

Digital Watermarking Based Image and Video Quality Evaluation

Sha Wang

Thesis submitted to the
Faculty of Graduate and Postdoctoral Studies
in partial fulfillment of the requirements
for a doctoral degree in Electrical and Computer Engineering

Ottawa-Carleton Institute for Electrical and Computer Engineering

School of Electrical Engineering and Computer Science

University of Ottawa

© Sha Wang, Ottawa, Canada, 2013

Abstract

Image and video quality evaluation is very important. In applications involving signal transmission, the Reduced- or No-Reference quality metrics are generally more practical than the Full-Reference metrics. Digital watermarking based quality evaluation emerges as a potential Reduced- or No-Reference quality metric, which estimates signal quality by assessing the degradation of the embedded watermark. Since the watermark contains a small amount of information compared to the cover signal, performing accurate signal quality evaluation is a challenging task. Meanwhile, the watermarking process causes signal quality loss.

To address these problems, in this thesis, a framework for image and video quality evaluation is proposed based on semi-fragile and adaptive watermarking. In this framework, adaptive watermark embedding strength is assigned by examining the signal quality degradation characteristics. The “Ideal Mapping Curve” is experimentally generated to relate watermark degradation to signal degradation so that the watermark degradation can be used to estimate the quality of distorted signals.

With the proposed framework, a quantization based scheme is first implemented in DWT domain. In this scheme, the adaptive watermark embedding strengths are optimized by iteratively testing the image degradation characteristics under JPEG compression. This iterative process provides high accuracy for quality evaluation. However, it results in relatively high computational complexity.

As an improvement, a tree structure based scheme is proposed to assign adaptive watermark embedding strengths by pre-estimating the signal degradation characteristics, which greatly improves the computational efficiency. The SPIHT tree structure and HVS masking are used to guide the watermark embedding, which greatly reduces

the signal quality loss caused by watermark embedding. Experimental results show that the tree structure based scheme can evaluate image and video quality with high accuracy in terms of PSNR, wPSNR, JND, SSIM and VIF under JPEG compression, JPEG2000 compression, Gaussian low-pass filtering, Gaussian noise distortion, H.264 compression and packet loss related distortion.

Contents

Abstract	i
Contents	iii
List of Tables	ix
List of Figures	x
List of Acronyms	xvi
Acknowledgement	xviii
1 Introduction	1
1.1 Image and video quality evaluation	1
1.1.1 Widely used quality evaluation metrics	2
1.1.2 Classifications of the quality metrics	12
1.2 Digital watermarking	13
1.2.1 A typical watermarking system and watermark classifications . .	13
1.2.2 The cover signals	18
1.2.3 Main applications of digital watermarking	19
1.2.4 The three watermarking criteria	21

1.3	Digital watermarking based quality evaluation	22
1.4	Contributions of the research	23
1.5	Thesis structure	27
2	Literature review	28
2.1	Watermarking based image quality evaluation	30
2.1.1	Schemes with image-feature-independent watermarks	32
2.1.2	Schemes with image-feature-dependent watermarks	36
2.2	Watermarking based video quality evaluation	40
2.2.1	Schemes with video-feature-independent watermarks	42
2.2.2	Schemes with video-feature-dependent watermarks	45
2.3	Summary	46
3	The proposed framework for watermarking based quality evaluation	48
3.1	The proposed framework for watermarking based quality evaluation . .	48
3.2	Discrete Wavelet Transform (DWT)	52
3.3	The proposed method to evaluate the image degradation characteristics	58
3.4	“Ideal Mapping Curve” – the degradation relationship between the wa- termark and images	63
3.5	The proposed method to evaluate the effectiveness of the watermarking based quality evaluation schemes	67
3.6	Summary	69
4	Image quality evaluation using quantization based watermarking	70
4.1	The quantization based watermark embedding process	70
4.1.1	Quantization method used for watermark embedding	74

4.1.2	Empirical initialization of the watermark embedding strength . . .	76
4.1.3	Iterative optimization of the watermark embedding strength . . .	79
4.2	The quantization based watermark extraction and quality evaluation . .	86
4.2.1	The watermark extraction	87
4.2.2	The generation of “Ideal Mapping Curve”	88
4.2.3	The watermarking based quality evaluation	91
4.3	Experimental results and accuracy evaluation	92
4.3.1	Summary of experiments	92
4.3.2	The original images and the original watermark	94
4.3.3	Experimental results	95
4.3.4	Advantages of the proposed quantization based scheme	101
4.4	Summary	102

5	Image quality evaluation using SPIHT tree structure and HVS based watermarking	104
5.1	The motivation of using tree structure	104
5.2	The tree structure based watermark embedding process	107
5.3	The tree structure based watermark embedder	109
5.3.1	The formation of the tree structure	109
5.3.2	The selection of trees and DWT coefficients	111
5.3.3	The watermark embedding	114
5.3.4	Security of the watermarking process	115
5.4	The estimation of the adaptive watermark embedding strength	116
5.4.1	The empirical watermark bit assignment	116
5.4.2	The HVS and bitplane based perceptual masking	121

5.5	The tree structure based watermark extraction and quality evaluation	129
5.6	The generation of the “Ideal Mapping Curve”	131
5.7	Summary	135
6	Experimental results of tree structure based image quality evaluation	137
6.1	The original images and the original watermark	138
6.2	Summary of the experiments	139
6.2.1	Choosing the range for the distortion strength	141
6.2.2	The empirical grouping thresholds	144
6.3	Experimental results and accuracy evaluation	146
6.3.1	Quality evaluation in terms of PSNR under different distortions	148
6.3.2	Quality evaluation in terms of wPSNR under different distortions	152
6.3.3	Quality evaluation in terms of Watson JND under different dis- tortions	156
6.3.4	Quality evaluation in terms of SSIM under different distortions .	160
6.3.5	Quality evaluation in terms of VIF under different distortions .	163
6.3.6	Summary of the evaluation accuracy	167
6.4	Advantages of the proposed scheme	170
7	Video quality evaluation using SPIHT tree structure and HVS based watermarking	172
7.1	The density of watermark embedding	175
7.2	Summary of the experiments on video quality evaluation	178
7.3	Quality evaluation in terms of PSNR under H.264 compression	179
7.3.1	A preliminary experiment	180
7.3.2	Experimental results tested on I-frames of CIF videos	183

7.3.3	Experimental results tested on I-frames of 720p videos	185
7.3.4	Experimental results tested on I- and P-frames of CIF videos . .	187
7.3.5	Experimental results tested on I- and P-frames of 720p videos .	189
7.4	Quality evaluation in terms of PSNR under packet loss	190
7.4.1	Experimental results tested under packet loss	194
7.5	Summary	196
8	Conclusions and future work	197
	Appendix A	200
	Bibliography	211

List of Tables

4.1	Summary of the quality evaluation accuracy in MAE.	101
4.2	Summary of the quality evaluation accuracy in Pearson correlation coefficient.	101
6.1	The empirical grouping thresholds used for the 16 sets of experiments. .	146
6.2	Accuracy of the experimental results in terms of PSNR achieved under different distortions.	151
6.3	The average quality (dB) of the distorted images in terms of PSNR under interested distortion strengths.	151
6.4	Accuracy of the experimental results in terms of wPSNR achieved under different distortions.	155
6.5	The average quality (dB) of the distorted images in terms of wPSNR under interested distortion strengths.	155
6.6	Accuracy of the experimental results in terms of JND achieved under different distortions.	159
6.7	The average quality (JND) of the distorted images in terms of JND under interested distortion strengths.	160
6.8	Accuracy of the experimental results in terms of SSIM achieved under different distortions.	163

6.9	The average quality of the distorted images in terms of SSIM under interested distortion strengths.	163
6.10	Accuracy of the experimental results in terms of VIF achieved under different distortions.	167
6.11	The average quality of the distorted images in terms of VIF under interested distortion strengths.	167
6.12	Summary of the quality evaluation accuracy in MAE.	168
6.13	Summary of the quality evaluation accuracy in Pearson correlation coefficient.	168
6.14	Summary of the quality evaluation accuracy in RMSE	169
6.15	Accuracy comparison between the quantization based scheme and the tree structure based scheme.	170
7.1	Quality evaluation in terms of PSNR on I-frames of 512×512 videos . .	183
7.2	Quality evaluation in terms of PSNR on I-frames of CIF videos	184
7.3	Quality evaluation in terms of PSNR on I-frames of 720p and 640p videos	187
7.4	Quality evaluation in terms of PSNR on I- and P-frames of CIF videos	188
7.5	Quality evaluation in terms of PSNR on I- and P-frames of 720p and 640p videos	190
7.6	Quality evaluation in terms of PSNR under packet loss	195

List of Figures

1.1	A contrast sensitivity function.	5
1.2	A frequency sensitivity threshold map.	6
1.3	Illustration of the calculated SSIM indices.	9
1.4	Illustration of the distorted images.	11
1.5	A digital watermarking system.	14
1.6	Watermark classifications.	16
1.7	A simple example of embedding watermark in an image.	17
1.8	Trade-off among the three watermarking criteria.	21
3.1	Watermark degradation under JPEG compression.	49
3.2	The general framework for signal-feature-independent watermarking based quality evaluation applications.	51
3.3	A DWT analysis bank.	56
3.4	The DWT down-sampling.	57
3.5	The DWT decomposed blocks.	57
3.6	The original Lena and Baboon with their wavelet decomposed images. .	60
3.7	The degradation characteristics of image Lena.	62
3.8	The degradation characteristics of image Baboon.	63
3.9	The “Ideal Mapping Curve” generator.	65

