MUX) to yield the embedding payload for the i -th image block, W_i .

The remaining procedure is analogous to the reference scheme. The watermark payload is then fed along with image block content $I_i^{(\text{ldr})}$, image block location i , and a security context κ to a hashing function:

$$H_i = h(I_i^{(\text{ldr})}, i, W_i, \kappa). \quad (5.2)$$

The resulting hash H_i is then embedded along with W_i into the i -th image block. This concludes the operation of the encoder.

Methods for Image Decoding

The operation of the adaptive self-embedding decoder is illustrated in Fig. 5.1b. The process starts with watermark extraction, which yields the extracted hashes

\hat{H}_i , and watermark payload \hat{W}_i , potentially different from the embedded original information. In the next step, the decoder recalculates the hashes:

$$H_i = h(I_i^{(w,ldr)}, i, \hat{W}_i, \kappa), \quad (5.3)$$

by repeating the analogous steps of the encoder. A comparison $H_i = \hat{H}_i$ results in a tampering (erasure) map, necessary for fountain decoding. The remaining authentic \hat{W}_i are demultiplexed (DEMUX) to obtain $\{\hat{Y}_i : H_i = \hat{H}_i\}$, and $\{\hat{Z}_i : H_i = \hat{H}_i\}$. Provided that the number of authentic symbols is sufficient for successful decoding of both streams, the decoder will be able to restore the approximate original content. Otherwise, the operation terminates, and the decoder yields only the tampering map E .

At first, the decoder recovers the quality descriptor \mathbf{q} . Based on the assigned reconstruction profiles, it then regenerates the reference blocks of authentic image regions, and constructs the corresponding reference symbols. The dependencies on the regenerated reference symbols are then removed from the previously extracted $\{\hat{Y}_i : H_i = \hat{H}_i\}$. The second fountain decoding yields the reference information of the tampered image blocks.

The last step is to restore their principal content. In order to eliminate disturbing noise boundaries between the watermarked and the restored image fragments, the L least significant bit-planes are set to 2^{L-1} , i.e., for $L = 3$, the LSBs are set to $4_{10} = 100_2$. Such an approach slightly lowers the average image distortion.

Communication of the Quality Descriptor

The quality descriptor is communicated to the decoder through the same channel as the reconstruction reference. The rate of the fountain code should be chosen accordingly to the anticipated maximal tampering rate. The complete quality descriptor is necessary, and the maximal tampering rate which allows for its successful retrieval is asymptotically:

$$\tilde{\gamma}_{qd} = 1 - \lambda_{qd}, \quad (5.4)$$

where λ_{qd} is the fountain code rate.

In this study, I consider the total number of descriptor bits per block to be always 8 bits. The quality descriptor is represented with either 2-bit, or 4-bit depth, for the content-adaptive, and the descriptor-adaptive schemes, respectively. Hence, the rate of the fountain code is $\frac{1}{4}$, and $\frac{1}{2}$, which correspond to $\tilde{\gamma}_{qd} = 0.75$, and $\tilde{\gamma}_{qd} = 0.50$.

Further improvement of $\tilde{\gamma}_{qd}$ could be obtained by employing source coding to compress the descriptor prior to channel coding. This issue is not addressed in this study.

Table 5.1: Reconstruction profiles for the descriptor-adaptive scheme.

Property	Profile no.							
	1	2	3	4	5	6	7	8
λ	0	1	1	2	3	1	2	3
Block texture	-	L	M	M	M	H	H	H
Guard	-	+	+	-	-	+	-	-

5.1.1 Definition of the Reconstruction Profiles

Adoption of multiple reconstruction profiles can be motivated by two main factors. Firstly, it is the improvement of the reconstruction fidelity, represented by the content-adaptive scheme. This scheme uses four profiles: a general one and 3 dedicated to low, medium and high-texture blocks. The profiles have identical reference rates, and differ by the quantization code-books and allocation matrices (Section 4.2.2).

Secondly, it is the need to enforce different restoration fidelity for various fragments of an image. At the cost of background content, certain fragments will be eligible for reconstruction with exceptionally high fidelity. Such a scenario is particularly useful for images with small areas with high-priority content, e.g., license number plates or human faces in video surveillance applications. This variant is supported by the descriptor-adaptive scheme. The utilized set of reconstruction profiles is presented in Table 5.1. It also contains profiles, which are optimized for blocks with various texture levels. The medium and high texture blocks can, however, be restored with three different fidelity levels. Three guard levels are defined (Section 3.2).

5.2 Experimental Evaluation

This section presents the results of experimental evaluation of the discussed self-embedding schemes: the reference, the content-adaptive, and the descriptor-adaptive. The former in a benchmark configuration with $L = 3$ is compared with 5 state-of-the-art alternative schemes in a common evaluation scenario.

In order to assess the efficiency of the trade-off between various system aspects, the schemes are considered in a number of various configurations, for three main watermark payloads, i.e., $L = 1$, $L = 2$, and $L = 3$. For $L = 1$, a single least significant bit-plane is used for watermark embedding. Hence, the embedding-inflicted distortion is:

$$20 \cdot \log_{10} \frac{255}{\sqrt{0.5}} \approx 51.1 \text{ dB.} \quad (5.5)$$

For the remaining configurations of $L = 2$ and $L = 3$, the expected embedding-inflicted distortion is respectively:

$$20 \cdot \log_{10} \frac{255}{\sqrt{2.5}} \approx 44.1 \text{ dB}, \quad (5.6)$$

$$20 \cdot \log_{10} \frac{255}{\sqrt{10.5}} \approx 37.9 \text{ dB}. \quad (5.7)$$

Further flexibility with respect to the embedding distortion can be easily achieved by using only selected bits, instead of complete bit-planes for information embedding.

5.2.1 Reference Self-Embedding Scheme

In this experiment, I perform large scale evaluation of various configurations of the reference scheme on the *ucid*, and the *bows* data-sets. The results presented in this chapter are expressed in PSNR of the restored fragments. The considered configurations, and the obtained results are collected in Table 5.2. The reported ranges correspond to the typical quality scores, and to the median. The typical scores are obtained by eliminating approximately 1% of the outliers from the empirical distribution, based on (4.9).

Fig. 5.2 shows a histogram of the reconstruction PSNR scores for the reference scheme ($L = 3$, $\lambda = 1$) on the complete *bows* data-set, i.e., 10,000 natural images. The average and the median PSNR are 37.0 dB and 37.3 dB, respectively. 99% of the obtained scores fall in the range [31.3; 41.5] dB. Box plots of the obtained fidelity scores are shown in Fig 5.3. On each box, the central mark is the median, and the edges correspond to the 25th and 75th percentiles. The whiskers extend to the most extreme data points, not considered outliers according to 4.9. The outliers are plotted individually. It can be observed that the reconstruction quality constantly improves with the increasing reference rate. Not only does the median increase, but also the range of typical scores quickly narrows.

Fig. 5.4 shows the trade-off between the reconstruction quality, and the maximum achievable tampering rates, defined by the theoretical success bounds. The presented results were obtained on the *bows* test set; the results for the *ucid* test set are nearly identical, and are therefore not presented. Solid lines represent the $\tilde{\gamma}_3$ bound, while dashed lines the optimistic $\tilde{\gamma}_2$ bound. Horizontal helper lines show successive success bounds for $\lambda = 1/4, 1/2, 1, 2, 3, \dots, 10$. The reconstruction quality scores are shown as both absolute median PSNR, and median PSNR relative to maximum achievable PSNR for each of the considered configurations. The latter stems from the watermarking-inflicted distortion (5.5)-(5.7), but due to the predefined LSB replacement, the distortion is reduced by half.

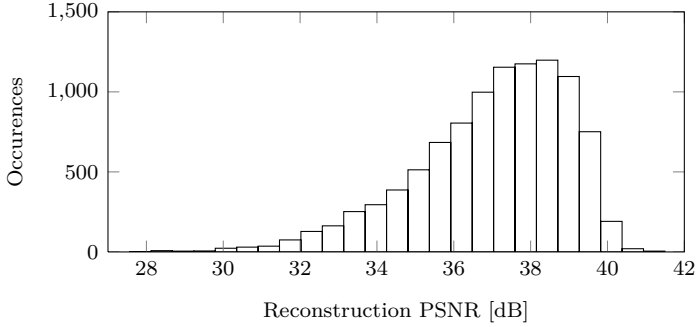


Fig. 5.2: Histogram of reconstruction PSNR on the *bows* data-set.

Configurations with $L = 3$ are the most efficient, both when absolute and relative quality scores are of concern. However, they do not allow for reconstruction with better fidelity than approx. 40 dB. In case better fidelity is needed, the $L = 2$ variants can be used, with only slightly worse trade-off efficiency. Significantly worse performance for $L = 1$ stems from a limited amount of watermark capacity, which carries the reference information.

An example of content authentication is shown in Fig. 5.5, which illustrates a manual tampering case, where some objects are removed from the facade of the building, and the sky is replaced with a cloudy pattern. The final tampering rate is approximately 23% of the image area, just below of the 25% theoretical limit for the considered configuration ($L = 2, \lambda = 3$). The reconstruction is successful. All of the tampered fragments are correctly detected, and the original appearance is properly restored. The final PSNR is greater than for the watermarked image, due to replacement of the least significant bit-planes with 10_2 .

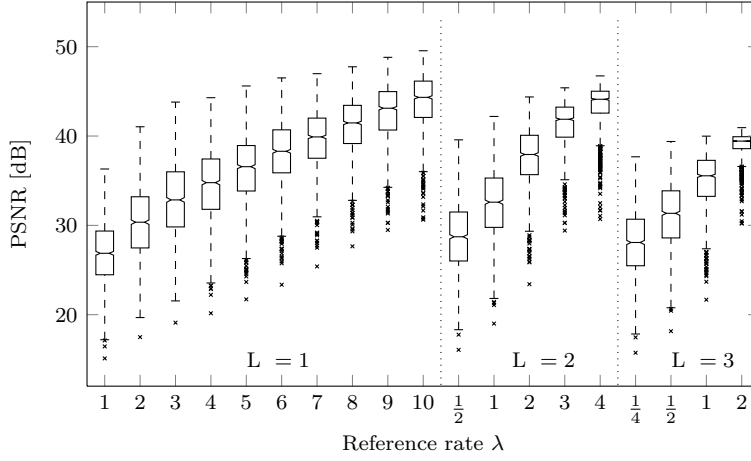
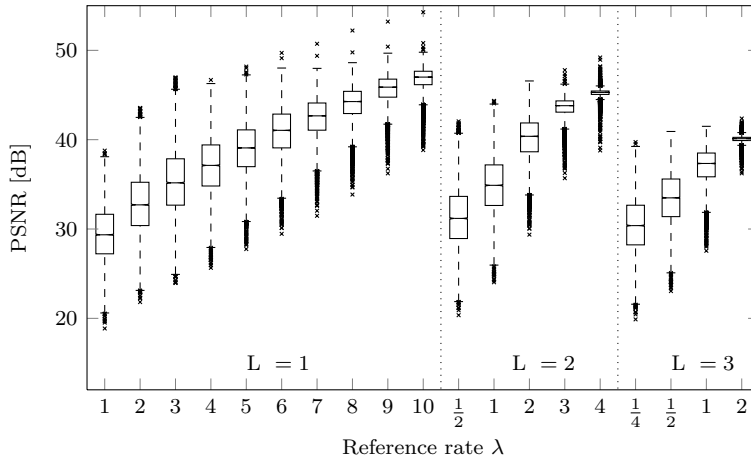
(a) *ucid* test set(b) *bows* test set

Fig. 5.3: Reconstruction quality for various configurations of the reference scheme, measured as the PSNR on the *ucid* and *bows* test sets.

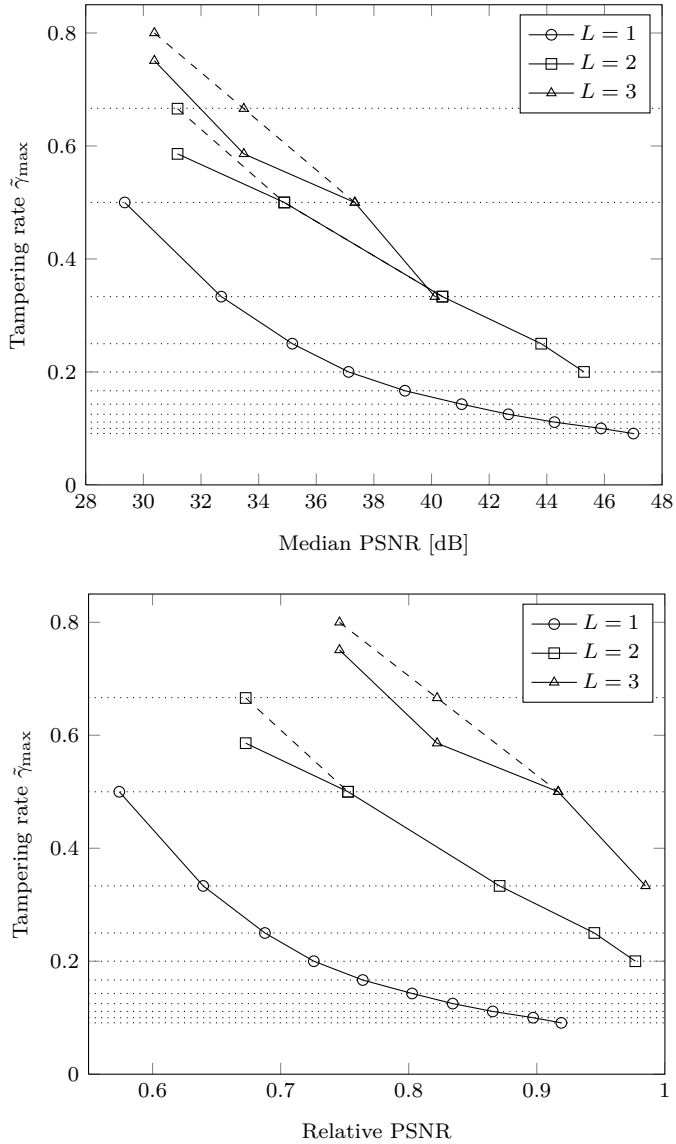


Fig. 5.4: Trade-off between the tampering rate and the median PSNR (top) and relative median PSNR (bottom) for the *bows* test set; solid lines represent the $\tilde{\gamma}_3$ bound; dashed lines represent the optimistic $\tilde{\gamma}_2$ bound.

Table 5.2: Reconstruction quality for various configurations of the reference self-embedding scheme, measured as the PSNR on the *ucid* and *bows* test sets..

Config. no.	λ	Rec. PSNR - <i>ucid</i> [dB] Range	Median	Rec. PSNR - <i>bows</i> [dB] Range	Median	Tamp. rate $\hat{\gamma}_{\max}$
$L = 1 \rightarrow B = 32$; embedding PSNR ≈ 51 dB						
1	1	17.1 - 35.3	26.1	20.6 - 38.1	29.4	50.0%
2	2	19.0 - 39.6	29.1	23.1 - 42.5	32.7	33.3%
3	3	20.8 - 42.6	31.5	24.9 - 45.6	35.2	25.0%
4	4	22.1 - 44.3	33.4	27.9 - 46.3	37.1	20.0%
5	5	24.6 - 45.6	35.5	30.8 - 47.3	39.1	16.7%
6	6	26.2 - 46.5	37.5	33.4 - 48.0	41.0	14.3%
7	7	29.1 - 47.0	39.6	36.5 - 48.0	42.7	12.5%
8	8	32.0 - 47.6	41.5	39.2 - 48.6	44.3	11.1%
9	9	34.3 - 48.5	43.6	41.7 - 49.7	45.9	10.0%
10	10	37.2 - 49.4	45.1	43.9 - 49.8	47.0	9.01%
$L = 2 \rightarrow B = 96$; embedding PSNR ≈ 44 dB						
11	$\frac{1}{2}$	18.2 - 37.6	27.7	21.9 - 40.7	31.2	58.6%
12	1	21.0 - 42.1	31.4	26.0 - 44.0	34.9	50.0%
13	2	26.6 - 44.4	37.2	33.8 - 46.6	40.4	33.3%
14	3	35.1 - 45.3	42.2	41.2 - 46.2	43.8	25.0%
15	4	41.3 - 46.7	44.7	44.5 - 46.0	45.3	20.0%
$L = 3 \rightarrow B = 160$; embedding PSNR ≈ 38 dB						
16	$\frac{1}{4}$	18.0 - 36.8	27.1	21.6 - 39.2	30.4	75.1%
17	$\frac{1}{2}$	20.4 - 39.5	30.2	25.1 - 40.9	33.5	58.6%
18	1	25.4 - 40.1	34.7	31.9 - 41.5	37.3	50.0%
19	2	37.2 - 40.9	39.6	39.4 - 40.8	40.1	33.3%

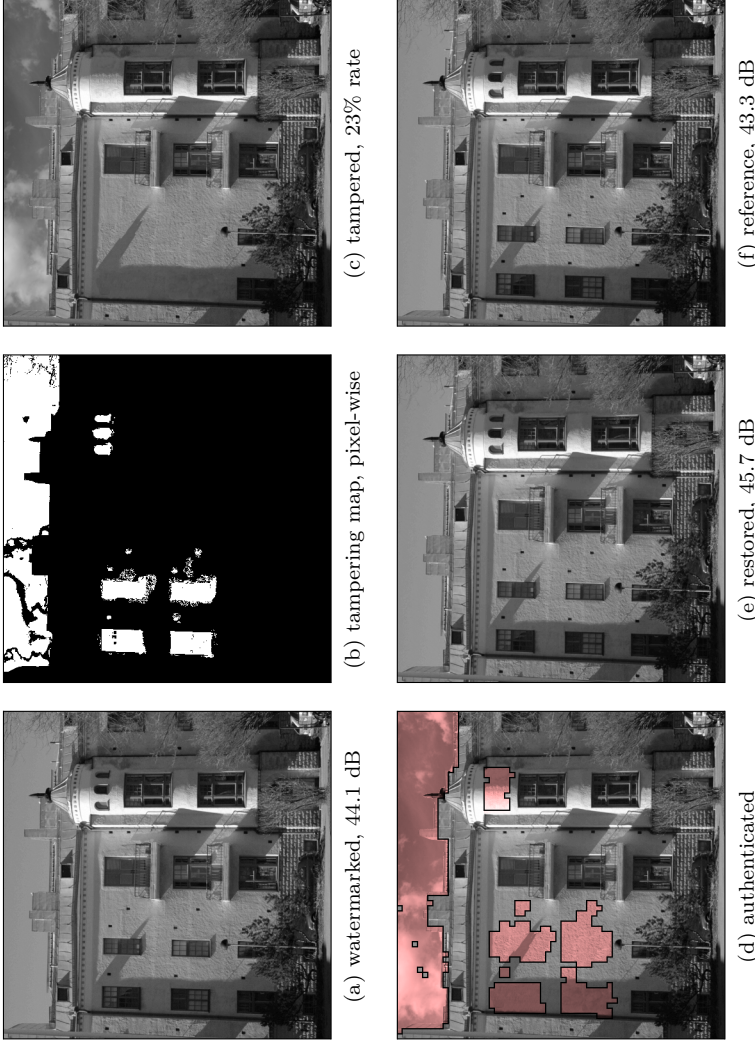


Fig. 5.5: Example reconstruction of maliciously tampered content; tampering rate $\tilde{\gamma} = 0.23$; image 9560 from the bows data set; $L = 2$, $\lambda = 3 \rightarrow \tilde{\gamma}_{\max} = 0.25$.

5.2.2 Comparison with State-of-the-Art Schemes

In this experiment, I compare the achievable reconstruction performance of the presented reference scheme ($L = 3, \lambda = 1$) with 5 state-of-the-art alternative schemes, both with constant [73, 76], and with flexible reconstruction quality [73–75]. The reconstruction performance, as reported in the original publications, does not allow for fair comparison of the adopted reconstruction approaches, as the test images and the tampering patterns are different. I reimplemented the schemes in a common evaluation framework¹. The least significant bits in the restored areas are set to 100₂ in all of the schemes.

Evaluation is performed on a set of 48 images of size 512×512 px, selected from 10,000 gray-scale natural images from the *bows* data-set. The selected images span the space of possible image characteristics, i.e., include dark, medium and bright images with various amount of details, measured as an average standard deviation of individual image blocks σ_i . Example images are shown in Fig. 5.6.

The images are watermarked with each of the considered schemes, and then randomly modified with tampering rates from 0.025 to 0.6 with a 0.025 step. Selection of image blocks for modification is the same for all of the schemes. The experiment is repeated 30 times, with different seeds for the pseudo-random number generator.

The average PSNR scores for selected tampering rates are collected in Table 5.3. The results are averaged over 48 images and 30 independent replications of the experiment. The presented reference scheme is capable of high-quality reconstruction, regardless of the observed tampering rate. The factual quality depends on the characteristics of the tampered content. While for low tampering rates two flexible schemes [74, 75] can deliver better performance, in case of extensive tampering the proposed algorithm is clearly superior. The threshold tampering rate is usually between $\tilde{\gamma} = 0.05$ and $\tilde{\gamma} = 0.2$, depending on the image itself.

The proposed algorithm is also beneficial when compared to constant-fidelity schemes. It delivers significantly better quality than [73], with only 9% worse tampering rate. An additional comment is necessary about the [76]-A scheme, which uses the remaining 5 most significant bit-planes as reference information, i.e., 320 bits per block. As a result, it allows for perfect recovery of these bit-planes, and the expected PSNR is 40.7 dB. The maximum tampering rate for this scheme is 0.24. With $b = 320$ ($\lambda = 2$), the success bound of the proposed method is

¹Due to reference values overflows, stemming from large DC coefficients, the original scheme from [74] exhibits prohibitively poor performance for dark and light images. The issue has been fixed in the presented evaluation by adjusting the quantization procedure in (8); $2f_t$ is used instead of f_t . While this operation limits the maximal reconstruction quality to approximately 38 dB, it allows for correct operation on dark and light images.

Table 5.3: Comparison of the reconstruction performance with state-of-the-art self-embedding schemes.

Scheme	Reconstruction PSNR [dB] for various $\tilde{\gamma}$						$\tilde{\gamma}_{\max}$
	0.050	0.100	0.200	0.300	0.400	0.475	
[74]	37.2	34.8	31.7	29.1	28.1	27.0	0.60
[75]	37.3	35.6	33.3	31.6	30.2	29.2	0.54
[76]-B	31.8	31.7	31.7	28.7	28.7	25.8	0.66
[73]	28.5	28.4	28.4	28.4	28.4	28.4	0.59
[76]-A	40.7	40.7	40.7	-	-	-	0.24
Proposed	36.4	36.4	36.4	36.4	36.4	36.4	0.50

$\tilde{\gamma}_2 = \tilde{\gamma}_3 = 0.33$. A detailed theoretical analysis of both reconstruction approaches is presented in Appendix B.

Individual reconstruction fidelity scores for example test images are shown in Fig. 5.7. The plots not only clearly show the threshold tampering rates, but also demonstrate the characteristic behavior of the reconstruction systems. Two flexible schemes [74, 75] reveal systematic deterioration of the reconstruction fidelity. The scheme [74] is more susceptible to the distributions of the details in the image. For images with large areas of solid low-detail blocks, the obtained curves may not be monotonic. The scheme [75] operates directly on pixel intensities, and is not affected by the problem. The plots also demonstrate the expected three distinct quality levels for the [76]-B scheme.

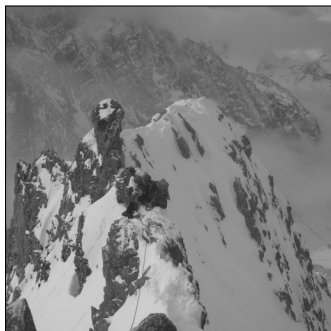
Reconstruction examples are shown in Fig. 5.8-5.12. For the sake of presentation clarity, the images were processed after down-sampling to 256×256 px, and are printed at 72 dpi clipped to a fraction of their height. The tampered area (marked in the top left sub-figure) is always rectangular, and the tampering rate $\tilde{\gamma}$ is either 0.28, unless stated otherwise. The proposed algorithm yielded the best quality. Large amount of highly textured blocks causes reference value saturation, most visible in [73], and [76]-B. The fidelity of [74] is limited by reconstruction artifacts, typical for this scheme when dealing with larger tampering rates. More reconstruction examples are available as supplementary multimedia materials.



(a) 6882



(b) 9011



(c) 131



(d) 4749

Fig. 5.6: Example test images from the *bows* data-set; original images of size 512×512 px, printed at 300 dpi.

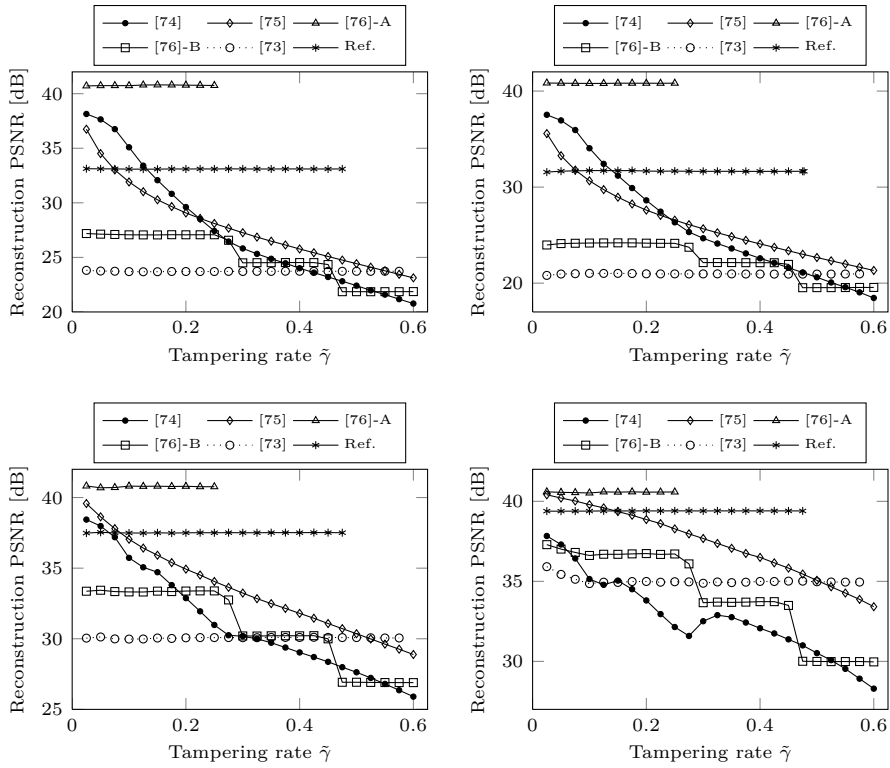


Fig. 5.7: Reconstruction PSNR for selected test images under varying tampering rates. From top left: 6882, 9011, 131, 4749.



(a) Original image & tampering



(b) Proposed reference scheme, 30.82 dB



(c) [74], 23.55 dB

Fig. 5.8: Reconstruction results for state-of-the-art self-embedding schemes for image 6882 from the *bows* data set (*cont.*).



(d) [76]-B, 24.73 dB



(e) [75], 25.11 dB



(f) [73], 21.40 dB

Fig. 5.8: Reconstruction results for state-of-the-art self-embedding schemes for image 6882 from the *bows* data set.



(a) Original image & tampering



(b) Proposed reference scheme, 30.69 dB

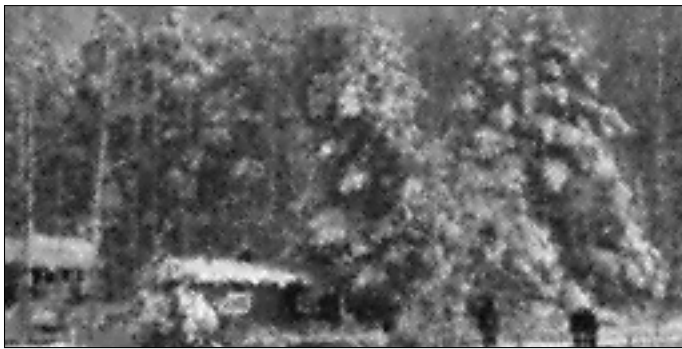


(c) [74], 19.69 dB

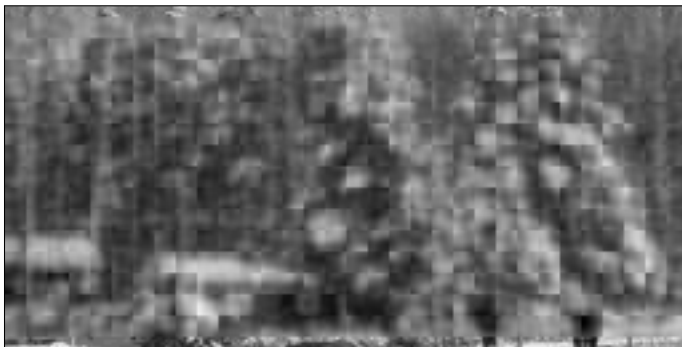
Fig. 5.9: Reconstruction results for state-of-the-art self-embedding schemes for image 6882 from the *bows* data set tampered with rate $\tilde{\gamma} = 0.46$ (*cont.*).



(d) [76]-B, 22.04 dB



(e) [75], 22.11 dB



(f) [73], 21.05 dB

Fig. 5.9: Reconstruction results for state-of-the-art self-embedding schemes for image 6882 from the *bows* data set tampered with rate $\tilde{\gamma} = 0.46$.



(a) Original image & tampering



(b) Proposed reference scheme, 31.20 dB



(c) [74], 24.33 dB

Fig. 5.10: Reconstruction results for state-of-the-art self-embedding schemes for image 9011 from the *bows* data set (*cont.*).