4.1	The quantization based image watermark embedding process.	71
4.2	Numeration of the 10 DWT decomposed subbands.	71
4.3	The quantization method used for watermark embedding. $\frac{1}{\Delta}$ is the quantization parameter.	75
4.4	The degradation characteristics of the watermark achieved using the initial watermark embedding strength.	78
4.5	The TDR-PSNR curves tested using the initial watermark embedding strength.	79
4.6	The optimization of the quantization parameters.	80
4.7	The implementation details of the quality gradient calculator.	81
4.8	The quantization based watermark extraction and quality evaluation process.	86
4.9	Generation of the “Ideal Mapping Curve” using a fixed watermark embedding strength in the quantization based scheme.	89
4.10	The principle of the quantization parameter optimization.	90
4.11	An example of the quantization parameter optimization on multiple tested quality points.	91
4.12	Linear interpolation.	92
4.13	The default parameters of Dctune2.0.	93
4.14	The 32×32 original watermark.	94
4.15	Illustration of the effect of watermark embedding.	96
4.16	The “Ideal mapping curves” generated for image quality evaluation in terms of PSNR, wPSNR, Watson JND and SSIM under JPEG compression.	97

4.17	Image quality evaluation in terms of PSNR and wPSNR achieved using both the initial watermark embedding strength and the optimized watermark embedding strengths.	99
4.18	Experimental results of image quality evaluation in terms of Watson JND and SSIM achieved using both the initial watermark embedding strength and the optimized watermark embedding strengths.	100
5.1	Illustration of the bitplane decomposed DWT image.	106
5.2	The SPIHT tree structure and HVS based image watermark embedding process.	107
5.3	Formation of the tree structure.	110
5.4	Illustration of the formed trees.	111
5.5	Illustration of the tree selection.	113
5.6	The tree selection from the three orientations.	114
5.7	Illustration of the secure reordering of the selected trees in Fig. 5.6 (a).	115
5.8	Illustration of the image complexity evaluation.	119
5.9	Illustration of the generated HVS masks.	125
5.10	Illustration of the mapping thresholds selection.	128
5.11	The SPIHT tree structure and HVS based watermark extraction and quality evaluation process.	129
5.12	Linear interpolation.	131
5.13	Comparisons of the “Ideal Mapping Curve” generated using different number of test images.	132
5.14	Generation of the “Ideal Mapping Curve” using adaptive watermark embedding strengths in the tree structure based scheme.	134

6.1	The original watermark.	138
6.2	Illustration of the effects of watermark embedding.	140
6.3	Example of the image quality evaluation under Gaussian low-pass filtering.	143
6.4	A 3×3 Gaussian filter mask with standard deviation, $\sigma = 0.5$	147
6.5	The generated “Ideal Mapping Curves” in terms of PSNR under JPEG compression, JPEG2000 compression, Gaussian low-pass filtering and Gaussian noise pollution.	148
6.6	Experimental results of image quality evaluation in terms of PSNR under different distortions with distortions strengths $ds \in [ls, \mathbf{hs}']$	149
6.7	Experimental results of image quality evaluation in terms of PSNR under different distortions with distortions strengths $ds \in [ls, \mathbf{hs}]$	150
6.8	The generated “Ideal Mapping Curves” in terms of wPSNR under JPEG compression, JPEG2000 compression, Gaussian low-pass filtering and Gaussian noise pollution.	152
6.9	Experimental results of image quality evaluation in terms of wPSNR under different distortions with distortions strengths $ds \in [ls, \mathbf{hs}']$	153
6.10	Experimental results of image quality evaluation tested in terms of wPSNR under different distortions with distortions strengths $ds \in [ls, \mathbf{hs}]$	154
6.11	The generated “Ideal Mapping Curves” in terms of Watson JND under JPEG compression, JPEG2000 compression, Gaussian low-pass filtering and Gaussian noise pollution.	156
6.12	Experimental results of image quality evaluation in terms of Watson JND under different distortions with distortions strengths $ds \in [ls, \mathbf{hs}']$	157
6.13	Experimental results of image quality evaluation in terms of Watson JND under different distortions with distortions strengths $ds \in [ls, \mathbf{hs}]$	158

6.14	The generated “Ideal Mapping Curves” in terms of SSIM under JPEG compression, JPEG2000 compression, Gaussian low-pass filtering and Gaussian noise pollution.	161
6.15	Experimental results of image quality evaluation in terms of SSIM under different distortions with distortions strengths $ds \in [ls, \mathbf{hs}']$	162
6.16	The generated “Ideal Mapping Curves” in terms of VIF under JPEG compression, JPEG2000 compression, Gaussian low-pass filtering and Gaussian noise pollution.	164
6.17	Experimental results of image quality evaluation in terms of VIF under different distortions with distortions strengths $ds \in [ls, \mathbf{hs}']$	165
6.18	Experimental results of image quality evaluation in terms of VIF under different distortions with distortions strengths $ds \in [ls, \mathbf{hs}]$	166
7.1	The framework for watermarking based video quality evaluation.	173
7.2	The SPIHT tree structure and HVS based video watermark embedding process.	175
7.3	The “Ideal Mapping Curve” generated to evaluate the quality of 512×512 videos under H.264 compression.	181
7.4	Illustration of the framewise quality evaluation on the I-frames of two sample 512×512 videos.	182
7.5	The “Ideal Mapping Curve” generated to evaluate the quality of CIF videos under H.264 compression.	183
7.6	Illustration of the framewise quality evaluation on the I-frames of one sample CIF video.	184

7.7	The “Ideal Mapping Curve” generated to evaluate the quality of 720p or 640p videos under H.264 compression.	185
7.8	Illustration of the framewise quality evaluation on the I-frames of one sample 720p video.	186
7.9	Illustration of the framewise quality evaluation on the I- and P-frames of one sample CIF video.	188
7.10	Illustration of the framewise quality evaluation on the I- and P-frames of one sample 720p video.	189
7.11	The weighted averaging error concealment on the intra-blocks corrupted by packet loss.	193
7.12	The “Ideal Mapping Curve” under packet loss.	194
7.13	Experimental results tested on CIF videos under packet loss.	195
8.1	Strategy of choosing empirical grouping thresholds for the tree structure based scheme.	199
A-1	The 150 original images used in the experiments.	204
A-2	The 10 video frames used to generate “Ideal Mapping Curves”.	205
A-3	The 12 original 512×512 video sequences.	207
A-4	The 12 original CIF video sequences.	208
A-5	The 9 original 720p or 640p video sequences.	210

List of Acronyms

AWGN	Additive White Gaussian Noise
CIF	Common Intermediate Format
DCT	Discrete Cosine Transform
DFT	Discrete Fourier Transform
DWT	Discrete Wavelet Transform
FFT	Fast Fourier Transform
FMT	Fourier Mellin Transform
GOP	Group of Pictures
HAS	Human Auditory System
HVS	Human Visual System
IDWT	Inverse Discrete Wavelet Transform
ITU	International Telecommunication Union
JND	Just Noticeable Difference
JPEG	Joint Photographic Experts Group
MAE	Mean Absolute Error
MOS	Mean Opinion Score
MPEG	Moving Picture Experts Group
MSE	Mean Squared Error
PAL	Phase Alternating Line
PSNR	Peak Signal-to-Noise Ratio
QCIF	Quarter Common Intermediate Format
QoS	Quality of Service

RMSE	Root Mean Squared Error
SPIHT	Set Partitioning in Hierarchical Trees
SSIM	Structural SIMilarity
TDR	True Detection Rate
VIF	Visual Information Fidelity
VQM	Video Quality Metric
wPSNR	weighted Peak Signal-to-Noise Ratio

Acknowledgement

I would like to thank my supervisor, Professor Jiying Zhao, for bringing the problem of digital watermarking to me, and for his patient guidance and feedbacks during every step of my work.

Special thanks to Dr. Speranza, Dr. Tam from Communication Research Centre and Dr. Dansereau from Carleton University for their valuable advice for my research work.

I would also like to thank Dong, April, and Adam for their encouragement and positive support.

Chapter 1

Introduction

1.1 Image and video quality evaluation

Image or video quality quantitatively expresses the degradation characteristics of the image or video signal against distortions or artifacts. With the rapidly increasing use of digital equipments, internet and cellular networks, digital images and videos are compressed using higher and higher compression ratios to save storage space and transmission bandwidth, which makes the quality evaluations to be more and more important. For digital communications, quality is a valuable indicator of the Quality of Service (QoS) and an important measure of the conditions of transmission channels. In the meanwhile, quality can be used to evaluate the performances of the image and video processing algorithms, and the efficiency of the digital equipments, such as digital camera, scanner, fax, etc. Moreover, the high quality image and video contents and technology are making a good impact on the society culture and economy.