(d) [76]-B, 24.00 dB

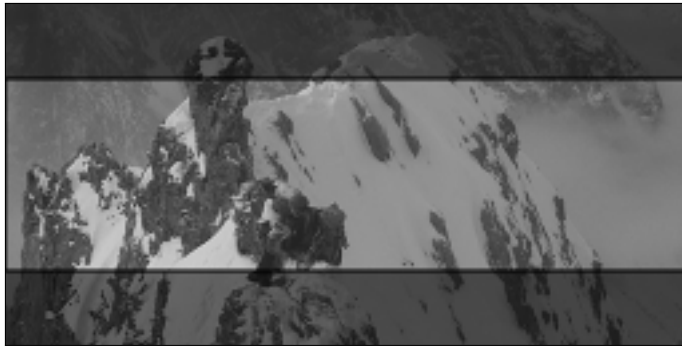


(e) [75], 25.57 dB



(f) [73], 20.80 dB

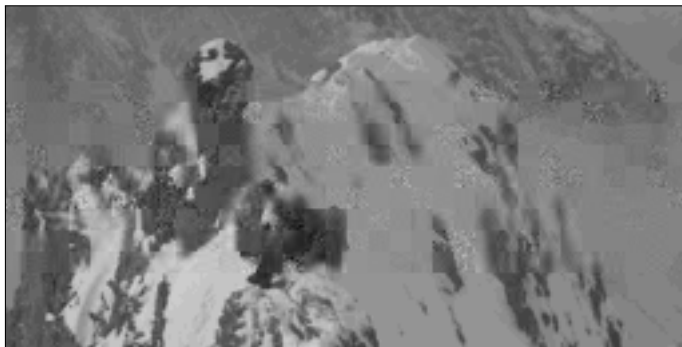
Fig. 5.10: Reconstruction results for state-of-the-art self-embedding schemes for image 9011 from the *bows* data set.



(a) Original image & tampering



(b) Proposed reference scheme, 36.21 dB

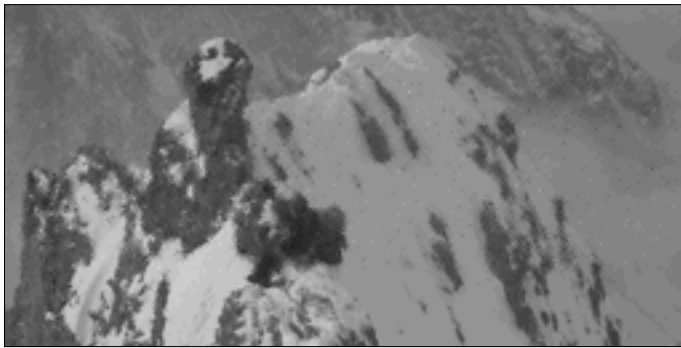


(c) [74], 28.35 dB

Fig. 5.11: Reconstruction results for state-of-the-art self-embedding schemes for image 131 from the *bows* data set (*cont.*).



(d) [76]-B, 31.07 dB



(e) [75], 31.35 dB



(f) [73], 27.74 dB

Fig. 5.11: Reconstruction results for state-of-the-art self-embedding schemes for image 131 from the *bows* data set.



(a) Original image & tampering



(b) Proposed reference scheme, 36.99 dB



(c) [74], 28.39 dB

Fig. 5.12: Reconstruction results for state-of-the-art self-embedding schemes for image 4749 from the *bows* data set (*cont.*).



(d) [76]-B, 29.29 dB



(e) [75], 31.69 dB



(f) [73], 26.73 dB

Fig. 5.12: Reconstruction results for state-of-the-art self-embedding schemes for image 4749 from the *bows* data set.

Table 5.4: Reconstruction performance for the content-adaptive scheme.

λ	Rec.PSNR - <i>ucid</i> [dB] Range	Median	Rec. PSNR - <i>bows</i> [dB] Range	Median
$L = 2 \rightarrow B = 88$; embedding PSNR ≈ 44 dB				
$\frac{1}{2}$	17.5 - 38.7	28.1	21.3 - 42.8	31.7
1	20.3 - 43.0	31.6	25.1 - 45.8	35.4
2	26.1 - 45.3	37.2	33.0 - 47.5	40.8
3	34.2 - 46.1	42.1	40.5 - 47.7	44.3
4	40.6 - 47.0	44.7	44.4 - 46.7	45.6
$L = 3 \rightarrow B = 152$; embedding PSNR ≈ 38 dB				
$\frac{1}{4}$	17.2 - 37.8	27.3	21.1 - 41.4	31.0
$\frac{1}{2}$	19.9 - 40.2	30.4	24.9 - 41.7	34.0
1	25.2 - 40.7	35.0	31.9 - 42.2	37.8
2	37.1 - 41.1	39.8	39.5 - 41.1	40.3

5.2.3 Content-Adaptive Self-Embedding Scheme

Introduction of content adaptivity allows for incorporation of quantization codebooks, and allocation matrices optimized for blocks with various texture levels. The considered content-adaptive scheme distinguishes low, medium, and high-texture blocks. In addition to the dedicated reconstruction profiles, the scheme also uses a general profile. Typically, such an approach would guarantee improvement of the reconstruction quality. However, since the quality descriptor needs to be communicated to the decoder, the effective reference rate is reduced. In the considered setup, the reference symbol length is decreased by 8.

The purpose of this experiment is to investigate whether the fidelity gain from content adaptivity can compensate the fidelity loss due to smaller reference rates. The evaluation is performed for $L = 2$, and $L = 3$. The results for the complete *ucid*. and *bows* data sets are collected in Table 5.4. The achievable improvement depends on the configuration of the system. For lower reference rates, the average improvement reaches nearly 0.8 dB. For higher rates it drops to 0.1-0.3 dB.

Fig. 5.13 and 5.14 show example reference images for the content-adaptive scheme, and their corresponding counterparts from the reference scheme. The images were obtained with $\lambda = 1/4$, i.e., $b = 38$, and $b = 40$ for the content-adaptive, and the reference schemes, respectively. The most visible difference is a significant reduction of the blocking artifacts, stemming from limited precision of the DC coefficients. This phenomenon is particularly well visible in solid areas.

In conclusion, the discussed technique is best suited for low reference rates. Both the objective fidelity gain, and the human-perceivable improvement are the most significant for such configurations. For higher reference rates, the improvement is diminished, and a single reconstruction profile seems sufficient.



(a) reference scheme, 33.08 dB



(b) content-adaptive scheme, 34.43 dB

Fig. 5.13: Reconstruction reference for image 7710 obtained with the reference ($b = 40$ bpb), and the content-adaptive schemes ($b = 38$ bpb).



(a) reference scheme, 32.73 dB



(b) content-adaptive scheme, 33.86 dB

Fig. 5.14: Reconstruction reference for image 7787 obtained with the reference ($b = 40$ bpb), and the content-adaptive schemes ($b = 38$ bpb).

5.2.4 Descriptor-Adaptive Self-Embedding Scheme

The descriptor-adaptive scheme uses multiple reconstruction profiles with various reference rates. Hence, the scheme has an additional degree of freedom for controlling the trade-off between the reconstruction quality and the tampering rate. In fact, arbitrary tampering rates from the range:

$$\tilde{\gamma} \in \left(\frac{1}{1 + \lambda_S}; \frac{1}{1 + \lambda_1} \right) \quad (5.8)$$

can be obtained with an appropriate profile assignment. λ_S and λ_1 denote the highest, and the lowest of the utilized reference rates.

The goal of this experiment is to compare the achievable reconstruction efficiency with the uniform-quality reference scheme. The latter is evaluated in three configurations, $\lambda = 1, 2$, and 3 . The adaptive scheme uses the same reference rates, but allows for assigning them to individual image blocks.

The encoder prepares a number of quality descriptors for a range of considered target tampering rates $\tilde{\gamma}_{\text{target}} \in [0.3, 0.5]$. The reconstruction success chart for randomly drawn tampering rates is shown in Fig. 5.15. Success and failure cases are marked with circles and crosses, respectively. It can be observed, that the scheme allows for successful reconstruction up to the desired target tampering rate. Example quality descriptors for selected tampering rates are shown in Fig. 5.16. The figure illustrates two example images, and their corresponding descriptors. The top row corresponds to uniform degradation, while the bottom row uses an importance map to guarantee high reconstruction quality for the license number plate. As a result, the license number remains perfectly clear, regardless of the desired tampering rate, and the reconstruction of the remaining regions. The descriptor was obtained for oblivious tampering.

Fig. 5.17 shows the prospective reconstruction result, i.e., the reference image, for both the adaptive, and the reference uniform-quality schemes. The protected image can be successfully restored up to 49%, and 50% of the image area for the former, and the latter schemes, respectively. While the reference scheme has higher effective payload (96 vs. 88 bits per block), it does not allow for clear reconstruction of the license number plate.

Due to additional bandwidth required for communication of the quality descriptor, the adaptive scheme uses lower effective payload for the reconstruction reference. This leads to slight degradation of the reconstruction quality at the end-points of the characteristics. At the same time, the ability to use better suited reconstruction profiles is capable of improving the quality. With properly chosen operation parameters, these effects can be compensated or even improved upon, depending on the image content.

Different behaviors of the trade-off are illustrated in Fig. 5.18, which shows scatter-plots of the reconstruction quality vs. the achievable tampering rates.

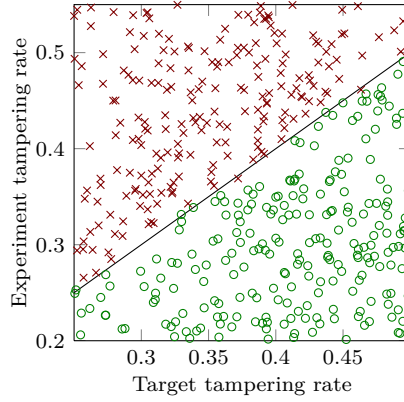


Fig. 5.15: Validation of the descriptor design objective based on 1,000 random reconstruction attempts with the descriptor-adaptive scheme.

The figure compares the efficiency of the descriptor-adaptive scheme, against the reference scheme. The former is considered for both the pessimistic, and the oblivious tampering. The efficiency of the trade-off is typically lower, when pessimistic tampering is of concern. The worst degradation can be observed for descriptors with dominating high-quality profiles, i.e., near $\tilde{\gamma} = 0.25$. This result is fully consistent with the theoretical analysis (Chapter 3). For some images, e.g., *7683* or *4749*, the efficiency can actually improve, as the target tampering rate approaches $\tilde{\gamma} = 0.5$. For oblivious tampering, the trade-off efficiency is usually better.

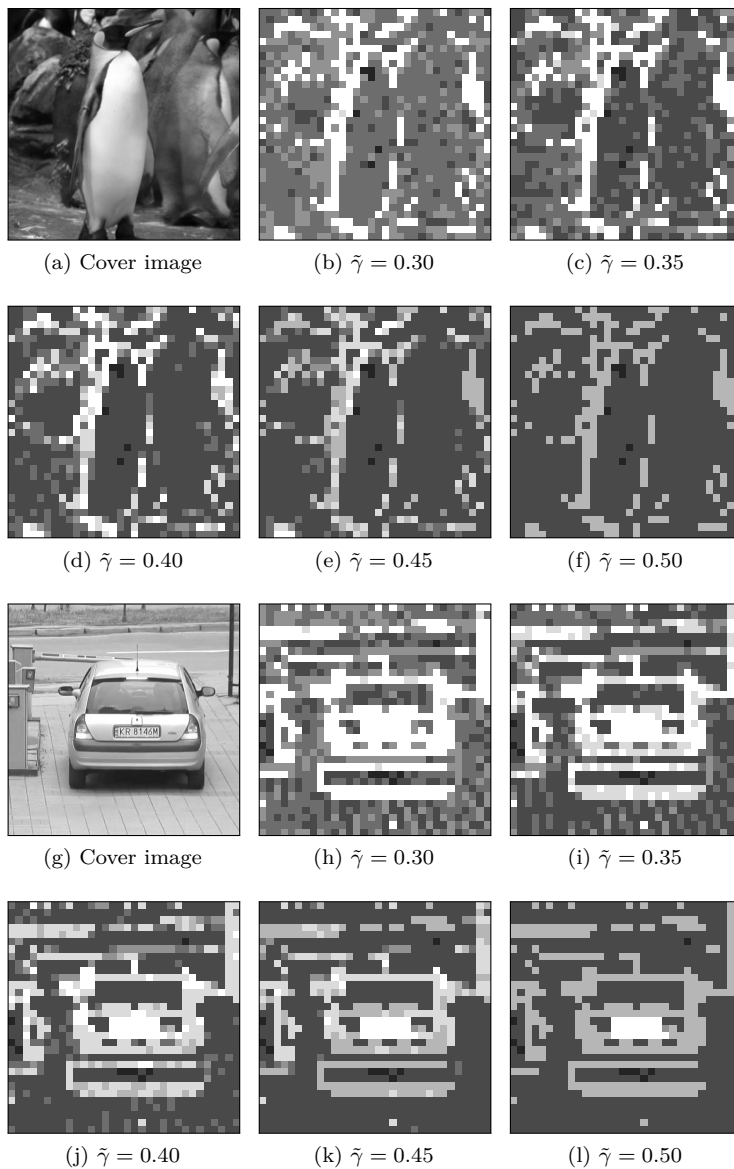
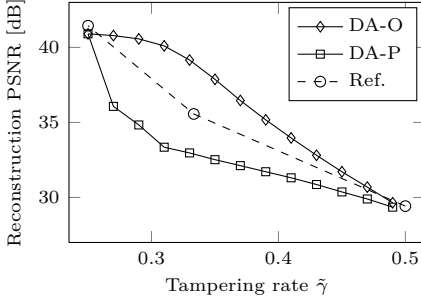


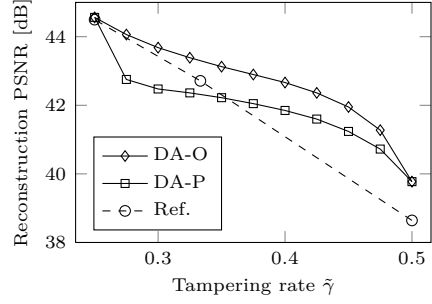
Fig. 5.16: Quality descriptors obtained for the descriptor-adaptive scheme for oblivious tampering without (b-f) and with (h-l) an importance map.



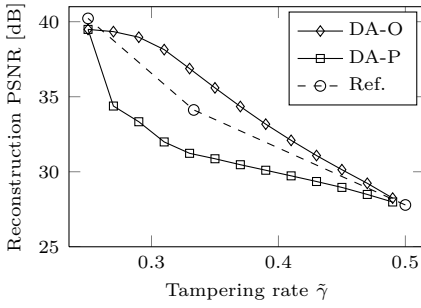
Fig. 5.17: Comparison of important content representation for the reference scheme, and the descriptor-adaptive scheme; same maximal tampering rate; license number plate chosen as important content.



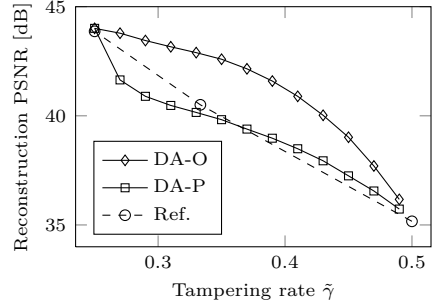
(a) image 6882



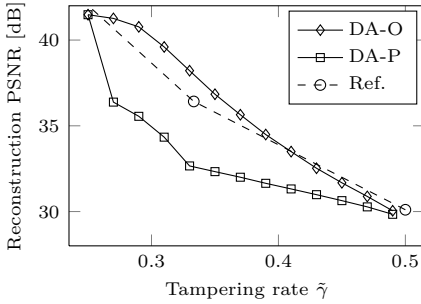
(b) image 4767



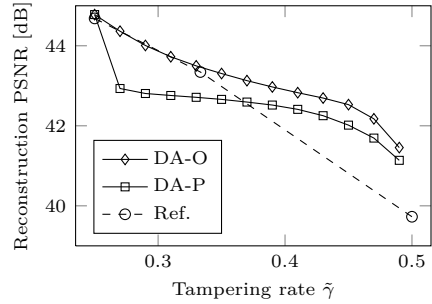
(c) image 9011



(d) image 131



(e) image 7683



(f) image 4749

Fig. 5.18: Scatter plots of the reconstruction quality vs. the tampering rate for the reference (Ref), and the descriptor-adaptive schemes for the pessimistic (DA-P) and oblivious tampering (DA-O).

Conclusions

The goal of the presented research was to analyze the content reconstruction problem in digital image authentication systems. The reconstruction was formulated as communication over an erasure channel, modified to take into account individual properties of the problem at hand. Communication theoretic tools were used for derivation of the reconstruction success bounds, and for further analysis of the reconstruction performance.

It was conclusively demonstrated that the proposed communication model accurately describes the real behavior, and the inherent trade-offs in content reconstruction systems. The adopted approach is immune to the most challenging problems, which crippled many earlier self-embedding schemes, i.e., the reconstruction dependency, and the reference waste problems. Due to the fact that the reference information is uniformly spread over the whole image, all authentic fragments contribute to the restoration of each tampered block. By ensuring that the watermark embedding function does not interfere with the reference generation procedure, and by exploiting the remaining authentic image regions, the knowledge about their appearance can be completely eliminated from the watermark. As a result, it becomes possible to use reference streams significantly longer than the available embedding capacity. The dissertation explains that certain parameters of the reconstruction process have critical importance, and should be properly chosen to guarantee optimal performance.

The proposed reconstruction model can be efficiently implemented with the use of digital fountain codes. Three example self-embedding schemes were implemented, and extensively evaluated with respect to the most important operation characteristics. With the same rate of reference information per block, the developed schemes allow for restoration with higher fidelity, or working with higher tampering rates, compared to existing alternative schemes. Recent independent work on self-embedding has also used a mechanism for distributing the reference information over the image [74, 76]. However, the adopted approach is different.

The reconstruction mechanism from [74, 76] divides the reconstruction problem into smaller sub-problems, involving reconstruction of randomly shuffled image fragments. As a result, the achievable tampering rates are smaller than the derived theoretical bounds. A detailed comparison of both approaches is presented in Appendix B.

The dissertation also addressed the emerging topic of adaptive self-embedding. Such schemes allow for defining the desired restoration fidelity on a per-block basis. Originally, such an approach was used to limit the length of the reference stream by assigning less reference bits to low-texture fragments. I extended the concept in order to allow for providing guarantees on certain reconstruction performance aspects. Specifically, the developed solution can ensure high-quality reconstruction of selected image fragments, and robustness against modifications up to a given target tampering rate. Such guarantees are of principal importance in a number of applications.

For this purpose, I developed a procedure for calculating the mapping between the image blocks and the reconstruction profiles, which maximizes the reconstruction quality given certain operation constraints. The procedure is based on the derived reconstruction model, and uses the introduced notion of reconstruction demand, to accurately estimate the applicable success bound.

I believe that adoption of the proposed model might enable the development of robust, easily customizable reconstruction systems. The presented self-embedding schemes show a practical implementation of a pro-active protection mechanism intended for loss-less digital images. In case of lossy-compressed images, the effective watermark capacity is severely reduced, and it becomes necessary to deal with prospective block classification, and watermark recovery errors. In fact, the proposed model can be easily extended to efficiently handle such problems. A possible starting point for further development in this direction is briefly discussed in Appendix E.

Achievements and Contributions

The achievements and contributions of the dissertation can be summarized as follows:

1. Exhaustive analysis of the current state-of-the-art was performed, with emphasis on different approaches to the content reconstruction problem, their achievable restoration performance, and inherent restoration trade-offs.
2. A new model of the content reconstruction problem was presented. The model is based on a revised erasure communication channel, and allows for adopting the tools from communication theory to analyze the success conditions, and the reconstruction performance of self-embedding schemes.

3. The proposed model uses a new mechanism for spreading the reference information over the whole image. Practical implementation of the proposed approach is based on digital fountain codes, and achieves better performance than existing state-of-the-art schemes.
4. A novel analysis of the reconstruction success conditions was performed, leading to clear analytical formulas for the appropriate success bounds.
5. The presented analysis gives better understanding of the inherent reconstruction trade-offs, as well as clear guidelines and conditions that need to be satisfied to guarantee optimal performance for various configurations of the system.
6. The concept of the *reconstruction demand* was introduced, to describe the impact of the tampering in the domain of image blocks on the necessary portions of the reference stream.
7. The concept of adaptive self-embedding was extended in order to provide guarantees on the desired reconstruction quality, and the supported tampering rates. Such an approach is of principal importance in a number of applications, where certain image fragments should be reconstructed with higher fidelity than others.
8. The impact of content adaptivity on the reconstruction success bounds was analyzed theoretically using the proposed self-recovery model, and the concept of reconstruction demand.
9. It was shown that in adaptive self-embedding, the highest-fidelity level is the dominant factor, which influences the achievable success bounds. Despite introducing reconstruction profiles with lower fidelity, the achievable success bounds are not necessarily improved.
10. An algorithm was proposed for the design of quality descriptors with an objective to maximize the reconstruction quality, given a set of constraints regarding the desired properties of the restoration process.
11. This study is the first to use a formal procedure for the selection of coefficient quantization precision. The issue becomes particularly important in adaptive schemes, where multiple allocation profiles need to be used.
12. An exhaustive reconstruction quality assessment was presented, including 5 state-of-the-art self-embedding schemes. The evaluation is performed on a representative test set of natural images, with identical tampering for all of the schemes.

In the light of the presented results, it can be stated that the thesis:

It is possible to model the content reconstruction problem as an erasure communication channel. Such a model allows for accurate formal analysis of the success bounds, and the inherent restoration trade-offs. A self-embedding scheme based on the proposed communication model delivers superior performance, compared to state-of-the-art reconstruction algorithms. High-quality reconstruction is possible even under extensive tampering.

has been proven.

Further Research Perspectives

While the presented content reconstruction model provides a solid foundation for analysis and construction of efficient self-embedding schemes, there are still certain aspects that need to be addressed to enable practical implementation of the described protection mechanism. Firstly, the presented schemes are intended for loss-less digital images. For commonly used lossy-compressed formats, e.g., JPEG, it becomes necessary to employ selective authentication and deal with block classification, and watermark extraction errors. The presented model can be easily extended to take such issues into account (Appendix E). However, further work in this direction still needs to be done.

Additionally, lossy-compressed formats are characterized by lower embedding capacity. While the proposed model allows for working with reference payloads significantly bigger than the available capacity, it might still be beneficial to consider alternative image representation methods, possibly better suited to low bit rates, e.g., Smoothlets, or X-Lets in general [48].

Secondly, the presented model needs to be extended to handle color images. Possible approaches include straightforward repetition of the complete reconstruction process for each color channel separately, and aggregation of the additional channels to strengthen the protection of the luminance component only. The decision will depend not only on the desired reconstruction mode, but also on the utilized image format. Factors like color space, and chromaticity compression settings will highly impact the final decision. A possible starting point for a discussion on the protection of color images is briefly given in Appendix E.

Thirdly, since large sensors are becoming commonly used in mainstream digital cameras, it remains an open problem how to efficiently handle very large images. While it is possible to divide the reconstruction problem into smaller, separate problems [75, 76], such an approach severely limits the achievable reconstruction performance (Appendix B), and still remains computationally challenging for images of several megapixels. The alternative presented in this dissertation, i.e., to

employ M-ary symbols instead of individual bits, is also not sufficient. A possible solution might be to develop a hybrid, properly balanced scheme, which combines both approaches.

Analogous problems can also be observed in traditional communication. Due to high computational complexity of the RLF decoding procedure, the code is rarely used in practice. Instead of the full Gaussian elimination, *belief propagation* is used for decoding [51, 52]. However, it requires a sparse generator matrix, which needs to satisfy certain requirements, expressed by means of a degree distribution. The most popular fountain code is the LT code [50]. It uses a *robust soliton distribution* to describe the degrees of successive code symbols. Unfortunately, the idea to employ a sparse generator matrix cannot be directly applied to the problem at hand. The encoder does not know the tampering locations, and the embedded watermark symbols would be quickly reduced to null, useless symbols during belief propagation. It might be possible to design a new degree distribution, better suited to the content reconstruction problem, but the problem still remains open.

The reference scheme presented in this dissertation should be considered as an illustration of a new self-recovery concept. Depending on the application requirements, it might be necessary to consider additional randomization or encryption mechanisms. In some applications, the content reconstruction functionality can further benefit from the use of RSA cryptography in a public key infrastructure, where the embedding and the verification keys are different [26].

This appendix discusses the most popular attacks on self-embedding schemes, and the necessary measures that need to be taken when implementing self-embedding mechanisms in real-world systems. Virtually all of the known attacks exploit vulnerabilities of the content authentication component. The primary goal is to enforce false negative classification of illegitimate image blocks. Even though a block itself might be intact, it might be misplaced in a different context. This is the fundamental assumption of the most popular attacks, i.e., the collage, and vector quantization (VQ) attacks [31]. The attack begins with identification of *equivalence classes*, i.e., sets of image blocks that can be replaced with each other without causing an alarm. Provided that the cardinality of such sets is large enough for any possible location in an image, it might be possible to use vector quantization to doctor a counterfeit image.

In certain applications, where large amounts of similar content are plentiful, the time needed to gather sufficient legitimate material might be short enough for the attack to be feasible. In video surveillance, where it might be desirable to remove selected objects from the scene, the abundant background content can easily be exploited to obtain a counterfeit. In fact, there exist algorithms originating from privacy protection research, which perform object deletion automatically in real time, by exploiting the background content models originally used in object detection and tracking [70].

With the knowledge of the internal construction of the algorithm, as well as its

parameters, it is possible to devise targeted attacks. If the embedding locations are known, it is possible to perform synchronized replacement of both the content, and the associated security information [29]. In many earlier schemes, the hashes were embedded with constant shift with respect to the original coordinates of the blocks. In such case, it is straightforward to perform malicious, unnoticeable replacement. The constant-average attack [11] exploits the knowledge about the block feature calculation mechanism, and adjusts the content until the counterfeit blocks are accepted by the authentication mechanism. A similar approach serves as a foundation for the XOR-equivalent attack [28], where a single bit-plane in a block suffices to be adjusted to conform the content to the expected hash.

In order to prevent straightforward implementation of targeted attacks, it has been argued that the details of digital watermarking algorithms should not be made publicly available. An observation which seems contradictory with the Kerckhoffs's principle from cryptography, stems from the fundamental differences between meaning of security in multimedia and in cryptographic applications [18]. In the latter there usually exists only a single, or at worst only a small set of feasible answers. In multimedia however, due to the imperfections of the human vision system, the same content is represented by a very large set of images. Various approximations of illegitimate content could be investigated to find a matching one.

The collage and VQ attacks are feasible in systems with oblivious block-based authentication mechanisms. Hence, successful implementation requires some context for proper authentication. In order to facilitate automatic operation two approaches are usually adopted. First, the context can be extracted from neighboring blocks, by taking their content into account when calculating the hashes [27, 30]. Such schemes usually require some tampering map correction mechanism, e.g., by means of voting [64]. An alternative approach is to automatically extract a global image feature, but it is not feasible if similar images are easily accessible.

The authentication context can also be provided manually by designing an appropriate usage protocol. An identifier, and a time-stamp can be used for both initialization of the pseudo-random number generator and in the cryptographic hash calculation procedure. The knowledge of such information at the decoder side is a typical case for many real-world environments.

The described self-embedding schemes are not vulnerable to simple targeted attacks. The embedded payload contains both the data hashes, and the reference information, which cannot be separated without knowing the secret key. Without providing a proper security context, it remains possible to create a successful collage, failing to alarm the detector if only whole blocks are replaced within their equivalence classes. However, in contrast to schemes with only tampering localization capabilities, self-embedding schemes, even when performing oblivious

block authentication could theoretically resist the collage and VQ attacks. A single image block contains reference information, which describes the content of other fragments of the image, which could also provide a context for authentication. In the proposed scheme, the presence of false negative authentication results renders the prospective reconstruction result indiscernible. The erroneous symbols will propagate artifacts over the whole reconstruction reference. By exploiting the reference information in the authentication step, it should be possible to provide robustness against collage and VQ attacks without the need for an external context. The design of an efficient procedure for this purpose remains an open research problem.

B

Comparison with Reference Sharing

In this appendix I compare the proposed reconstruction method, with the reference sharing mechanism described in [76]. I will show that both techniques are asymptotically equivalent, and are essentially two different approaches to practical implementation of the same high-level paradigm. The fundamental concept involves spreading of the reference information over the whole image, and exploitation of the remaining authentic content to aid the restoration process. For the sake of retaining consistent notation with the referenced paper, the symbols L , and $q(\cdot)$ will be used in their original meaning. The symbols will be defined, when necessary, and their new meaning remains valid in this appendix only.

I consider the scheme [76]-A, which features uniform reconstruction quality, and uses 320 most significant bits per 8×8 px image block as reference information. The bits are randomly permuted, and organized into L -bit subsets. A N' pixel image contains $5N'/L$ such subsets. To fit into the available watermark capacity, the subsets are then projected onto $L/2$ -bit vectors with the use of a pseudo-random binary matrix of size $L/2 \times L$. After concatenation, the reconstruction reference is randomly permuted, and embedded into 3 least significant bit-planes.

Therefore, the method divides the reconstruction problem into smaller problems, where the spreading mechanism is applied locally to randomly selected image portions. The probability of successful reconstruction is:

$$P_S = P_{LI}^{5N'/L}, \quad (\text{B.1})$$

where P_{LI} is the probability of success within a single subset. It is calculated with the use of two binomial distributions, and a recursive formula for the probability that a random binary matrix is of sufficient rank. Asymptotically, I consider the case of $L = 5N'$, i.e., with a single subset covering the whole image. Such a configuration is equivalent to the proposed approach operating on 1-bit symbols (Section 2.3).

If there exists only a single subset, the amount of tampered, and extractable elements no longer has a stochastic character, and the problem resolves to solving a $(1 - \tilde{\gamma}) \cdot 5N'/2 \times \tilde{\gamma} \cdot 5N'$ linear system in $\text{GF}(2)$ arithmetic. It becomes possible if the number of columns is at most equal to the number of rows, i.e.:

$$(1 - \tilde{\gamma}_z) \cdot 5N'/2 = \tilde{\gamma}_z \cdot 5N' \Rightarrow \tilde{\gamma}_z = \frac{1}{3}. \quad (\text{B.2})$$

The obtained asymptotic bound is identical to $\tilde{\gamma}_2$ in our approach (2.11b). I will now analyze the reconstruction success probability, and demonstrate how it converges to a threshold in $\tilde{\gamma}_z$. Since the original calculation procedure from [76] is not feasible for the considered problem size, I will use it in a different form. Let $E[\tilde{\gamma}]$ denote the probability of tampering a single image block, and also the expected tampering rate. The success probability for a single subset is:

$$P_{LI} = \sum_{i=0}^{L/2} \sum_{j=0}^L P_v(i) P_{nT}(j) (1 - q(i, j)), \quad (\text{B.3})$$

where:

$$P_{nT}(j) = e^{\ln\binom{L}{j} + j \cdot \ln(E[\tilde{\gamma}]) + (L-j) \cdot \ln(1-E[\tilde{\gamma}])}, \quad (\text{B.4a})$$

$$P_v(i) = e^{\ln\binom{L/2}{i} + i \cdot \ln(1-E[\tilde{\gamma}]) + (L/2-i) \cdot \ln(E[\tilde{\gamma}])}. \quad (\text{B.4b})$$

In such form, P_{nT} and P_v can be efficiently calculated by using the logarithmic gamma function to obtain the binomial coefficients. The term $1 - q(i, j)$ denotes the probability that a random binary matrix of size $i \times j$ has sufficient rank. Instead of the recursive formula from [76], I approximate it as:

$$q(i, j) \approx \begin{cases} 1, & \text{if } j > i, \\ 2^{-i}, & \text{if } j = 1, \\ 0.712, & \text{if } j = i, \\ 2^{j-i}, & \text{otherwise.} \end{cases} \quad (\text{B.5})$$

The approximation is founded on boundary analysis [10]:

Lemma 1. *Let V be a vector space of dimension n over $\text{GF}(q)$ and let $m \geq n$. Then, the probability that m random vectors in V span the whole space V is:*

$$\prod_{i=1}^n \left(1 - \frac{1}{q^{m-n+i}}\right) \geq \begin{cases} 0.288, & \text{if } m = n \text{ and } q = 2, \\ 1 - \frac{1}{q^{m-n}(q-1)}, & \text{otherwise.} \end{cases}$$

Equivalently, this also bounds the probability that a random $m \times n$ matrix over $\text{GF}(q)$ has rank n .