Although the image and video contents are eventually evaluated by humans, it is impossible to let human observers to do all the quality evaluation work. In the

past few decades, lots of efforts has been carried out to develop the quality metrics to make the evaluated quality meaningful. Because image is much smaller and less complex than video signals and is more suitable to be used to develop complex quality metrics, numerous quality metrics are proposed in the literature to evaluate image quality. The commonly used quality metrics for video signals are the Mean Squared Error (MSE) and Peak-Signal-to-Noise Ratio (PSNR). In the following sections, several widely used or popular quality metrics, such as Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Peak-Signal-to-Noise Ratio (PSNR), weighted Peak Signal-to-Noise Ratio (wPSNR), Just Noticeable Difference (JND), Structural SIMilarity (SSIM), and Visual Information Fidelity (VIF) are presented.

1.1.1 Widely used quality evaluation metrics

In this section, the quality metrics MSE, RMSE, PSNR, wPSNR, JND, and SSIM are presented. These metrics conduct quality evaluation by comparing the distorted signal and the original signal.

1.1.1.1 MSE and RMSE

With the MSE quality metric, it is assumed that the original signal is a perfect signal and any difference appears in another signal comparing to the original signal is treated as error. The MSE measures the average of the squares of errors.

MSE is defined as:

$$MSE = \frac{1}{MN} \sum_{i=1}^M \sum_{j=1}^N \left(I(i, j) - \hat{I}(i, j) \right)^2 \quad (1.1)$$

The RMSE is defined as the square root of MSE:

$$RMSE = \sqrt{\frac{1}{MN} \sum_{i=1}^M \sum_{j=1}^N \left(I(i, j) - \hat{I}(i, j) \right)^2} \quad (1.2)$$

where, I and $\hat{I}$ respectively are the original image and distorted image; (i, j) are the coordinates of the current pixel in the image; M and N are respectively the numbers of rows and columns of the image; MN is the total number of pixels in the image. In literature, sometimes, MSE and RMSE are used to evaluate video quality. In this case, I and $\hat{I}$ are the original video frame and the distorted video frame.

1.1.1.2 PSNR

PSNR is a MSE based quality metric and performs quality evaluation by comparing the pixelwise differences between the distorted image or video frames and the original image or video frames. Till now, PSNR is one of the most widely used quality evaluation metrics.

PSNR is defined as:

$$PSNR = 20 \log_{10} \frac{MAX}{\sqrt{\frac{1}{MN} \sum_{i=1}^M \sum_{j=1}^N \left(I(i, j) - \hat{I}(i, j) \right)^2}} \quad (1.3)$$

where, MAX is the maximum possible pixel value of an image or video frame. For an 8 bits/pixel grey-scale image, MAX is equal to 255. For a normalized image or video frame, MAX equals to 1.

1.1.1.3 wPSNR

The wPSNR is developed based on the PSNR by taking one subjective perception factor into account. It employs a contrast sensitivity function to weight the pixelwise difference

between the distorted image and the original image. As we know that human eyes are more sensitive to contrast than absolute luminance in an image, wPSNR performs quality evaluation closer to human perception compared to PSNR. The wPSNR is mathematically expressed using Equ. (1.4).

$$wPSNR = 20 \log_{10} \frac{MAX}{\sqrt{\frac{1}{MN} \sum_{i=1}^M \sum_{j=1}^N \left(CSF(f, \theta) \otimes \left(I(i, j) - \hat{I}(i, j) \right)^2 \right)}} \quad (1.4)$$

where, $CSF(\bullet)$ is the contrast sensitivity function; f is the spatial frequency and defines the relationship between the horizontal spatial frequency (u) and vertical spatial frequency (v) which can be mathematically expressed as $f = \sqrt{u^2 + v^2}$; f , u and v are in cycles/degree; $\theta = \tan^{-1}(v/u)$ is the angle in degree with respect to the horizontal axis; $\otimes$ indicates the filtering operation.

In paper [1], Makoto Miyahara et al. proposed to model the spatial frequency response of the contrast sensitivity function, $CSF(f, \theta)$, as follows:

$$\begin{aligned} CSF(f, \theta) &= CSF(f) \cdot O(f, \theta) \\ &= \left(1.5e^{-\frac{\pi^2}{450}f^2} - e^{-\frac{2\pi^2}{225}f^2} \right) \cdot \frac{1 + e^{-\frac{4\pi}{15}(f-f_o)} \cos^4 2\theta}{1 + e^{-\frac{4\pi}{15}(f-f_o)}} \end{aligned} \quad (1.5)$$

where, $CSF(f)$ is a contrast sensitivity function giving more weight to the low frequency coefficients; $O(f, \theta)$ performs the additional attenuation for the coefficients with spatial frequency $f > f_o$, $f_o = 11.13$ cycles/degree is an empirical parameter.

The contrast sensitivity function generated using Equ. (1.5) is illustrated in Fig. 1.1.

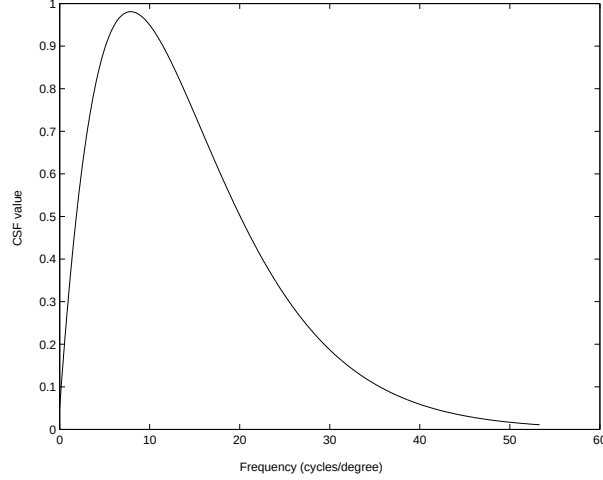


Figure 1.1: A contrast sensitivity function.

1.1.1.4 Watson JND

The Watson model [2] assesses the image visual quality by estimating the multiples of Just Noticeable Differences (JNDs) between the original signal and the distorted signal. According to the mechanism of the human visual system, three factors are considered in the Watson model in order to more closely approximate the perceptual quality of an image/video signal. And the three factors are: a frequency sensitivity function, a luminance masking function and a contrast masking function. Meanwhile, a pooling mechanism is used to combine all the estimated local perceptual distances together to achieve a global perceptual distance. The Watson model is designed in the Discrete Cosine Transform (DCT) domain to accommodate the JPEG compression process.

1. The frequency sensitivity function

The frequency sensitivity function estimates the human visual thresholds for one JND's distortion. It is derived in [3] utilizing a number of parameters, such as the distance between the viewer and the image and resolution of the image. For

one 8×8 DCT block, one threshold is estimated for one coefficient. That means if the current coefficient is changed by the amount of its threshold, the distortion caused by this change may reach one JND. And an example set of thresholds for a DCT block can be illustrated using Fig. 1.2. In the figure, x -axis and y -axis respectively indicate the coefficients' horizontal and vertical indices in a block. The z -axis represents the threshold values of the coefficients. The coordinate, $(0, 0)$, in the figure corresponds to the DC component of the current DCT block. The lower thresholds in the sensitivity map indicate that the human eyes are more sensitive to the changes on these frequency components.

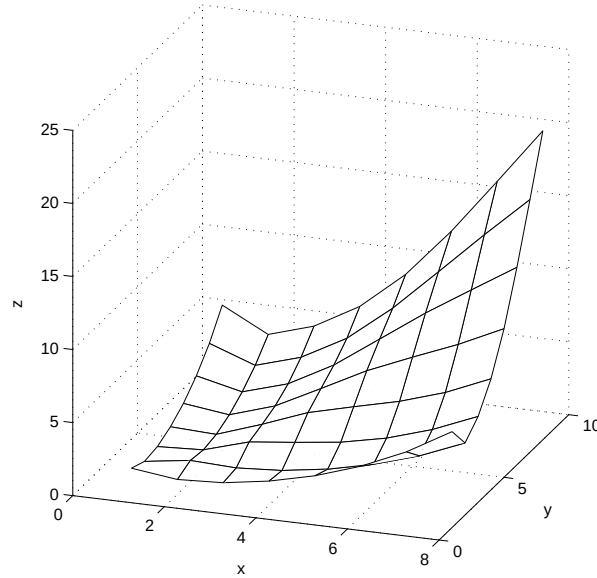


Figure 1.2: A frequency sensitivity threshold map.

2. The luminance masking function

The luminance masking function estimates the luminance masking thresholds for the block DCT coefficients. For a DCT block, if the intensity of its DC component is brighter, the block will be much more tolerant to alterations. According the

Watson model, the luminance masking thresholds for a DCT block can be adjusted using Equ. (1.6).

$$t_L(i, j, k) = t(i, j)(C_o(0, 0, k)/C_{0,0})^{0.649} \quad (1.6)$$

where, t , $C_o(0, 0, k)$, i and j are respectively the frequency sensitivity threshold map, the DC component, the row and column of a coefficient of the k^{th} DCT block; $C_{0,0}$ is the average value of the DC components in the image. The constant 0.649 is a suggested value that bridges the frequency sensitivity thresholds to the luminance masking thresholds.

3. The contrast masking function

For the same DCT block, the contrast masking thresholds can be achieved using Equ. (1.7).

$$s(i, j, k) = \max\{t_L(i, j, k), |C_o(i, j, k)|^{0.7} t_L(i, j, k)^{0.3}\} \quad (1.7)$$

The contrast masking thresholds are derived utilizing the frequency sensitivity thresholds and the luminance masking thresholds. For one DCT coefficient, it is a comprehensive threshold that indicates how much alteration that can be applied to the current frequency component while not being notified.

The Watson distance can be obtained by pooling all the local perceptual distances together using Equ. (1.9).

$$d(i, j, k) = \frac{e(i, j, k)}{s(i, j, k)} \quad (1.8)$$

$$D_{wat}(c_o, c_w) = \left(\sum_{i,j,k} |d(i, j, k)|^4 \right)^{\frac{1}{4}} \quad (1.9)$$

where, d is the local perceptual distance; e is the differences between the original image and the distorted image in block DCT domain; $D_{wat}(c_o, c_w)$ is the estimated Watson distance expressed as multiples of JNDs between the original image and the distorted image.

1.1.1.5 SSIM

Compared to the quality metrics presented in the previous few subsections, a more sophisticated quality metric, the Structural SIMilarity (SSIM), is proposed in [4] with the assumption that the perceived image distortions can be well approximated by a measure of the structural information changes. The wPSNR and Watson JND quality metrics evaluate the perceived image quality by empirically scaling the pixelwise differences between the original image and the distorted image. The SSIM evaluates the perceived image quality by evaluating the structural changes between the original image and the distorted image.

Considering that, with a typical viewing distance, the human eyes can perceive one local area of an image with high resolution instantaneously, the SSIM quality metric is proposed to apply on image locally within a 8×8 window. The local SSIM index is mathematically calculated by evaluating the luminance, contrast and structural information of the corresponding local regions in the original image and the distorted image using Equ. (1.10). In the equation, $l(\cdot)$, $c(\cdot)$, $s(\cdot)$ and $f(\cdot)$ respectively are the luminance function, contrast function, structural function and the pooling function; x and y are the corresponding local regions of the original image and the distorted im-

age, respectively; C_1 and C_2 are two empirical parameters introduced to guarantee the stability of the SSIM metric when the means and variances of x and y are approaching zero; L denotes the dynamic range of the pixel values of the cover image; k_1 and k_2 are set up as 0.01 and 0.02 by default.

$$\begin{aligned}
 SSIM(x, y) &= f(l(x, y), c(x, y), s(x, y)) \\
 &= \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)} \\
 &= \frac{(2\mu_x\mu_y + (k_1L)^2)(2\sigma_{xy} + (k_2L)^2)}{(\mu_x^2 + \mu_y^2 + (k_1L)^2)(\sigma_x^2 + \sigma_y^2 + (k_2L)^2)}
 \end{aligned} \tag{1.10}$$

In the local SSIM quality metric, the luminance, contrast and structural functions are assumed independent to each other. The SSIM index map calculated on image Barbara is illustrated in Fig. 1.3 (b).



(a) The original image Barbara.



(b) The SSIM index map of image Barbara (SSIM window size 8×8).

Figure 1.3: Illustration of the calculated SSIM indices.

After evaluating the structural changes locally, the degraded quality of a distorted image is assessed by averaging the calculated local SSIM indices:

$$MSSIM(\hat{I}, I) = \frac{1}{N} \sum_{i=1}^N SSIM(x_i, y_i) \quad (1.11)$$

where, $MSSIM$ is the evaluated quality in SSIM between the original image I and the distorted image $\hat{I}$; x_i is the current evaluating local region in the original image; y_i is the region corresponding to x_i in the distorted image; N is the total number of local regions in the image. The calculated MSSIM index locates in $[0, 1]$.

1.1.1.6 VIF

The Visual Information Fidelity (VIF) is a quality metric for natural images which provides quality scores with higher correlations to the subjective quality comparing to the wPSNR, Watson JND and SSIM as addressed in [5][6]. The VIF conducts image quality assessment in the DWT domain by evaluating the similarity of the image visual information extracted from the reference image and the distorted image.

In the VIF algorithm, to quantify the image visual information, both the reference image and the distorted image are modeled using random fields. In the DWT domain, the reference image, C , is modeled using the Gaussian scalar mixture random field, $\mathcal{C} = \mathcal{S} \cdot \mathcal{U}$, where $\mathcal{S}$ is a positive scalar random field and $\mathcal{U}$ is a Gaussian random field. The distorted image, D , is modeled as $\mathcal{D} = \mathcal{G}\mathcal{C} + \mathcal{V}$, where, $\mathcal{G}$ is a scalar random field and causes image blur, $\mathcal{V}$ is a Gaussian random field and is used to model the white Gaussian noise. By combining the image blur and additive Gaussian noise, the stochastic distortion model $\mathcal{G}\mathcal{C} + \mathcal{V}$ is used to approximate and quantify the perceptual annoyance caused by any possible type of image distortion, such as JPEG, JPEG2000, Gaussian noise addition and image blur. An example is illustrated in Fig. 1.4. Fig. 1.4 (a) is a distorted image compressed by the JPEG2000 compression. Fig. 1.4 (b) is an

image synthesized using the reference image and the stochastic distortion model, which has similar perceptual annoyance as Fig. 1.4 (a).



Figure 1.4: Illustration of the distorted images.

The VIF measure is defined as a ratio of the distorted image visual information over the reference image visual information:

$$VIF = \frac{I(\mathcal{C}; \mathcal{F})}{I(\mathcal{C}; \mathcal{E})} \quad (1.12)$$

where, $\mathcal{E}$ is the perceived reference image filtered by the HVS mask and is modeled as $\mathcal{E} = \mathcal{C} + \mathcal{N}$. Correspondingly, $\mathcal{F}$ is the perceived distorted image and is modeled as $\mathcal{F} = \mathcal{D} + \mathcal{N}'$. $\mathcal{N}$ and $\mathcal{N}'$ are two zero-mean Gaussian random fields modeling the loss of the image information caused by the HVS masking. $I(\mathcal{C}; \mathcal{F})$ is the mutual information of the perceived distorted image and the reference image quantifying the image visual information extracted from the distorted image. $I(\mathcal{C}; \mathcal{E})$ is the mutual information of the perceived reference image and the reference image quantifying the image visual information extracted from the reference image.

For distorted images, the VIF metric provides quality ratings varying in the range of $[0,1]$. For the contrast enhanced images, the VIF quality ratings will be larger than 1 indicating that the enhanced images have superior quality comparing to the

corresponding reference images.

1.1.2 Classifications of the quality metrics

Based on the dependence on a reference image, the quality evaluation metrics can be divided into three categories: the Full-Reference quality metrics, the Reduced-Reference quality metrics, and the No-reference quality metrics [7]. All the quality metrics introduced in Section 1.1.1 are Full-reference objective quality metrics.

The Full-Reference quality metrics evaluate signal quality by comparing the differences between the distorted signal and the original signal. A requirement to use the Full-Reference quality metrics is that the original signal must be present when the quality of the distorted signal is assessed. The Reduced-Reference quality metrics work requiring partial information or reduced reference features of the original reference signal. The reduced reference features can be rough (low data rate) or fine (high data rate) and can directly affect the accuracy of the quality evaluation. The rough reference features can be easily transmitted to the receiver side and results in a less accurate quality evaluation. With fine reference features, the quality evaluation will be much more accurate. However, the fine reference features will cause a heavy burden on transmission. The third type of quality metrics is the No-Reference quality metrics. In the past decade, lots of efforts have been put in research on the No-Reference quality metrics. These quality metrics need lots of research work in the metric development stage to achieve thorough knowledge of the characteristics of the natural signal such as how they react to different distortions.

Among the three categories, the Full-Reference metrics provide the most accurate quality evaluation results. However, their dependence on the original signal results in big inconvenience in certain applications when the original signal is not available. To

this end, the digital watermarking-based quality metrics have emerged as an appealing approach for Reduced- or No-reference quality metrics [8][9].

1.2 Digital watermarking

The rapid development of new information technologies has improved the ease of access to digital information. It also leads to the problem of illegal copying and redistribution of digital media. The concept of digital watermarking came up while trying to solve the problems related to the management of intellectual property of media. Digital watermarking is a type of technique to imperceptibly insert additional information into signals through slight modification of the data. For different purposes, the additional information, which is also called the digital watermark, can be embedded into the signals robustly or fragilely and will undergo the same distortions as the host signals. The retrieved digital watermark serves multiple purposes, such as the proof of the ownership and copyright and the indication of content tampering. In recent years, the watermarking based quality evaluation becomes a more and more promising research topic because of the convenience it provides to evaluate image or video quality without the access of the original signal.

1.2.1 A typical watermarking system and watermark classifications

As shown in Fig. 1.5, a typical watermarking system essentially consists of a watermark embedding process and a watermark extraction process.