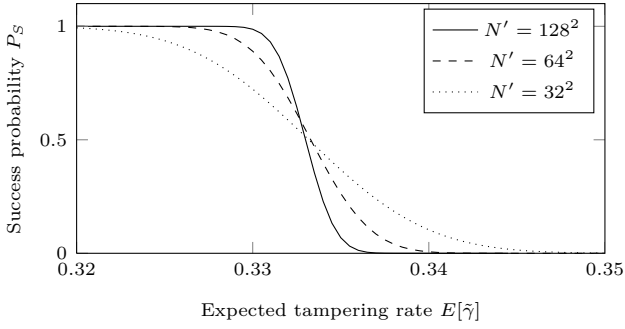
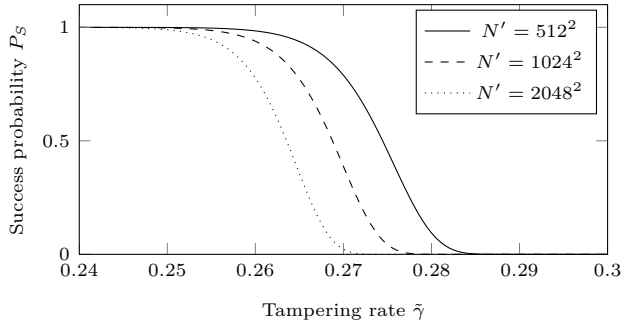
(a) Asymptotic variant, $L = 5N'$ (b) Original variant, $L = 512$

Fig. B.1: Probability of successful recovery of the reference sharing mechanism for different image sizes.

With a fixed $L = 5N'$, the success probability has only one degree of freedom, i.e., the size of the image. As N' increases, the reconstruction becomes more probable around $\tilde{\gamma}_z$. Fig. B.1a shows P_S vs. $E[\tilde{\gamma}]$ for different images sizes. Just as expected, the slope becomes steeper, and the curve approaches a threshold in $\tilde{\gamma}_z$.

It can be concluded that the proposed spreading technique, and the one from [76] are essentially two different methods of practical implementation of the same high-level concept. Instead of dividing the reconstruction problem into smaller fragments, our approach uses M-ary symbols for reference information processing in a single spreading process.

Our approach has three major benefits. Firstly, it does not suffer from an inherent performance penalty, and can still reach the optimistic success bound $\tilde{\gamma}_2$. Given the same rate of reference information, it allows for working with higher

tampering rates, e.g., for 320 bits per block ($\lambda = 2$), the difference in the maximal tampering rate is 0.33 vs. 0.24. Secondly, the method from [76] is sensitive to image size, and the performance deteriorates as N' increases (Fig. B.1b). Thirdly, the proposed approach, formulated in terms of digital fountain coding, can be conveniently analyzed with communication theoretic tools. The applicable success bounds are expressed in terms of well-defined formulas for an arbitrary configuration of the system.

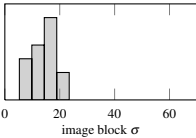
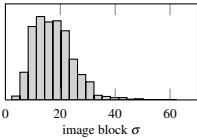
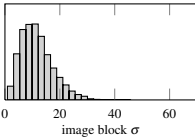
C

Image Tests Sets

The experiments carried out in this study were performed on three popular tests sets of images with natural statistics: *sipi*, *ucid*, and *bows*. From each of the test sets, a representative sub-set is selected for quantization code-book and distortion model training. The selection was made to include bright, subdued, and dark images spanning a wide range of possible texture levels. The amount of texture was measured as a mean sample standard deviation of individual image blocks σ .

A summary of relevant information about each of the considered test sets is collected in Table C.1. The reported values of the image block σ correspond to the minimum, the median, and the maximum values after removing 1% of the outliers according to (4.9).

Table C.1: Summary of the considered data sets.

	sipi	ucid	bows
# images	15	1,338	10,000
Resolution	512×512 px	512×384 px	512×512 px
Image mode	grayscale	average of RGB	grayscale
Block σ	5.29/15.33/23.46	2.30/16.83/41.26	0.91/10.78/28.53
Training set σ	5.29/15.33/23.46	7.26/19.03/31.27	3.14/12.03/20.69
Statistics			

The *sipi* set contains 15 images commonly used in image processing research, selected from the USC SIPI data set [1]. The training set is the same as the test set, and contains all 15 images (Fig. C.1).

The *bows* test set comes from the 2nd break our watermarking system (BOWS) contest [5], and contains 10,000 uncompressed gray-scale 512×512 px images. The corresponding training set contains 16 representative images (Fig. C.2).

The *ucid* test set contains 1,338 images from the uncompressed colour image database (UCID) [59]. The images are 512×384 px, and have been converted to gray-scale by averaging the red, green, and blue components. The training set contains 15 representative image (Fig. C.3).



Fig. C.1: Training images from the *sipi* data set.



Fig. C.2: Training images from the *bows* data set.



Fig. C.3: Training images from the *ucid* data set.

D

Solutions to the Optimal Reference Allocation Problem

This appendix collects the obtained solutions to the reference payload allocation problem for the most important configurations of the considered self-embedding schemes. The presented results are shown by means of allocation matrices \mathbf{V} , obtained by the proposed optimization procedure described in Algorithm 3, and by a general integer nonlinear programming (INLP) solver.

The results obtained with both solvers are nearly identical. The differences stem from the additional flexibility of the proposed procedure, which solves the full 64-element optimization, instead of the simplified 15-element version of the problem. For reference, the tables also show intermediate results, with allocation vectors obtained from various starting points for the INLP solver. For the sake of presentation clarity, the value of the objective function for each of the vectors is denoted briefly as θ . For more information, please refer to Section 4.1.

One of the main advantages of the proposed algorithm is the computation speed. Intuitively, the time required to perform the optimization increases along with the number of reference bits b . The INLP solver typically yields a solution in between 4 seconds and 2 minutes. An extended version of the problem, with additional constraints for the minimal allowed coefficient precision, the optimization takes several minutes for b around 40, and becomes prohibitively time consuming around $b = 80$. The proposed solver yields the result below 2 s, even for the highest meaningful b , regardless of the constraint on the minimum allowed coefficient precision. The calculations were performed on a dual core 2.67 GHz Core i5 processor.

Table D.1: Solutions to the optimal resource allocation problem for the Lloyd-Max code-book, $L = 3$ and the *bows* training set (*cont.*).

Dedicated solver	INLP, simplified optimization
$b = 40$ bpb, $\theta(\mathbf{V}) = 1.625$	$b = 40$ bpb, $\theta(\mathbf{V}) = 1.625$
$\begin{bmatrix} 5 & 4 & 3 & 2 & 2 & 0 & 0 & 0 \\ 4 & 3 & 2 & 2 & 0 & 0 & 0 & 0 \\ 3 & 2 & 2 & 0 & 0 & 0 & 0 & 0 \\ 2 & 2 & 0 & 0 & 0 & 0 & 0 & 0 \\ 2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$	$\begin{bmatrix} 5 & 4 & 3 & 2 & 2 & 0 & 0 & 0 \\ 4 & 3 & 2 & 2 & 0 & 0 & 0 & 0 \\ 3 & 2 & 2 & 0 & 0 & 0 & 0 & 0 \\ 2 & 2 & 0 & 0 & 0 & 0 & 0 & 0 \\ 2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$
INLP solution chosen from:	
$\mathbf{v}_{\text{exp}}^+$	$[5, 4, 3, 2, 2] \quad \theta = 1.62$
$\mathbf{v}_{\text{lin}}^+$	$[6, 3, 3, 2, 1, 1] \quad \theta = 1.75$
$\mathbf{v}_{\text{ref}}^+$	$[5, 4, 3, 2, 2] \quad \theta = 1.62$
$b = 80$ bpb, $\theta(\mathbf{V}) = 0.773$	$b = 80$ bpb, $\theta(\mathbf{V}) = 0.791$
$\begin{bmatrix} 6 & 5 & 4 & 3 & 3 & 2 & 2 & 0 \\ 5 & 4 & 3 & 3 & 2 & 1 & 0 & 0 \\ 4 & 3 & 3 & 2 & 2 & 0 & 0 & 0 \\ 3 & 3 & 2 & 2 & 0 & 0 & 0 & 0 \\ 3 & 2 & 2 & 0 & 0 & 0 & 0 & 0 \\ 2 & 2 & 0 & 0 & 0 & 0 & 0 & 0 \\ 2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$	$\begin{bmatrix} 7 & 4 & 4 & 3 & 3 & 2 & 2 & 0 \\ 4 & 4 & 3 & 3 & 2 & 2 & 0 & 0 \\ 4 & 3 & 3 & 2 & 2 & 0 & 0 & 0 \\ 3 & 3 & 2 & 2 & 0 & 0 & 0 & 0 \\ 3 & 2 & 2 & 0 & 0 & 0 & 0 & 0 \\ 2 & 2 & 0 & 0 & 0 & 0 & 0 & 0 \\ 2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$
INLP solution chosen from:	
$\mathbf{v}_{\text{exp}}^+$	$[7, 4, 4, 3, 3, 2, 2] \quad \theta = 0.79$
$\mathbf{v}_{\text{lin}}^+$	$[7, 6, 4, 3, 2, 2, 1, 1] \quad \theta = 0.83$
$\mathbf{v}_{\text{ref}}^+$	$[7, 4, 4, 3, 3, 2, 2] \quad \theta = 0.79$

Table D.1: Solutions to the optimal resource allocation problem for the Lloyd-Max code-book, $L = 3$ and the *bows* training set.

Dedicated solver	INLP, simplified optimization
$b = 160$ bpb, $\theta(\mathbf{V}) = 0.230$	$b = 160$ bpb, $\theta(\mathbf{V}) = 0.230$
$\begin{bmatrix} 8 & 6 & 5 & 4 & 4 & 3 & 3 & 3 \\ 6 & 5 & 4 & 4 & 3 & 3 & 3 & 2 \\ 5 & 4 & 4 & 3 & 3 & 3 & 2 & 2 \\ 4 & 4 & 3 & 3 & 3 & 2 & 2 & 0 \\ 4 & 3 & 3 & 3 & 2 & 2 & 0 & 0 \\ 3 & 3 & 3 & 2 & 2 & 0 & 0 & 0 \\ 3 & 3 & 2 & 2 & 0 & 0 & 0 & 0 \\ 3 & 2 & 2 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$	$\begin{bmatrix} 8 & 6 & 5 & 4 & 4 & 3 & 3 & 3 \\ 6 & 5 & 4 & 4 & 3 & 3 & 3 & 2 \\ 5 & 4 & 4 & 3 & 3 & 3 & 2 & 2 \\ 4 & 4 & 3 & 3 & 3 & 2 & 2 & 0 \\ 4 & 3 & 3 & 3 & 2 & 2 & 0 & 0 \\ 3 & 3 & 3 & 2 & 2 & 0 & 0 & 0 \\ 3 & 3 & 2 & 2 & 0 & 0 & 0 & 0 \\ 3 & 2 & 2 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$
INLP solution chosen from:	
$\mathbf{v}_{\text{exp}}^+$	$[8, 6, 5, 4, 4, 3, 3, 2, 2]$ $\theta = 0.23$
$\mathbf{v}_{\text{lin}}^+$	$[7, 6, 5, 4, 4, 3, 3, 2, 2, 1, 1]$ $\theta = 0.24$
$\mathbf{v}_{\text{ref}}^+$	$[8, 6, 5, 4, 4, 3, 3, 2, 2]$ $\theta = 0.23$
$b = 320$ bpb, $\theta(\mathbf{V}) = 0.018$	$b = 320$ bpb, $\theta(\mathbf{V}) = 0.019$
$\begin{bmatrix} 8 & 8 & 7 & 7 & 6 & 6 & 5 & 5 \\ 8 & 7 & 7 & 6 & 6 & 5 & 5 & 4 \\ 7 & 7 & 6 & 6 & 5 & 5 & 5 & 4 \\ 7 & 6 & 6 & 5 & 5 & 5 & 4 & 4 \\ 6 & 6 & 5 & 5 & 4 & 4 & 4 & 3 \\ 6 & 5 & 5 & 5 & 4 & 4 & 3 & 3 \\ 5 & 5 & 5 & 4 & 4 & 3 & 3 & 3 \\ 5 & 5 & 4 & 4 & 3 & 3 & 3 & 2 \end{bmatrix}$	$\begin{bmatrix} 8 & 8 & 8 & 7 & 6 & 5 & 5 & 5 \\ 8 & 8 & 7 & 6 & 5 & 5 & 5 & 5 \\ 8 & 7 & 6 & 5 & 5 & 5 & 5 & 4 \\ 7 & 6 & 5 & 5 & 5 & 5 & 4 & 4 \\ 6 & 5 & 5 & 5 & 5 & 4 & 4 & 3 \\ 5 & 5 & 5 & 5 & 4 & 4 & 3 & 3 \\ 5 & 5 & 5 & 4 & 4 & 3 & 3 & 3 \\ 5 & 5 & 4 & 4 & 3 & 3 & 3 & 3 \end{bmatrix}$
INLP solution chosen from:	
$\mathbf{v}_{\text{exp}}^+$	$[8, 8, 8, 7, 6, 5, 5, 5, 5, 4, 4, 3, 3, 3, 3]$ $\theta = 0.02$
$\mathbf{v}_{\text{lin}}^+$	$[8, 8, 7, 7, 6, 6, 5, 5, 5, 4, 4, 3, 3, 2, 2]$ $\theta = 0.02$
$\mathbf{v}_{\text{ref}}^+$	$[8, 8, 7, 7, 7, 6, 5, 5, 4, 4, 4, 3, 3, 3, 2]$ $\theta = 0.02$

Table D.2: Solutions to the optimal resource allocation problem for the uniform code-book, $L = 3$ and the *bows* training set (*cont.*).