The key components of Fig. 1.5 are denoted as follows: a digital watermark W , the host signal C , a security key K , the watermarked signal C_w , the distorted watermarked

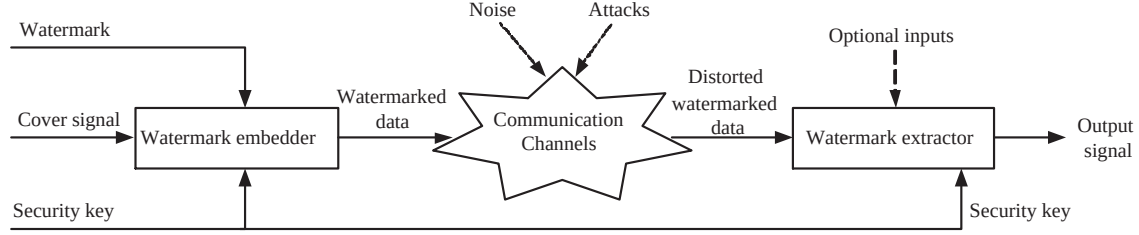


Figure 1.5: A digital watermarking system.

signal C'_w , the output signal O , the watermark embedder $e(\cdot)$ and the watermark extractor $d(\cdot)$. Thus, the watermark embedding process can be defined as:

$$C_w = e(C, W, K) \quad (1.13)$$

The watermark extraction process can be expressed as:

$$O = d(C'_w, K, \dots) \quad (1.14)$$

In the watermark embedding process, the inputs are the host signal, a watermark and a user-defined security key. The host signal can be an image, video, audio or other media expressions. To serve different purposes, the watermark can be a grey-scale image, a binary pattern or a number sequence, meaningful or random. The security key enhances the security of the whole watermarking system and guarantees the authorized access to the watermarking information. The output of the watermark embedding process is the watermarked signal.

The inputs to the watermark extraction process are the distorted watermarked signal and the security key. The optional input is the original watermark. As discussed in [2], the watermark extraction process has two functionalities: first, it detects or extracts the embedded watermark from the distorted watermarked signal; second, it does the

bitwise evaluations of the extracted watermark by referring to the original watermark to achieve a meaningful measurement. In this case, the output of the watermark extraction process can be the detected existence of the watermark, the extracted watermark or a desired measurement.

For a typical watermarking system, several requirements should be satisfied:

1. The watermark can be extracted from C_w or C'_w with or without explicit knowledge of C .
2. C_w should be as close to C as possible.
3. If C_w is unmodified, the extracted watermark W' , should exactly matches the original watermark, W .
4. By comparing to the original watermark, the extracted watermark, W' , should be able to reflect a meaningful measurement to serve the purpose of watermark embedding.

According to the way a watermark is embedded, it can be generally categorized as the visible watermark and the invisible watermark. The classifications of the watermark is illustrated in Fig. 1.6.

The visible watermark plays a less important role than the invisible watermark. It is partially transparently placed over the content of the cover work and can be hardly removed without damaging the cover work.

To work without significantly affecting the quality of the cover signal, the digital watermark is usually designed to be embedded completely invisible. According to the strength that a watermark is embedded, the invisible watermark can be further categorized into three groups: the robust watermark, semi-fragile watermark and fragile

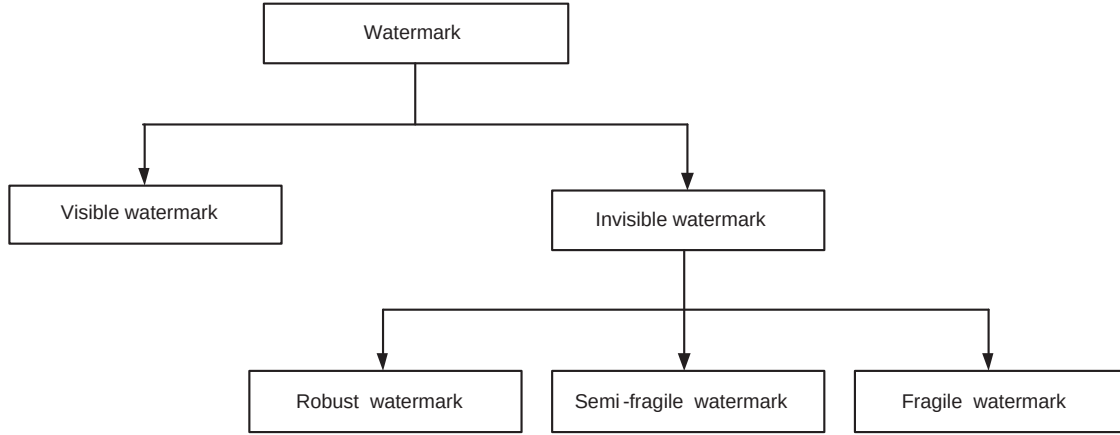


Figure 1.6: Watermark classifications.

watermark. For robust watermarking, the watermark is usually embedded with strong embedding strength and can survive a vast amount of the intentional or unintentional attacks. For fragile watermarking, the watermark is embedded with a very weak embedding strength so that the embedded watermark will be damaged even the watermarked image is very lightly distorted. The fragile watermark should be able to indicate the possible location that the watermarked image is tampered. The semi-fragile watermarking is between the robust watermarking and the fragile watermarking. It is designed to survive the legitimate distortions to the cover signals but be destroyed by illegitimate distortions [10]. The extracted semi-fragile watermark is expected to be able to reflect how much the watermarked image is degraded.

In a watermarking scheme, the watermark can be embedded in the spatial domain or in the frequency domain of the cover signal. Implementing the watermarking process in the spatial domain is simpler than in the frequency domain. However, the embedded watermark is easier to be removed using the method like pixelwise forgery attack. More advanced watermarking techniques are to embed watermark in the frequency domain of the cover signal, such as the Discrete Fourier Transform (DFT) domain [11][12][13],

the Discrete Cosine Transform (DCT) domain [14][15] and the Discrete Wavelet Transform (DWT) domain [16][17]. Furthermore, the watermark can be embedded with a fixed watermark embedding strength or with adaptive watermark embedding strength. The adaptive embedding strength usually obeys some desirable characteristics, such as Human Visual System (HVS) or Human Auditory System (HAS)[18][19], or image degradation characteristics under distortions [20][21]. The watermarking schemes working with adaptive watermark embedding strength are more sophisticated than those working with fixed watermark embedding strength. The watermarking schemes with considerations of HVS or HAS cause less quality degradation to the cover signal.

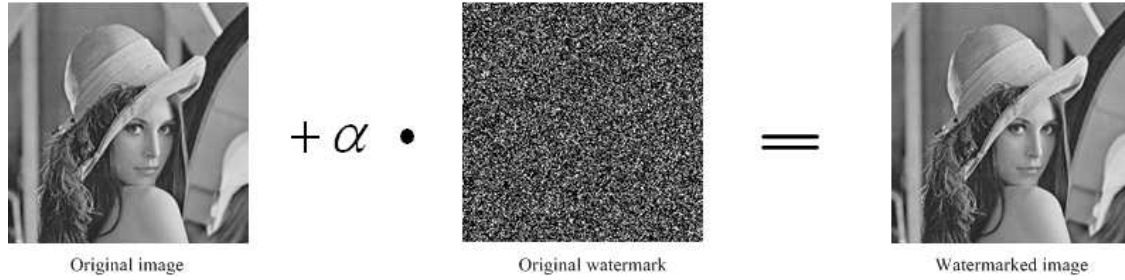


Figure 1.7: A simple example of embedding watermark in an image.

A simple example of image watermark embedding in spatial domain is illustrated in Fig. 1.7. The original watermark is normal distributed and is of the same size as the image. The watermark is linearly embedded into the original image. The constant α is a fixed watermark embedding strength and controls the vulnerability of the embedded watermark. The bigger the α , the stronger the watermark embedding strength and the more robust the embedded watermark, and vice versa. In this example, α equals to 1. The quality of the watermarked image is 48.59 dB in PSNR. In the watermarking based applications, when the quality of the watermarked image is higher than 45 dB in PSNR, the quality loss caused by the watermarking process is insignificant.

1.2.2 The cover signals

According to the cover signals that the watermark is embedded into, the digital watermarking applications can be divided into different categories.

1. Digital audio watermarking. The audio watermarking means that the watermark is embedded into the audio signal inaudibly and is transmitted along with the cover audio signal. Audio watermarking is based on the psycho-acoustical approach of perceptual audio coding techniques [22]. It exploits the properties of the human ear by embedding one or more key-dependent watermark signals below the audible masking threshold.
2. Digital image watermarking. In literature, most of the researches about digital watermarking are on image watermarking because that it is simpler to develop and analyze the watermarking schemes on images than on videos. In the meanwhile, there are so many images available on the internet free of charge and without any copyright protection.
3. Digital video watermarking [23][24][25]. Raw video signal consists of a sequence of still images. Theoretically, all the image watermarking schemes can be directly applied to the video signals. However, because video signal is usually compressed into IBP frames, further development is needed to make the image watermarking schemes efficient for the BP video frames. For security purpose, different security keys can be used for different video frames or different shots. In the meanwhile, the video watermark should be able to resist different types of attacks such as frame averaging, frame dropping, and frame swapping.
4. 3D virtual objects watermarking. The most important component for watermark

embedding in both VRML (Virtual Reality Modelling Language) and MPEG-4 is the 3D polygonal mesh [26][27][28]. The shape of a 3D polygonal mesh is defined by two components, vertex coordinate and vertex topology. Vertex coordinates combined with vertex topology define more complex geometrical primitives such as lines and polygons. These components are the most important targets for embedding in 3D mesh polygonal meshes.