Dedicated solver	INLP, simplified optimization
$b = 40$ bpb, $\theta(\mathbf{V}) = 2.141$	$b = 40$ bpb, $\theta(\mathbf{V}) = 2.141$
$\begin{bmatrix} 6 & 5 & 4 & 3 & 0 & 0 & 0 & 0 \\ 5 & 4 & 3 & 0 & 0 & 0 & 0 & 0 \\ 4 & 3 & 0 & 0 & 0 & 0 & 0 & 0 \\ 3 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$	$\begin{bmatrix} 6 & 5 & 4 & 3 & 0 & 0 & 0 & 0 \\ 5 & 4 & 3 & 0 & 0 & 0 & 0 & 0 \\ 4 & 3 & 0 & 0 & 0 & 0 & 0 & 0 \\ 3 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$
INLP solution chosen from:	
$\mathbf{v}_{\text{exp}}^+$	$[6, 5, 4, 3] \quad \theta = 2.14$
$\mathbf{v}_{\text{lin}}^+$	$[6, 3, 3, 2, 1, 1] \quad \theta = 2.95$
$\mathbf{v}_{\text{ref}}^+$	$[6, 5, 4, 3] \quad \theta = 2.14$
$b = 80$ bpb, $\theta(\mathbf{V}) = 1.137$	$b = 80$ bpb, $\theta(\mathbf{V}) = 1.149$
$\begin{bmatrix} 6 & 6 & 5 & 4 & 3 & 0 & 0 & 0 \\ 6 & 5 & 4 & 3 & 3 & 0 & 0 & 0 \\ 5 & 4 & 3 & 3 & 0 & 0 & 0 & 0 \\ 4 & 3 & 3 & 0 & 0 & 0 & 0 & 0 \\ 4 & 3 & 0 & 0 & 0 & 0 & 0 & 0 \\ 3 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$	$\begin{bmatrix} 6 & 5 & 5 & 4 & 3 & 3 & 0 & 0 \\ 5 & 5 & 4 & 3 & 3 & 0 & 0 & 0 \\ 5 & 4 & 3 & 3 & 0 & 0 & 0 & 0 \\ 4 & 3 & 3 & 0 & 0 & 0 & 0 & 0 \\ 3 & 3 & 0 & 0 & 0 & 0 & 0 & 0 \\ 3 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$
INLP solution chosen from:	
$\mathbf{v}_{\text{exp}}^+$	$[6, 5, 5, 4, 3, 3] \quad \theta = 1.15$
$\mathbf{v}_{\text{lin}}^+$	$[7, 6, 4, 3, 2, 2, 1, 1] \quad \theta = 1.47$
$\mathbf{v}_{\text{ref}}^+$	$[6, 5, 5, 4, 3, 3] \quad \theta = 1.15$

Table D.2: Solutions to the optimal resource allocation problem for the uniform code-book, $L = 3$ and the *bows* training set.

Dedicated solver	INLP, simplified optimization
$b = 160$ bpb, $\theta(\mathbf{V}) = 0.415$	$b = 160$ bpb, $\theta(\mathbf{V}) = 0.415$
$\begin{bmatrix} 7 & 7 & 6 & 5 & 5 & 4 & 4 & 3 \\ 7 & 6 & 5 & 5 & 4 & 4 & 3 & 0 \\ 6 & 5 & 5 & 4 & 4 & 3 & 0 & 0 \\ 5 & 5 & 4 & 4 & 3 & 0 & 0 & 0 \\ 5 & 4 & 4 & 3 & 0 & 0 & 0 & 0 \\ 4 & 4 & 3 & 0 & 0 & 0 & 0 & 0 \\ 4 & 3 & 0 & 0 & 0 & 0 & 0 & 0 \\ 3 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$	$\begin{bmatrix} 7 & 7 & 6 & 5 & 5 & 4 & 4 & 3 \\ 7 & 6 & 5 & 5 & 4 & 4 & 3 & 0 \\ 6 & 5 & 5 & 4 & 4 & 3 & 0 & 0 \\ 5 & 5 & 4 & 4 & 3 & 0 & 0 & 0 \\ 5 & 4 & 4 & 3 & 0 & 0 & 0 & 0 \\ 4 & 4 & 3 & 0 & 0 & 0 & 0 & 0 \\ 4 & 3 & 0 & 0 & 0 & 0 & 0 & 0 \\ 3 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$
INLP solution chosen from:	
$\mathbf{v}_{\text{exp}}^+$	$[7, 7, 6, 5, 5, 4, 4, 3]$ $\theta = 0.42$
$\mathbf{v}_{\text{lin}}^+$	$[7, 6, 5, 4, 4, 3, 3, 2, 2, 1, 1]$ $\theta = 0.55$
$\mathbf{v}_{\text{ref}}^+$	$[7, 7, 6, 5, 5, 4, 4, 3]$ $\theta = 0.42$
$b = 320$ bpb, $\theta(\mathbf{V}) = 0.045$	$b = 320$ bpb, $\theta(\mathbf{V}) = 0.045$
$\begin{bmatrix} 8 & 8 & 7 & 7 & 6 & 6 & 5 & 5 \\ 8 & 7 & 7 & 6 & 6 & 5 & 5 & 4 \\ 7 & 7 & 6 & 6 & 5 & 5 & 4 & 4 \\ 7 & 6 & 6 & 5 & 5 & 4 & 4 & 4 \\ 6 & 6 & 5 & 5 & 4 & 4 & 4 & 4 \\ 6 & 5 & 5 & 5 & 4 & 4 & 4 & 3 \\ 5 & 5 & 5 & 4 & 4 & 4 & 3 & 3 \\ 5 & 5 & 4 & 4 & 4 & 3 & 3 & 0 \end{bmatrix}$	$\begin{bmatrix} 8 & 8 & 7 & 7 & 6 & 6 & 5 & 5 \\ 8 & 7 & 7 & 6 & 6 & 5 & 5 & 4 \\ 7 & 7 & 6 & 6 & 5 & 5 & 4 & 4 \\ 7 & 6 & 6 & 5 & 5 & 4 & 4 & 4 \\ 6 & 6 & 5 & 5 & 4 & 4 & 4 & 4 \\ 6 & 5 & 5 & 4 & 4 & 4 & 4 & 3 \\ 5 & 5 & 4 & 4 & 4 & 4 & 3 & 3 \\ 5 & 4 & 4 & 4 & 4 & 3 & 3 & 3 \end{bmatrix}$
INLP solution chosen from:	
$\mathbf{v}_{\text{exp}}^+$	$[8, 8, 7, 7, 6, 6, 5, 5, 4, 4, 4, 4, 3, 3, 3]$ $\theta = 0.05$
$\mathbf{v}_{\text{lin}}^+$	$[8, 8, 8, 7, 6, 6, 5, 5, 4, 4, 4, 3, 3, 3, 4]$ $\theta = 0.05$
$\mathbf{v}_{\text{ref}}^+$	$[8, 8, 7, 7, 6, 6, 5, 5, 4, 4, 4, 4, 3, 3, 3]$ $\theta = 0.05$

Table D.3: Solutions to the optimal resource allocation problem for the Lloyd-Max code-book, $L = 2$ and the *bows* training set (*cont.*).

Dedicated solver	INLP, simplified optimization
$b = 48$ bpb, $\theta(\mathbf{V}) = 5.376$	$b = 48$ bpb, $\theta(\mathbf{V}) = 5.376$
$\begin{bmatrix} 6 & 4 & 3 & 3 & 2 & 0 & 0 & 0 \\ 5 & 3 & 3 & 2 & 0 & 0 & 0 & 0 \\ 3 & 3 & 2 & 0 & 0 & 0 & 0 & 0 \\ 3 & 2 & 0 & 0 & 0 & 0 & 0 & 0 \\ 2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$	$\begin{bmatrix} 7 & 4 & 3 & 2 & 2 & 1 & 0 & 0 \\ 4 & 3 & 2 & 2 & 1 & 0 & 0 & 0 \\ 3 & 2 & 2 & 1 & 0 & 0 & 0 & 0 \\ 2 & 2 & 1 & 0 & 0 & 0 & 0 & 0 \\ 2 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$
INLP solution chosen from:	
$\mathbf{v}_{\text{exp}}^+$	$[7, 4, 3, 2, 2, 1] \quad \theta = 5.38$
$\mathbf{v}_{\text{lin}}^+$	$[7, 4, 3, 2, 2, 1] \quad \theta = 5.38$
$\mathbf{v}_{\text{ref}}^+$	$[7, 4, 3, 2, 2, 1] \quad \theta = 5.38$
$b = 96$ bpb, $\theta(\mathbf{V}) = 2.322$	$b = 96$ bpb, $\theta(\mathbf{V}) = 2.326$
$\begin{bmatrix} 7 & 5 & 4 & 4 & 3 & 3 & 2 & 0 \\ 5 & 4 & 4 & 3 & 3 & 2 & 0 & 0 \\ 4 & 4 & 3 & 3 & 2 & 0 & 0 & 0 \\ 4 & 3 & 3 & 2 & 0 & 0 & 0 & 0 \\ 3 & 3 & 2 & 0 & 0 & 0 & 0 & 0 \\ 3 & 2 & 0 & 0 & 0 & 0 & 0 & 0 \\ 2 & 2 & 0 & 0 & 0 & 0 & 0 & 0 \\ 2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$	$\begin{bmatrix} 7 & 5 & 4 & 3 & 3 & 3 & 2 & 1 \\ 5 & 4 & 3 & 3 & 3 & 2 & 1 & 0 \\ 4 & 3 & 3 & 3 & 2 & 1 & 0 & 0 \\ 3 & 3 & 3 & 2 & 1 & 0 & 0 & 0 \\ 3 & 3 & 2 & 1 & 0 & 0 & 0 & 0 \\ 3 & 2 & 1 & 0 & 0 & 0 & 0 & 0 \\ 2 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$
INLP solution chosen from:	
$\mathbf{v}_{\text{exp}}^+$	$[7, 5, 4, 3, 3, 3, 2, 1] \quad \theta = 2.33$
$\mathbf{v}_{\text{lin}}^+$	$[6, 5, 4, 3, 3, 2, 2, 1] \quad \theta = 2.43$
$\mathbf{v}_{\text{ref}}^+$	$[7, 5, 4, 3, 3, 3, 2, 1] \quad \theta = 2.33$

Table D.3: Solutions to the optimal resource allocation problem for the Lloyd-Max code-book, $L = 2$ and the *bows* training set.

Dedicated solver	INLP, simplified optimization
$b = 192$ bpb, $\theta(\mathbf{V}) = 0.526$	$b = 192$ bpb, $\theta(\mathbf{V}) = 0.526$
$\begin{bmatrix} 8 & 7 & 6 & 5 & 4 & 4 & 3 & 3 \\ 7 & 6 & 5 & 4 & 4 & 3 & 3 & 3 \\ 6 & 5 & 4 & 4 & 3 & 3 & 3 & 2 \\ 5 & 4 & 4 & 3 & 3 & 3 & 2 & 2 \\ 4 & 4 & 3 & 3 & 3 & 2 & 2 & 0 \\ 4 & 3 & 3 & 3 & 2 & 2 & 0 & 0 \\ 3 & 3 & 3 & 2 & 2 & 0 & 0 & 0 \\ 3 & 3 & 2 & 2 & 0 & 0 & 0 & 0 \end{bmatrix}$	$\begin{bmatrix} 8 & 7 & 6 & 5 & 4 & 4 & 3 & 3 \\ 7 & 6 & 5 & 4 & 4 & 3 & 3 & 3 \\ 6 & 5 & 4 & 4 & 3 & 3 & 3 & 2 \\ 5 & 4 & 4 & 3 & 3 & 3 & 2 & 2 \\ 4 & 4 & 3 & 3 & 3 & 2 & 2 & 0 \\ 4 & 3 & 3 & 3 & 2 & 2 & 0 & 0 \\ 3 & 3 & 3 & 2 & 2 & 0 & 0 & 0 \\ 3 & 3 & 2 & 2 & 0 & 0 & 0 & 0 \end{bmatrix}$
INLP solution chosen from:	
$\mathbf{v}_{\text{exp}}^+$	$[8, 6, 5, 4, 4, 3, 3, 2, 2] \quad \theta = 0.53$
$\mathbf{v}_{\text{lin}}^+$	$[8, 7, 6, 5, 4, 4, 3, 3, 2, 2, 1, 1] \quad \theta = 0.54$
$\mathbf{v}_{\text{ref}}^+$	$[8, 7, 6, 5, 4, 4, 3, 3, 3, 2, 2] \quad \theta = 0.53$
$b = 288$ bpb, $\theta(\mathbf{V}) = 0.110$	$b = 288$ bpb, $\theta(\mathbf{V}) = 0.111$
$\begin{bmatrix} 8 & 8 & 7 & 6 & 6 & 5 & 5 & 4 \\ 8 & 7 & 6 & 6 & 5 & 5 & 4 & 4 \\ 7 & 6 & 6 & 5 & 5 & 4 & 4 & 4 \\ 6 & 6 & 5 & 5 & 4 & 4 & 4 & 3 \\ 6 & 5 & 5 & 4 & 4 & 4 & 3 & 3 \\ 5 & 5 & 4 & 4 & 4 & 3 & 3 & 2 \\ 5 & 4 & 4 & 4 & 3 & 3 & 2 & 2 \\ 4 & 4 & 4 & 3 & 3 & 3 & 2 & 2 \end{bmatrix}$	$\begin{bmatrix} 8 & 8 & 7 & 6 & 6 & 5 & 5 & 4 \\ 8 & 7 & 6 & 6 & 5 & 5 & 4 & 4 \\ 7 & 6 & 6 & 5 & 5 & 4 & 4 & 4 \\ 6 & 6 & 5 & 5 & 4 & 4 & 4 & 3 \\ 6 & 5 & 5 & 4 & 4 & 4 & 3 & 3 \\ 5 & 5 & 4 & 4 & 4 & 3 & 3 & 2 \\ 5 & 4 & 4 & 4 & 3 & 3 & 2 & 2 \\ 4 & 4 & 4 & 3 & 3 & 2 & 2 & 3 \end{bmatrix}$
INLP solution chosen from:	
$\mathbf{v}_{\text{exp}}^+$	$[8, 8, 7, 6, 6, 5, 5, 4, 4, 4, 3, 3, 2, 2, 3] \quad \theta = 0.11$
$\mathbf{v}_{\text{lin}}^+$	$[8, 8, 7, 7, 6, 5, 5, 4, 4, 3, 3, 3, 3, 2, 2] \quad \theta = 0.11$
$\mathbf{v}_{\text{ref}}^+$	$[8, 8, 7, 6, 6, 5, 5, 4, 4, 4, 3, 3, 2, 2, 3] \quad \theta = 0.11$

Table D.4: Solutions to the optimal resource allocation problem for the Lloyd-Max code-book, $L = 2$ and the *bows* training set (*cont.*).

Dedicated solver	INLP, simplified optimization
$b = 44$ bpb, $\theta(\mathbf{V}) = 5.784$	$b = 44$ bpb, $\theta(\mathbf{V}) = 5.873$
$\begin{bmatrix} 6 & 4 & 3 & 3 & 2 & 0 & 0 & 0 \\ 4 & 3 & 2 & 2 & 0 & 0 & 0 & 0 \\ 3 & 3 & 2 & 0 & 0 & 0 & 0 & 0 \\ 3 & 2 & 0 & 0 & 0 & 0 & 0 & 0 \\ 2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$	$\begin{bmatrix} 6 & 4 & 4 & 2 & 2 & 0 & 0 & 0 \\ 4 & 4 & 2 & 2 & 0 & 0 & 0 & 0 \\ 4 & 2 & 2 & 0 & 0 & 0 & 0 & 0 \\ 2 & 2 & 0 & 0 & 0 & 0 & 0 & 0 \\ 2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$
INLP solution chosen from:	
$\mathbf{v}_{\text{exp}}^+$	$[6, 4, 4, 2, 2] \quad \theta = 5.87$
$\mathbf{v}_{\text{lin}}^+$	$[6, 5, 3, 2, 1, 1] \quad \theta = 6.11$
$\mathbf{v}_{\text{ref}}^+$	$[6, 4, 4, 2, 2] \quad \theta = 5.87$
$b = 88$ bpb, $\theta(\mathbf{V}) = 2.610$	$b = 88$ bpb, $\theta(\mathbf{V}) = 2.616$
$\begin{bmatrix} 7 & 5 & 4 & 4 & 3 & 2 & 2 & 0 \\ 5 & 4 & 4 & 3 & 2 & 2 & 0 & 0 \\ 4 & 4 & 3 & 2 & 2 & 0 & 0 & 0 \\ 4 & 3 & 3 & 2 & 0 & 0 & 0 & 0 \\ 3 & 2 & 2 & 0 & 0 & 0 & 0 & 0 \\ 3 & 2 & 0 & 0 & 0 & 0 & 0 & 0 \\ 2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$	$\begin{bmatrix} 7 & 5 & 4 & 3 & 3 & 3 & 2 & 0 \\ 5 & 4 & 3 & 3 & 3 & 2 & 0 & 0 \\ 4 & 3 & 3 & 3 & 2 & 0 & 0 & 0 \\ 3 & 3 & 3 & 2 & 0 & 0 & 0 & 0 \\ 3 & 3 & 2 & 0 & 0 & 0 & 0 & 0 \\ 3 & 2 & 0 & 0 & 0 & 0 & 0 & 0 \\ 2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$
INLP solution chosen from:	
$\mathbf{v}_{\text{exp}}^+$	$[7, 5, 4, 3, 3, 3, 2] \quad \theta = 2.62$
$\mathbf{v}_{\text{lin}}^+$	$[8, 5, 4, 4, 3, 2, 1, 1] \quad \theta = 2.77$
$\mathbf{v}_{\text{ref}}^+$	$[7, 5, 4, 3, 3, 3, 2] \quad \theta = 2.62$

Table D.4: Solutions to the optimal resource allocation problem for the Lloyd-Max code-book, $L = 2$ and the *bows* training set.

Dedicated solver	INLP, simplified optimization
$b = 176$ bpb, $\theta(\mathbf{V}) = 0.675$	$b = 176$ bpb, $\theta(\mathbf{V}) = 0.677$
$\begin{bmatrix} 8 & 6 & 5 & 5 & 4 & 4 & 3 & 3 \\ 6 & 5 & 5 & 4 & 4 & 3 & 3 & 2 \\ 5 & 5 & 4 & 4 & 3 & 3 & 2 & 2 \\ 5 & 4 & 4 & 3 & 3 & 2 & 2 & 0 \\ 4 & 4 & 3 & 3 & 2 & 2 & 0 & 0 \\ 4 & 3 & 3 & 2 & 2 & 2 & 0 & 0 \\ 3 & 3 & 2 & 2 & 2 & 0 & 0 & 0 \\ 3 & 2 & 2 & 2 & 0 & 0 & 0 & 0 \end{bmatrix}$	$\begin{bmatrix} 7 & 7 & 5 & 5 & 4 & 4 & 3 & 3 \\ 7 & 5 & 5 & 4 & 4 & 3 & 3 & 2 \\ 5 & 5 & 4 & 4 & 3 & 3 & 2 & 2 \\ 5 & 4 & 4 & 3 & 3 & 2 & 2 & 1 \\ 4 & 4 & 3 & 3 & 2 & 2 & 1 & 0 \\ 4 & 3 & 3 & 2 & 2 & 1 & 0 & 0 \\ 3 & 3 & 2 & 2 & 1 & 0 & 0 & 0 \\ 3 & 2 & 2 & 1 & 0 & 0 & 0 & 0 \end{bmatrix}$
INLP solution chosen from:	
$\mathbf{v}_{\text{exp}}^+$	$[7, 7, 5, 5, 4, 4, 3, 3, 2, 2, 1] \quad \theta = 0.67$
$\mathbf{v}_{\text{lin}}^+$	$[8, 7, 5, 5, 4, 4, 3, 2, 2, 1, 1, 1] \quad \theta = 0.72$
$\mathbf{v}_{\text{ref}}^+$	$[7, 7, 5, 5, 4, 4, 3, 3, 2, 2, 1] \quad \theta = 0.67$
$b = 264$ bpb, $\theta(\mathbf{V}) = 0.164$	$b = 264$ bpb, $\theta(\mathbf{V}) = 0.164$
$\begin{bmatrix} 8 & 8 & 6 & 6 & 5 & 5 & 4 & 4 \\ 8 & 6 & 6 & 5 & 5 & 4 & 4 & 4 \\ 6 & 6 & 5 & 5 & 4 & 4 & 4 & 3 \\ 6 & 5 & 5 & 4 & 4 & 4 & 3 & 3 \\ 5 & 5 & 4 & 4 & 4 & 3 & 3 & 3 \\ 5 & 4 & 4 & 4 & 3 & 3 & 3 & 2 \\ 4 & 4 & 4 & 3 & 3 & 3 & 2 & 2 \\ 4 & 4 & 3 & 3 & 3 & 2 & 2 & 0 \end{bmatrix}$	$\begin{bmatrix} 8 & 8 & 6 & 6 & 5 & 5 & 4 & 4 \\ 8 & 6 & 6 & 5 & 5 & 4 & 4 & 4 \\ 6 & 6 & 5 & 5 & 4 & 4 & 4 & 3 \\ 6 & 5 & 5 & 4 & 4 & 4 & 3 & 3 \\ 5 & 5 & 4 & 4 & 4 & 3 & 3 & 3 \\ 5 & 4 & 4 & 4 & 3 & 3 & 3 & 2 \\ 4 & 4 & 4 & 3 & 3 & 3 & 2 & 2 \\ 4 & 4 & 3 & 3 & 3 & 2 & 2 & 0 \end{bmatrix}$
INLP solution chosen from:	
$\mathbf{v}_{\text{exp}}^+$	$[8, 8, 6, 6, 5, 5, 4, 4, 4, 3, 3, 3, 2, 2] \quad \theta = 0.16$
$\mathbf{v}_{\text{lin}}^+$	$[8, 8, 7, 6, 5, 5, 4, 4, 4, 3, 3, 2, 2, 1] \quad \theta = 0.16$
$\mathbf{v}_{\text{ref}}^+$	$[8, 8, 6, 6, 5, 5, 4, 4, 4, 3, 3, 3, 2, 2] \quad \theta = 0.16$

E

Miscellaneous Practical Implementation Issues

This appendix briefly discussed prospective extensions of the proposed content reconstruction model, which allow for operation on color images, and for handling block classification, and watermark extraction errors. The latter are of particular importance when dealing with lossy-compressed formats.

Extension to color images can operate in one of two possible modes. Assuming gray-scale reconstruction is sufficient, the additional embedding capacity can be used to provide more redundancy for the luminance channel. Then, assuming no rounding errors, the restoration condition becomes:

$$n_c \gamma \geq \lambda(1 - \gamma) \Rightarrow \gamma \geq \lambda(n_c + \lambda)^{-1} \quad (\text{E.1})$$

where n_c denotes the number of available same-capacity channels. For $n_c = 2$, this improves the supported tampering rates by up to 17% of the image area.

Alternatively, the additional capacity can be used for the reconstruction of the chrominance channels. The reconstruction is performed independently from the luminance component, and the success bounds is (E.1) for $n_c = 1$.

Due to coefficient rounding errors during prospective image editing, unintentional bit flips either in the blocks' reference information, or the embedded payload, make it possible for authentic image blocks to be misclassified as tampered. Such blocks would be restored in the decoder, and would unnecessarily limit the achievable tampering rates. Given false positive probability f_p , the restoration condition becomes:

$$(1 - f_p)\gamma \geq \lambda(1 - \gamma) + \lambda\gamma f_p \quad (\text{E.2a})$$

$$\gamma \geq \lambda(1 - f_p + \lambda(1 - f_p))^{-1}. \quad (\text{E.2b})$$

In order to distinguish blocks with erroneous content, and corrupted payload, a dual-hash mechanism could be adopted. In case an image blocks is authentic,

Table E.1: Reconstruction success bounds in the presence of false positive classification errors.

Mode	λ	$\tilde{\gamma}_{\max}$ for symbol error rate p_e				
		0.0	0.01	0.05	0.10	0.15
Single-hash	1	50.0	49.5	47.4	44.4	41.2
Single-hash	2	33.3	32.7	29.8	25.9	21.6
Single-hash	3	25.0	24.2	21.1	16.7	11.8
Single-hash	4	20.0	19.2	15.8	11.1	5.9
False positive rate $f_p = 0.01$						
Dual-hash	1	49.7	49.5	48.5	47.1	45.7
Dual-hash	2	32.9	32.7	31.7	30.6	29.3
Dual-hash	3	24.4	24.2	23.5	22.5	21.5
Dual-hash	4	19.4	19.2	18.5	17.7	16.8
False positive rate $f_p = 0.05$						
Dual-hash	1	48.7	48.5	47.4	45.9	44.4
Dual-hash	2	31.0	30.8	29.8	28.6	27.3
Dual-hash	3	22.1	21.9	21.1	20.0	18.9
Dual-hash	4	16.7	16.5	15.8	14.9	14.0

yet contains an invalid payload, it will not be reconstructed. Let p_e denote the watermark symbol error rate. Then, the reconstruction condition becomes:

$$(1 - p_e)\gamma \geq \lambda(1 - \gamma) + \lambda\gamma f_p \quad (\text{E.3a})$$

$$\gamma \geq \lambda(1 - p_e + \lambda(1 - f_p))^{-1}. \quad (\text{E.3b})$$

The false positive classification errors are typically significantly less frequent than watermark symbol errors, i.e., $f_p \ll p_e$. Hence, the introduction of a dual-hash mechanism allows for higher tampering rates. Table E.1 collects the theoretically achievable tampering rates for both a single and a dual-hash configuration.

References

- [1] The sipi dataset, university of southern california. <http://sipi.usc.edu/database/>. Visited on 26 March 2012.
- [2] Photoshop cs6 / content-aware fill. <http://www.adobe.com/products/photoshop/content-aware-fill.html>. Visited on 29 July 2012.
- [3] Datamark. <http://www.datamark.com.sg/>. Visited on 29 July 2012.
- [4] Handy photolab / touch retouch. http://www.handyphotolab.com/application_tr. Visited on 29 July 2012.
- [5] The dataset from the 2nd bows contest. <http://bows2.ec-lille.fr/>, 2007. Visited on 26 March 2012.
- [6] Chowdary. B. Adsumilli, Mylène. C. Q. Farias, Sanjit. K. Mitra, and Marco. Carli. A robust error concealment technique using data hiding for image and video transmission over lossy channels. *IEEE Transactions on Circuits and Systems for Video Technology*, 15(11):1394–1406, 2005. doi: 10.1109/TCSVT.2005.856933.
- [7] Sergio Bravo-Solorio and Asoke K. Nandi. Secure fragile watermarking method for image authentication with improved tampering localisation and self-recovery capabilities. *Signal Processing*, 91(4):728 – 739, 2011. ISSN 0165-1684. doi: 10.1016/j.sigpro.2010.07.019.
- [8] Sergio Bravo-Solorio, Chang-Tsun Li, and Asoke K. Nandi. Watermarking with low embedding distortion and self-propagating restoration capabilities. In *Proc. of IEEE International Conference on Image Processing*, Orlando, FL, 2012. URL <http://sbravo.x10.mx/home/repo/bravo-icip12.pdf>.

- [9] Sergio Bravo-Solorio, Chang-Tsun Li, and Asoke K. Nandi. Watermarking method with exact self-propagating restoration capabilities. In *Proc. of IEEE International Workshop on Information Forensics and Security*, Tenerife, 2012. URL <http://sbravo.x10.mx/home/repo/bravo-wifs12.pdf>.
- [10] Richard P. Brent, Shuhong Gao, and Alan G. B. Lauder. Random Krylov spaces over finite fields. *SIAM Journal on Discrete Mathematics*, 16(2): 276–287, February 2003. ISSN 0895-4801. doi: 10.1137/S089548010139388X.
- [11] Chin-Chen Chang, Yi-Hsuan Fan, and Wei-Liang Tai. Four-scanning attack on hierarchical digital watermarking method for image tamper detection and recovery. *Pattern Recognition*, 41(2):654 – 661, 2008. ISSN 0031-3203. doi: 10.1016/j.patcog.2007.06.003.
- [12] Abbas Cheddad. *Steganoflage : A New Image Steganography Algorithm*. PhD thesis, University of Ulster, 2009. URL <http://theses.eurasip.org/theses/443/steganoflage-a-new-image-steganography-algorithm/>.
- [13] Abbas Cheddad, Joan Condell, Kevin Curran, and Paul Mc Kevitt. A secure and improved self-embedding algorithm to combat digital document forgery. *Signal Processing*, 89:2324–2332, December 2009. ISSN 0165-1684. doi: 10.1016/j.sigpro.2009.02.001.
- [14] Brian Chen and Gregory W. Wornell. Quantization index modulation: a class of provably good methods for digital watermarking and information embedding. *IEEE Transactions on Information Theory*, 47(4):1423–1443, 2001. doi: 10.1109/18.923725.
- [15] Fan Chen, Hongjie He, Yaoran Huo, and Hongxia Wang. Self-recovery fragile watermarking scheme with variable watermark payload. In YunQing Shi, Hyoun-Joong Kim, and Fernando Perez-Gonzalez, editors, *Digital Forensics and Watermarking*, volume 7128 of *Lecture Notes in Computer Science*, pages 142–155. Springer Berlin Heidelberg, 2012. ISBN 978-3-642-32204-4. doi: 10.1007/978-3-642-32205-1_13.
- [16] Minghua Chen, Yun He, and R.L. Lagendijk. A fragile watermark error detection scheme for wireless video communications. *IEEE Transactions on Multimedia*, 7(2):201–211, 2005. doi: 10.1109/TMM.2005.843367.
- [17] Ingemar Cox, Matthew Miller, Jeffrey Bloom, Jessica Fridrich, and Ton Kalker. *Digital Watermarking and Steganography*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 2nd edition, 2007. ISBN 0123725852, 9780123725851.