5. Others, such as text watermarking [29], software watermarking [30], and database watermarking [31].

In this thesis, most of the discussions will be focused on digital image and video watermarking.

1.2.3 Main applications of digital watermarking

The main applications of the digital watermarking are listed as follows:

1. Copyright protection.

The problem of copyright protection comes with the fast-growing illegitimate redistributions of images and videos. In this application, the watermark carries the owner's copyright information and is robustly embedded into the image or video signal invisibly so that the watermark can still be extracted even when the watermarked signal is severely damaged [32][33][34].

2. Proof of ownership.

A further step from copyright protection for the watermarking technique is to prove the ownership. The watermark should be unambiguous. The ambiguity attack is to embed additional watermark to claim to be the owner. The watermark

used as a proof of ownership should be able to resolve the problem of the rightful ownership [35][36].

3. Content authentication.

The purpose of this application is to tell whether the content has been changed and the fragile watermarking contributes the most to this application [37][38]. The watermark is used to monitor even the smallest alteration to its cover signal. It is embedded with the weakest embedding strength and is expected to become undetectable if the cover signal is modified.

4. Quality evaluation.

To evaluate signal quality, a semi-fragile watermark is usually embedded into the cover signal. The embedded watermark is expected to survive the legal signal operations and be damaged after the illegal operations. After certain distortion, the quality of the degraded signal is evaluated by assessing the degradation of the watermark.

5. Broadcast monitoring.

In this application, the watermark is embedded into the cover signals in order to help the advertisers to identify when and where their advertisements are broadcasted by recognizing embedded watermark [39].

6. Copy control.

The watermark exists to prevent people from illegally copying the copyrighted content [40].

1.2.4 The three watermarking criteria

For an invisible watermarking system, three criteria are usually used to evaluate its performance:

1. Invisibility: Invisibility means that the watermark is expected to be embedded into the host signal with insignificant impact on the quality of the signal. Therefore, compared to the original signal, the higher the quality of the watermarked signal, the better.

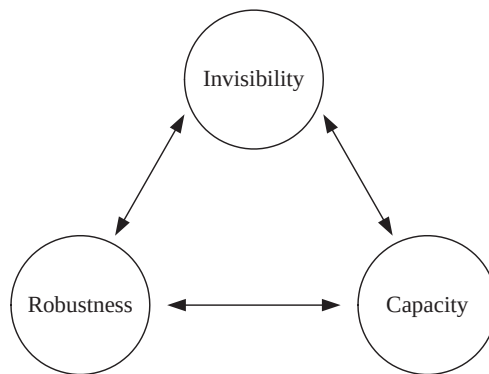


Figure 1.8: Trade-off among the three watermarking criteria.

2. Robustness: For different applications, the watermark is expected to be embedded with appropriate robustness.
3. Capacity: The watermarking capacity indicates the amount of watermark bits that can be embedded into the host signal and that can be reliably extracted afterwards. The higher the watermarking capacity is, the more potential the watermarking scheme has.

As illustrated in Fig. 1.8, a good trade-off among the three criteria is desired in the design of a digital watermarking system.

1.3 Digital watermarking based quality evaluation

To the contrary of the Full-Reference quality metrics, the digital watermarking based quality evaluation schemes emerge as a feasible approach to develop the Reduced- or No-reference quality metrics. In these schemes, usually a semi-fragile watermark is embedded in the cover signal and the embedded watermark is used to track the degradation of the watermarked signal under different distortions. After distortion, the degraded watermark is extracted and the degraded quality of the distorted signal is evaluated by assessing the degradation of the watermark compared to the original watermark. Thus, the original watermark needs to be either transmitted to or known by the receiver side. Compared to the cover signal, the watermark is usually very small in size. If the watermark needs to be transmitted, it will not cause significant cost to the transmission. This makes the watermarking based quality metrics more practical for the quality evaluation applications.

With the watermarking based quality metrics, the signal quality is normally estimated in terms of an objective quality metric, such as PSNR or JND, or a subjective quality metric, such as MOS. Due to the small size of the watermark, the most challenging task of the watermarking based quality metric is to accurately evaluate signal quality using the embedded watermark. Therefore, the accuracy of the quality evaluation is one criteria to assess the effectiveness of the watermarking based quality metric. In this thesis, “quality evaluation accuracy” assesses how close the quality scores provided by the proposed watermarking based quality metric to those of the reference quality metrics, such as PSNR, wPSNR, JND and SSIM. The higher the quality estimation accuracy, the better the performance of the watermarking based quality metric, and vice versa. The phrase “watermark embedding” indicates the data alterations in

the cover signal, which will directly introduce quality loss to the cover signal. Thus, the quality loss of the cover signal caused by the watermark embedding is also used to evaluate the performance of the watermarking based quality metric. Meanwhile, the watermarking based quality metrics should provide quality evaluation highly correlated with a wide variety of existing evaluation metrics. For example, the watermarking based quality metrics assess signal quality in terms of one Full-Reference quality metric, such as PSNR or SSIM. It is desirable if the watermarking based quality metrics can evaluate signal quality in terms of any interested quality metrics. In this case, the watermarking based quality metrics are expected to be as more flexible as possible. Because the watermarking based quality metrics are used to evaluate the quality of audio, image, video and 3D signal, etc. The high computational efficiency is more ideal for quality evaluations of the huge-sized signals or real time applications.

Based on the presentation made above, four criteria summarized in the following can be used to assess the performance of the watermarking based quality evaluation schemes:

1. The accuracy of the quality evaluation.
2. The quality loss caused by the watermark embedding process.
3. The flexibility of the watermarking based quality evaluation.
4. The computational efficiency.

1.4 Contributions of the research

In this thesis, research on the digital watermarking based quality evaluation is presented. The main contributions of the research can be summarized as follows:

1. In this thesis, a general digital watermarking based quality evaluation framework is proposed. This framework can be used to evaluate the quality of different signals (such as image, video and audio) by assessing the degradation of the watermark.
2. The concept of “Ideal Mapping Curve” has been introduced for the first time in literature. It quantifies the degradation relationship between the watermark and the cover signal. It makes the watermarking based quality evaluation of a distorted signal possible.
3. This thesis addresses two ways to estimate adaptive watermark embedding strength by evaluating the quality degradation characteristics of the cover signal under distortions, which are the quantization based scheme and the SPIHT tree structure and HVS based scheme. With the adaptive watermark embedding strength, the accuracy of the quality evaluation is greatly increased.
4. In literature, the quantization based algorithm is the first to assign adaptive watermark embedding strength by employing an iterative optimization process. It laid a solid basis for the further research on watermarking based quality evaluation.
5. The proposed quality evaluation scheme using SPIHT tree structure and HVS based watermarking is first introduced in literature. The watermark is embedded in selected bitplanes of selected SPIHT trees in the DWT domain. The HVS mask calculated using the DWT coefficients is experimentally mapped to HVS bitplane mask, which indicates the highest bitplane that can be used for watermark embedding without introducing perceptual change. The tree structure scheme provides high computational efficiency and insignificant quality degradation caused by the

watermark embedding.

6. The proposed framework makes quality evaluation flexible. With the “Ideal Mapping Curve”, the signal quality can be evaluated in terms of any interested Full-Reference quality metrics, such as PSNR, wPSNR, Watson JND and SSIM. The SPIHT tree structure and HVS based quality evaluation scheme has great potential to be further developed for other quality metrics.

Journal articles generated from the research:

1. Sha Wang, Dong Zheng, Jiying Zhao, Wa James Tam, and Filippo Speranza, “Adaptive watermarking and tree structure based image quality estimation”, submitted to IEEE Transactions on Multimedia, under 2nd review.
2. Sha Wang, Dong Zheng, Jiying Zhao, Wa James Tam, and Filippo Speranza, “An image quality evaluation method based on digital watermarking”, IEEE Transactions on Circuits and Systems for Video Technology, Vol.17, No.1, pp.98-105, January 2007.
3. Sha Wang, Dong Zheng, Jiying Zhao, Wa James Tam, and Filippo Speranza, “An Accurate Method for Image Quality Evaluation Using Digital Watermarking”, IEICE Electronics Express (ELEX), Vol.2, No.20, pp.523-529, October 2005.

Conference papers generated from the research:

4. Sha Wang, Dong Zheng, Jiying Zhao, James Tam, and Filippo Speranza, “Digital Watermarking and Tree Structure Based Image Quality Estimation”, ICIMCS’11, pp. 74-77, August 5-7, 2011, Chengdu, China.

5. Sha Wang, Dong Zheng, Jiying Zhao, Wa James Tam, and Filippo Speranza, “Digital Watermarking and Perceptual Model Based Video Quality Measurement”, IEEE Instrumentation and Measurement Technology Conference, Ottawa, Ontario, Canada, pp.1729-1734, 17-19 May 2005.
6. Sha Wang, Dong Zheng, Jiying Zhao, Wa James Tam, and Filippo Speranza, “Video quality measurement using digital watermarking”, IEEE International Workshop on Haptic, Audio and Visual Environments and their Applications, Ottawa, Ontario, Canada, pp.183-188, 2-3 October 2004.
7. Sha Wang, Jiying Zhao, Wa James Tam, and Filippo Speranza, Image Quality Measurement by Using Watermarking Based on Discrete Wavelet Transform, The 32nd Biennial Symposium on Communications, Kingston, Ontario, Canada, pp.210-212, June 1-3, 2004.

Other related publications:

8. Dong Zheng, Sha Wang, and Jiying Zhao, “An RST invariant image watermarking algorithm with mathematical modeling and analysis of the watermarking processes, IEEE Transactions on Image Processing”, Vol. 18, No. 5, pp. 1055-1068, May 2009.
9. Sha Wang, Dong Zheng, and Jiying Zhao, “Error analysis for an RST invariant image watermarking algorithm”, IEEE Instrumentation and Measurement Technology Conference, pp. 504-507, May 2010, Austin, USA.
10. Dong Zheng, Sha Wang, and Jiying Zhao, “Mathematical modeling and stochastic analysis for RST invariant watermarking algorithm”, The 5th IET Visual Infor-

mation Engineering 2008 Conference (VIE'08), Xian, China, July 29 – August 1, 2008.

1.5 Thesis structure

The thesis is organized as follows. In Chapter 2, we review the watermarking based image and video quality evaluation schemes in literature. In Chapter 3, we propose a general watermarking framework that can be used for the quality evaluation applications, such as image quality evaluation and video quality evaluation applications. Based on this framework, two image quality evaluation schemes have been implemented. In Chapter 4, we present the image quality evaluation scheme using the quantization based watermarking. The corresponding experimental results are listed and discussed as well. In Chapter 5, we present the image quality evaluation scheme using the SPIHT tree structure and HVS based watermarking. Discussions about the improvements of the tree structure based scheme over the quantization based scheme are addressed as well. In Chapter 6, the experimental results achieved with the tree structure based scheme are listed and discussed. In Chapter 7, we present the video quality evaluation scheme using the SPIHT tree structure and HVS based watermarking and list the experimental results. In Chapter 8, we conclude the thesis and discuss the future work. In Appendix A, the original images and original videos used for Chapter 4 through 7 are shown.

Chapter 2

Literature review

The watermarking techniques are originally developed to protect the originality of the multimedia. The most popular applications of the watermarking techniques before 2005 were the copyright protection and multimedia content authentication. In the copyright protection application, the watermark is usually embedded robustly in the cover signal and the rotation, scaling and translation resistant features of the embedded watermark are desired [41][42][43][44][45]. In the multimedia content authentication application, fragile or semi-fragile watermarking is usually used [46][47][48].

In 2002, the concept of using a tracing watermark to evaluate signal quality was first proposed by the International Telecommunication Union (ITU) [49]. The ITU recommended that a semi-fragile watermark can be embedded into the multimedia to trace the quality degradation of the cover signal and the degraded quality of the cover signal can be evaluated by examining the degradation of the watermark. The most significant benefit of [49] is that it is unnecessary to have the original signal present to evaluate the quality of the distorted signals. With this recommendation, lots of research had been carried out to test audio quality [22], image quality [14][20][50][51][52], video

quality [53][54][55][56][57] and QoS of the mobile communications [58][59][60][61][62] using embedded watermark(s).

The watermarking process can be implemented in the spatial domain for simplicity or in the transformed domain. In literature, almost all of the watermarking schemes are developed in the transformed domain to take advantages of their special characteristics. Furthermore, it is hard to remove the watermark embedded in the transformed domain of the cover signal. The commonly used transformations are the frequency transformations, such as Discrete Fourier Transform (DFT), Discrete Wavelet Transform (DWT), Discrete Cosine Transform (DCT) and Fourier Mellin Transform (FMT).

Depending on the techniques, the watermark embedding process causes more or less quality loss to the cover signals. The more robust the watermark is embedded, the more significant quality loss will be caused by the watermarking process, and vice versa. The watermark can be embedded into the pixel values or frequency coefficients linearly for simplicity. In contrast, the watermark can be embedded with the guidance of the perceptual models to reduce the quality loss of the cover signal caused by the watermark embedding process. For image and video signals, the perceptual model can be a noise visibility function, a contrast sensitivity function, the Watson perceptual model or the Human Visual System (HVS). Among these four perceptual models, the HVS is the most sophisticated and has the best effect to model the property of human vision. As reported in [63], the HVS includes the frequency sensitivity, luminance sensitivity, color and contrast masking, etc. Thus, the HVS can provide better masking effects for the watermark embedding. For the color channel model, it is found that the human eyes are less sensitive to the alterations in the blue channel. Therefore, for the color image watermarking, it is desired to embed the watermark in the blue component of the image. Similarly, for audio signals, the model of Human Auditory System (HAS) can be used.

In [22], a watermarking based scheme is proposed to evaluate audio quality. Because different audio signals have different degradation characteristics, to accurately evaluate the audio quality, the watermark is embedded into different audio signals with different embedding strengths. In this paper, an iterative adjustment procedure is involved to automatically adjust the watermarking embedding strength by recursively testing the degradation characteristics of the audio signals.

The watermarking techniques can also be used to evaluate the Quality of Service (QoS) of mobile communications. As presented in [62], the QoS of the mobile communications includes the wireless link diagnostic, signal quality monitoring and QoS based billing service. The watermarking based QoS evaluation is usually tested on the MPEG-2 or MPEG-4 video sequence using the spread spectrum method. Thus, the watermarking based QoS evaluation schemes will be reviewed in Section 2.2.

In the following sections, we will focus on reviewing the literatures of the watermarking based image quality evaluation and the watermarking based video quality evaluation.

2.1 Watermarking based image quality evaluation

Compared to the video signals, images are smaller in size, less complex in signal structure and available free of charge on internet. Thus, images are more desirable being used to develop the theoretically complex quality metrics. In literature, lots of full reference quality metrics are proposed for image quality evaluation. In contrast, for the quality evaluations of video signals, the widely used quality metrics are still the MSE and PSNR.

In the application of the watermarking based image quality evaluation, the qual-

ity of a distorted image is evaluated by evaluating the degradation of the embedded watermark. As presented in [49], the watermark degradation evaluated by comparing the distorted watermark and the original watermark can not represent the quality of the distorted image by itself. It is always experimentally related to the existing Full-Reference quality metrics like PSNR. Therefore, the quality of the distorted image evaluated using the watermarking based scheme is usually expressed in terms of an existing objective or subjective quality metric, such as PSNR, SSIM or MOS. However, no matter which quality metric it is related to, the watermarking based quality evaluation metric is always treated as an objective quality metric. In this application, the preferred frequency domains for the watermark embedding are the DCT domain and the DWT domain. To evaluate the quality of the JPEG compressed images, the DCT domain is an inherent choice. On the other hand, the DWT provides both the spatial information and the frequency information of the cover image and is more suitable to develop a relatively complex quality metrics.

The watermark embedding strength usually controls the vulnerability of the watermark and the watermark degradation characteristics under distortions. The same watermark embedded with different embedding strengths results in different watermark degradation trends. The more accurate the watermark degradation reflecting the quality degradation of the cover image, the more accurate the image quality evaluation. Thus, the choosing of the watermark embedding strengths directly affects the accuracy of the quality evaluation. Different images have different image contents which make their degradations characteristics different under distortions. Therefore, to achieve the highest accuracy of quality evaluation, the watermark embedding strengths are desired to be different for different images.

In literature, depending on the type of the watermark, the proposed watermarking

based image quality evaluation schemes can be categorized into two groups: schemes with image-feature-independent watermarks and schemes with image-feature-dependent watermarks. In the first category, the watermark is usually an independent sequence. It can be a random binary sequence, a meaningful binary image or a noise-like signal obeying specific distribution. For the second category, some extracted features of the cover image or part of the image are used as the watermark. These two groups of schemes will be respectively reviewed in the following two sub-sections.

2.1.1 Schemes with image-feature-independent watermarks

Paper [14] is one of the first publications about the watermarking based image quality evaluation. In the paper, a randomly generated binary watermark is embedded into the middle-frequency coefficients in the DCT domain using a look-up table method. With this method, each of the possible integer DCT coefficient is randomly assigned one binary bit. For one selected DCT coefficient, if the watermark bit is different from its associated binary bit, this coefficient will be modified to the closest DCT coefficient associated with the binary bit equals to the watermark bit. The 8×8 DCT blocks are empirically divided into the low frequency, middle frequency and high frequency blocks. The vulnerability of the watermark can be changed by embedding the watermark into different numbers of blocks with different frequencies. This scheme is tested on image Barbara under JPEG compression, Gaussian noise distortion and salt pepper noise distortion. The experimental results show the embedded watermark degrades monotonically with the increasing of the distortion strength. However, the quality of the distorted image is not quantitatively evaluated. Another factor needs to be improved for this DCT based scheme is that the watermark embedding process causes significant quality loss to the cover image. As illustrated in this paper, the quality of the water-

marked image is 32.65 dB in PSNR which is not acceptable. This is resulted by the look-up table method because the binary bits associated with the DCT coefficients are randomly assigned. The random binary pattern causes bigger alterations on the DCT coefficients compared to the 0101 patterned binary sequence.