-
- [18] Ingemar J. Cox, Gwenaél Doërr, and Teddy Furon. Watermarking is not cryptography. In YunQing Shi and Byeungwoo Jeon, editors, *Digital Watermarking*, volume 4283 of *Lecture Notes in Computer Science*, pages 1–15. Springer Berlin Heidelberg, 2006. ISBN 978-3-540-48825-5. doi: 10.1007/11922841_1.
 - [19] Chenwei Deng, Weisi Lin, Bu-Sung Lee, and Chiew Tong Lau. Robust image coding based upon compressive sensing. *IEEE Transactions on Multimedia*, 14(2):278–290, 2012. doi: 10.1109/TMM.2011.2181491.
 - [20] D.L. Donoho. Compressed sensing. *IEEE Transactions on Information Theory*, 52(4):1289–1306, 2006. doi: 10.1109/TIT.2006.871582.
 - [21] Joachim J. Eggers, Robert Bäuml, Roman Tzschoppe, and Bernd Girod. Scalar costa scheme for information embedding. *IEEE Transactions on Signal Processing*, 51(4):1003–1019, 2003. doi: 10.1109/TSP.2003.809366.
 - [22] Dominik Engel, Thomas Stütz, and Andreas Uhl. A survey on jpeg2000 encryption. *Multimedia Systems*, 15:243–270, 2009. ISSN 0942-4962. doi: 10.1007/s00530-008-0150-0.
 - [23] Jessica Fridrich. Robust hash functions for digital watermarking. In *Proc. of the International Conference on Information Technology: Coding and Computing*, pages 178–183, 2000. doi: 10.1109/ITCC.2000.844203.
 - [24] Jessica Fridrich and Miroslav Goljan. Images with self-correcting capabilities. In *Proc. of IEEE International Conference on Image Processing*, 1999. doi: 10.1109/ICIP.1999.817228.
 - [25] Gürkan Gür, Yücel Altug, Emin Anarim, and Fatih Alagöz. Image error concealment using watermarking with subbands for wireless channels. *IEEE Communications Letters*, 11(2):179–181, 2007. doi: 10.1109/LCOMM.2007.061055.
 - [26] Ammar M. Hassan, Yassin M. Y. Hasan, Ayoub Al-Hamadi, Mohamed A. A. Wahab, and Bernd Michaelis. A novel public key self-embedding fragile watermarking technique for image authentication. In *Proc. of IEEE International Conference on Image Processing*, pages 1253–1256, 2009. ISBN 978-1-4244-5653-6. doi: 10.1109/ICIP.2009.5413551.
 - [27] Hong-Jie He, Jia-Shu Zhang, and Heng-Ming Tai. Self-recovery fragile watermarking using block-neighborhood tampering characterization. In Stefan Katzenbeisser and Ahmad-Reza Sadeghi, editors, *Information Hiding*, volume 5806 of *Lecture Notes in Computer Science*, pages 132–145. Springer Berlin Heidelberg, 2009. ISBN 978-3-642-04430-4. doi: 10.1007/978-3-642-04431-1_10.

- [28] Hongjie He and Jiashu Zhang. Cryptanalysis on majority-voting based self-recovery watermarking scheme. In *Proc. International Conference on Multimedia Information Networking and Security*, 2009. doi: 10.1109/MINES.2009.218.
- [29] Hongjie He, Jiashu Zhang, and Hongxia Wang. Synchronous counterfeiting attacks on self-embedding watermarking schemes. *International Journal of Computer Science and Network Security*, 6(1B):251–257, 2006.
- [30] HongJie He, Fan Chen, Heng-Ming Tai, Ton Kalker, and Jiashu Zhang. Performance analysis of a block-neighborhood-based self-recovery fragile watermarking scheme. *IEEE Transactions on Information Forensics and Security*, 7(1):185–196, 2012. doi: 10.1109/TIFS.2011.2162950.
- [31] Matthew J. Holliman and Nasir D. Memon. Counterfeiting attacks on oblivious block-wise independent invisible watermarking schemes. *IEEE Transactions on Image Processing*, 9(3):432–441, 2000. doi: 10.1109/83.826780.
- [32] Wien Hong and Tung-Shou Chen. A novel data embedding method using adaptive pixel pair matching. *IEEE Transactions on Information Forensics and Security*, 7(1):176–184, 2012. doi: 10.1109/TIFS.2011.2155062.
- [33] Yaoran Huo, Hongjie He, and Fan Chen. Alterable-capacity fragile watermarking scheme with restoration capability. *Optics Communications*, 285: 1759–1766, 2012. doi: 10.1016/j.optcom.2011.12.044.
- [34] Paweł Korus and Andrzej Dziech. A novel approach to adaptive image authentication. In *Proc. of IEEE International Conference on Image Processing*, Brussels, 2011. doi: 10.1109/ICIP.2011.6116243.
- [35] Paweł Korus and Andrzej Dziech. Reconfigurable self-embedding with high quality restoration under extensive tampering. In *Proc. of IEEE International Conference on Image Processing*, Orlando, FL, 2012. doi: 10.1109/ICIP.2012.6467329.
- [36] Paweł Korus and Andrzej Dziech. Efficient method for content reconstruction with self-embedding. *IEEE Transactions on Image Processing*, 22(3):1134–1147, March 2013. doi: 10.1109/TIP.2012.2227769.
- [37] Paweł Korus, Wojciech Szmuc, and Andrzej Dziech. A scheme for censorship of sensitive image content with high-quality reconstruction ability. In *Proc. of IEEE International Conference on Multimedia and Expo*, Singapore, 2010. doi: 10.1109/ICME.2010.5583410.

-
- [38] Paweł Korus, Jarosław Białas, Piotr Olech, and Andrzej Dziech. A high-capacity annotation watermarking scheme. In *Multimedia Communications, Services and Security*, volume 149 of *Communications in Computer and Information Science*, pages 1–9. Springer Berlin Heidelberg, 2011. ISBN 978-3-642-21512-4. doi: 10.1007/978-3-642-21512-4_1.
- [39] Paweł Korus, Lucjan Janowski, and Piotr Romaniak. Automatic quality control of digital image content reconstruction schemes. In *Proc. of IEEE International Conference on Multimedia and Expo*, Barcelona, 2011. doi: 10.1109/ICME.2011.6011872.
- [40] Paweł Korus, Jarosław Białas, and Andrzej Dziech. A new approach to high-capacity annotation watermarking based on digital fountain codes. *Multimedia Tools and Applications*, pages 1–19, 2012. ISSN 1380-7501. doi: 10.1007/s11042-011-0986-8.
- [41] Koert Kuipers. Bnb - branch and bound solver for a mixed-integer nonlinear programming. <http://www.mathworks.com/matlabcentral/fileexchange/95-bnb>, 2003.
- [42] Jaejin Lee and Chee Sun Won. Authentication and correction of digital watermarking images. *Electronics Letters*, 35(11):886–887, 1999. doi: 10.1049/el:19990642.
- [43] Jaejin Lee and Chee Sun Won. Image integrity and correction using parities of error control coding. In *Proc. of IEEE International Conference on Multimedia and Expo*, volume 3, pages 1297–1300, 2000. doi: 10.1109/ICME.2000.871004.
- [44] Tien-You Lee and Shinfeng D. Lin. Dual watermark for image tampering detection and recovery. *Pattern Recognition*, 41:3497–3506, 2008. doi: 10.1016/j.patcog.2008.05.003.
- [45] Guangzhen Li, Yoshimichi Ito, Xiaoyi Yu, Naoko Nitta, and Noboru Babaguchi. A discrete wavelet transform based recoverable image processing for privacy protection. In *Proc. of IEEE International Conference on Image Processing*, pages 1372–1375, 2008. doi: 10.1109/ICIP.2008.4712019.
- [46] Guangzhen Li, Yoshimichi Ito, Xiaoyi Yu, Naoko Nitta, and Noboru Babaguchi. Recoverable privacy protection for video content distribution. *EURASIP Journal on Information Security*, 2009, January 2009. ISSN 1687-4161. doi: 10.1155/2009/293031.

- [47] Ching-Yung Lin and Shih-Fu Chang. Sari: Self-authentication-and-recovery image watermarking system. In *ACM International Conference on Multimedia*, pages 628–629, New York, NY, USA, 2001. ISBN 1-58113-394-4. doi: 10.1145/500141.500266.
- [48] Anna Lisowska. Smoothlets - multiscale functions for adaptive representation of images. *IEEE Transactions on Image Processing*, 20(7):1777–1787, 2011. doi: 10.1109/TIP.2011.2108662.
- [49] Stuart Lloyd. Least Squares Quantization in PCM. *IEEE Transactions on Information Theory*, 28(2):129 – 137, mar 1982. ISSN 0018-9448. doi: 10.1109/TIT.1982.1056489.
- [50] Michael Luby. Lt codes. In *Proceedings of the 43rd Symposium on Foundations of Computer Science*, FOCS '02, pages 271–280, Washington, DC, USA, 2002. IEEE Computer Society. ISBN 0-7695-1822-2. doi: 10.1109/SFCS.2002.1181950.
- [51] David J. MacKay. *Information Theory, Inference, and Learning Algorithms*. Cambridge University Press, 2003. ISBN 978-0521642989. URL <http://www.inference.phy.cam.ac.uk/mackay/itila/>.
- [52] David J. MacKay. Fountain codes. *IEE Proceedings Communication*, 152(6), 2005. doi: 10.1049/ip-com:20050237.
- [53] Antonis Mouhtaropoulos, Chang-Tsun Li, and Marthie Grobler. Proactive digital forensics: The ever-increasing need for standardization. In *Proc. of 2012 European Intelligence and Security Informatics Conference (EISIC)*, Odense, 2012. doi: 10.1109/EISIC.2012.66.
- [54] Zhenxing Qian and Guorui Feng. Inpainting assisted self recovery with decreased embedding data. *IEEE Signal Processing Letters*, 17(11):929–932, November 2010. ISSN 1070-9908. doi: 10.1109/LSP.2010.2072991.
- [55] Zhenxing Qian and Tong Qiao. Image self-embedding with large-area restoration capability. In *Proc. of IEEE International Multimedia Information Networking and Security Conference*, pages 649–652, 2010. doi: 10.1109/MINES.2010.141.
- [56] Zhenxing Qian, Guorui Feng, Xinpeng Zhang, and Shuozhong Wang. Image self-embedding with high-quality restoration capability. *Digital Signal Processing*, 21(2):278–286, March 2011. ISSN 1051-2004. doi: 10.1016/j.dsp.2010.04.006.

-
- [57] Chuan Qin, Chin-Chen Chang, and Pei-Yu Chen. Self-embedding fragile watermarking with restoration capability based on adaptive bit allocation mechanism. *Signal Processing*, 92(4):1137 – 1150, 2012. ISSN 0165-1684. doi: 10.1016/j.sigpro.2011.11.013.
- [58] Amir Said. Measuring the strength of partial encryption schemes. In *International Conference on Image Processing*, pages II: 1126–1129, 2005. doi: 10.1109/ICIP.2005.1530258.
- [59] Gerald Schaefer and Michal Stich. Ucid - an uncompressed colour image database. In *Storage and Retrieval Methods and Applications for Multimedia 2004*, Proceedings of SPIE, pages 472–480, 2004. URL <http://homepages.lboro.ac.uk/~cogs/datasets/UCID/ucid.html>.
- [60] Gerard Sierksma. *Linear and Integer Programming - Theory and Practice*. Marcel Dekker, 2-nd edition, 2002. ISBN 978-0824706739.
- [61] David Slepian and Jack Wolf. Noiseless coding of correlated information sources. *IEEE Transactions on Information Theory*, 19(4):471–480, July 1973. ISSN 0018-9448. doi: 10.1109/TIT.1973.1055037.
- [62] Jun Tian. Reversible data embedding using a difference expansion. *IEEE Transactions on Circuits and Systems for Video Technology*, 13(8):890–896, 2003. doi: 10.1109/TCSVT.2003.815962.
- [63] Hui Wang, Anthony T. S. Ho, and Xi Zhao. A novel fast self-restoration semi-fragile watermarking algorithm for image content authentication resistant to jpeg compression. In *Proceedings of the 10th international conference on Digital-Forensics and Watermarking, IWDW'11*, pages 72–85, Berlin, Heidelberg, 2012. Springer-Verlag. ISBN 978-3-642-32204-4. doi: 10.1007/978-3-642-32205-1_8.
- [64] Ming-Shi Wang and Wei-Che Chen. A majority-voting based watermarking scheme for color image tamper detection and recovery. *Computer Standards & Interfaces*, 29(5):561–570, 2007. ISSN 0920-5489. doi: 10.1016/j.csi.2006.11.009.
- [65] Shuenn-Shyang Wang and Sung-Lin Tsai. Automatic image authentication and recovery using fractal embedding and image inpainting. *Pattern Recognition*, 41:701–712, 2008. doi: 10.1016/j.patcog.2007.05.012.
- [66] Aaron D. Wyner and Jacob Ziv. The rate-distortion function for source coding with side information at the decoder. *IEEE Transactions on Information Theory*, 22:1–10, 1976. doi: 10.1109/TIT.1976.1055508.

- [67] Shao Yafei, Zhang Li, Wu Guowei, and Lin Xinggang. Reconstruction of missing blocks in image transmission by using self-embedding. In *Proc. Int Intelligent Multimedia, Video and Speech Processing Symp*, pages 535–538, 2001. doi: 10.1109/ISIMP.2001.925451.
- [68] Chun-Wei Yang and Jau-Ji Shen. Recover the tampered image based on vq indexing. *Signal Processing*, 90:331–243, 2010. doi: 10.1016/j.sigpro.2009.07.007.
- [69] Xiaoyi Yu, Kenta Chinomi, Takashi Koshimizu, Naoko Nitta, Yoshimichi Ito, and Noboru Babaguchi. Privacy protecting visual processing for secure video surveillance. In *Proc. of IEEE International Conference on Image Processing*, pages 1672–1675, 2008. doi: 10.1109/ICIP.2008.4712094.
- [70] Wei Zhang, S.S. Cheung, and Minghua Chen. Hiding privacy information in video surveillance system. In *Proc. of IEEE International Conference on Image Processing*, volume 3, 2005. doi: 10.1109/ICIP.2005.1530530.
- [71] Xinpeng Zhang and Shuozhong Wang. Fragile watermarking with error free restoration capability. *IEEE Transactions on Multimedia*, 10(8), 2008. doi: 10.1109/TMM.2008.2007334.
- [72] Xinpeng Zhang and Shuozhong Wang. Fragile watermarking scheme using a hierarchical mechanism. *Signal Processing*, 89:675–679, 2009. doi: 10.1016/j.sigpro.2008.10.001.
- [73] Xinpeng Zhang, Shuozhong Wang, and Guorui Feng. Fragile watermarking scheme with extensive content restoration capability. In *Proc. of International Workshop on Digital Watermarking*, 2009. doi: 10.1007/978-3-642-03688-0_24.
- [74] Xinpeng Zhang, Zhenxing Qian, Yanli Ren, and Guorui Feng. Watermarking with flexible self-recovery quality based on compressive sensing and composite reconstruction. *IEEE Transactions on Information Forensics and Security*, 6(4):1223–1232, 2011. doi: 10.1109/TIFS.2011.2159208.
- [75] Xinpeng Zhang, Shuozhong Wang, Zhenxing Qian, and Guorui Feng. Self-embedding watermark with flexible restoration quality. *Multimedia Tools and Applications*, 54:385–395, 2011. ISSN 1380-7501. doi: 10.1007/s11042-010-0541-z.
- [76] Xinpeng Zhang, Shuozhong Wang, Zhenxing Qian, and Guorui Feng. Reference sharing mechanism for watermark self-embedding. *IEEE Transactions on Image Processing*, 20(2):485–495, 2011. doi: 10.1109/TIP.2010.2066981.

- [77] Xunzhan Zhu, Anthony T.S. Ho, and Pina Marziliano. A new semi fragile image watermarking with robust tampering restoration using irregular sampling. *Signal Processing : Image Communication*, 22(5), 2007. doi: 10.1016/j.image.2007.03.004.