Based on the study of [14] and [16], a scheme is proposed in [51] in the DWT domain using a quantization based scheme. With the quantization based method, the odd DWT coefficients are assigned binary bit 0. In contrast, the even DWT coefficients are assigned binary bit 1. Thus, a 0101 patterned binary sequence is produced for the DWT coefficients. This scheme is tested on image Barbara under JPEG compression. The experimental results only show the watermark degradation as well. The quality of the distorted image is not numerically evaluated either. Compared to the scheme in [14], the scheme in [51] reduced the quality loss caused by the watermark embedding process, which is over 40 dB in PSNR.

The research about the scheme in [51] is extended in [20][50][52]. In [50], the relationship between the degradation of the watermark and the degradation of the cover image is proposed as the “Ideal Mapping Curve” and is quantitatively expressed, for the first time in literature. In [20], the “Ideal Mapping Curve” is generated using 20 different images. The degradation of the watermark is evaluated by computing the True Detection Rate (TDR) of the distorted watermark compared to the original watermark. The quality of the degraded image is evaluated by mapping the calculated TDR to a possible quality value referring to the “Ideal Mapping Curve” using linear interpolation. The TDR is defined as an extension of the False Detection Rate (FDR) proposed in [49]. The TDR locates in $[0, 1]$ and $TDR + FDR = 1$. In [20][50], the adaptive watermark embedding strength is iteratively tested in the watermark embedding process. The watermark embedding strength that results in the least quality evaluation

error is assigned to the cover image and the watermarked image generated using the assigned watermark embedding strength is output the watermark embedding process. As mentioned in [4], it is critically important to test the effectiveness of a quality metric across images. The scheme in [20][52] is tested on 100 different textured images under JPEG compression. The quality of the distorted images is evaluated in terms of PSNR, wPSNR and Watson JND. The experimental results clearly show that the evaluated quality is highly correlated with the quality calculated using the Full-Reference quality metrics. Thus, the watermarking based quality metric proposed in [20][50][51][52] is effective under JPEG compression and is effective enough to test any image even if it is not in the sender's image database. In [14][20][50][51][52], the watermark embedding strengths are assigned without considering the human perception characteristics. Therefore, the quality of the watermarked images can be further improved by involving the human perception characteristics in this scheme.

One algorithm derived from [14] and [51] is proposed in [64]. The watermark embedding is implemented in the middle-high frequency DCT coefficients using an improved quantization method. The quantization steps are computed with the guidance of the Watson perceptual model. In this scheme, the false negative probability of the watermark extraction is experimentally tested at the sender side and works as *a priori* at the receiver side. On the other hand, the false negative probability is theoretically generated as a function of the watermark bit error rate and the distortion distances between the distorted watermarked DCT coefficients and the undistorted watermarked DCT coefficients. At the receiver side, the watermark bit error rate is calculated by comparing the original watermark and the extracted distorted watermark. Therefore, the theoretical false negative probability function is simplified as the function of the distortion distance. Then, the pixelwise distortion distance is evaluated by approaching

the theoretical false negative probability function to the experimentally tested false negative probability using the curve fitting technique. The quality of the distorted image is evaluated in terms of PSNR using the pixelwise distortion distances which are statistically estimated for the watermarked DCT coefficients. This scheme is tested under JPEG compression and JPEG2000 compression with the experimental results showing the effectiveness of the proposed scheme. However, one limitation of this scheme needs to be pointed out: as illustrated in [64], for different images, the false negative probability function needs to be experimentally re-tested and should be known at the receiver side so that the distortion distances can be estimated. The same algorithm is further tested on the video signal under H.264 compression [65].

Another similar scheme reported in [66] is based on the studies in [20]. The watermarking process is implemented in the DCT domain using the spread spectrum method. An “Ideal Mapping Curve” is generated as well to relate the watermark degradation to the quality degradation of the cover image. A normal distributed watermark is embedded into the middle-frequency DCT coefficients with the guidance of the Watson perceptual model. An iterative process is involved in the scheme as well to minimize the quality evaluation error. The scheme is tested under JPEG compression in terms of PSNR. Compared to [20], one improvement made in [66] is that the perceptual model is used for the watermark embedding and the quality loss caused by the watermarking process is expected to be reduced. One disadvantage of this scheme is that the “Ideal Mapping Curve” is generated using 20 different images and the proposed scheme is tested on the same set of images. In this situation, the experimental testing can not prove the generality of the scheme for the images not in the sender’s image database.

In [67], a multiple-watermarking based image quality evaluation scheme is proposed in the DWT domain. This scheme is developed to evaluate image quality in terms

of PSNR under JPEG compression. Three-level DWT is applied on the test images. Three different watermarks are linearly embedded into three empirically selected DWT subbands. Similar to [20], an iterative process is used in this scheme to optimize the watermark embedding strengths. Then, the optimal watermark embedding strengths are scaled by the HVS perceptual models respectively generated for the three selected subbands. In this scheme, the degraded quality of the distorted images is designed to be evaluated in four scales. After distortion, if all the three watermarks are detected, the quality of the distorted image is evaluated as higher than 40 dB in PSNR which the authors assume the very good quality. If two watermarks are detected, the quality is evaluated as between 35 dB and 40 dB in PSNR which is assumed a good quality. If only one watermark is detected, the quality is evaluated as between 30 dB and 35 dB in PSNR which is treated as acceptable quality. Otherwise, if no watermarks are detected, the quality is evaluated as lower than 30 dB in PSNR which is assumed poor quality. These quality scales are divided empirically. This scheme can provide a rough quality evaluation and the evaluators should accept a 5-dB quality evaluation error.

2.1.2 Schemes with image-feature-dependent watermarks

Instead of the binary watermarks, some researchers use image features as the watermark. In such schemes, the image features are usually extracted and embedded into the original image. After distortion, the reference image features are tried to be reconstructed. The quality of the distorted image is assessed by comparing the distorted image features with the reconstructed reference image features. Therefore, it is expected that the image features should be semi-fragile so that the degradation of the image features can reflect the quality degradation of the cover image. On the other hand, after distortion, the reference image features should be guaranteed to be extracted with

the least distortion, which can be accomplished to embed the reference image features in a relatively robust way or using the error resilience techniques. Compared to the image-feature-independent watermarking techniques, the benefits of this type of scheme are the unnecessary of transmitting the original watermark to the receiver side and the possibility to partly recover the distorted image using the embedded information. However, at the same time, the distortion occurred to the embedded features need to be considered. The difference between the original image features and the reconstructed reference image features may reduce the accuracy of the image quality evaluation.

In [68], the authors proposed to use the statistical features extracted from the three-scale four-orientation steerable pyramid decomposed subbands of the cover image as the watermark. As stated, the marginal distribution of the coefficients in an individual decomposed subband is approximated using a two-parameter generalized Gaussian distribution, where the two parameters control the shape of the distribution. The difference between the real distribution of the steerable pyramid decomposed coefficients in one subband and the approximated generalized Gaussian distribution is calculated using the Kullback-Leibler distance. The two distribution parameters and the distribution distance are selected as the statistical features to be embedded in the original image. To improve the correct watermark extraction at the receiver side, the image features are coded using both the cyclic redundancy check code and the BCH code. After distortion, the embedded features are extracted and the quality of the distorted image is evaluated by estimating the distortion distance between the distorted image and the original image using the extracted features. The experimental results show the effectiveness of the scheme.

Another watermarking and image feature based quality evaluation scheme is presented in [69]. The scheme is implemented using a relatively complicated Discrete

Wavelet decomposition of the cover image. After applying the two level DWT to the cover image, the binary edge image extracted from the 2nd level approximation subband using Canny edge detector is employed as the binary watermark. Then, on each of the 1st level DWT detail subband, a further 2-level DWT is applied. For example, after applying 2-level DWT on the subband CH_1 , the further decomposed subbands $[hCA_2, hCH_2, hCD_2, hCV_2, hCH_1, hCD_1, hCV_1]$ are obtained. The watermark is linearly embedded into the hCA_2 . Similarly, the same binary watermark is linearly embedded into the dCA_2 and vCA_2 subbands, respectively. The special way of watermark embedding balances the invisibility and the robustness of the embedded watermark. In the meanwhile, the redundancy of the watermark embedding is three. In this scheme, the distortions that the watermarked images will go through are assumed as the White Gaussian noise distortion and JPEG compression. Both of these two distortions and the watermark signal are treated as noise. The singular value decomposition, which is famous for noise reduction, is used to estimate the reference image from the distorted watermarked image. With the estimated reference image and the distorted watermarked image, the distorted watermark is obtained by averaging the three redundant watermark extracted from the three 1st level detail subbands. The reference watermark is reconstructed by iteratively refining the evaluation of the reference image using the rank-k evaluation technique. The reconstructed watermark having the highest structural similarity with the distorted watermark is seemed as the optimal reconstructed watermark. The quality of the distorted image is assessed by evaluating the correlation coefficient calculated between the optimal reconstructed reference watermark and the distorted watermark. The experimental results show that the scheme works more efficiently with the additive white Gaussian noise than JPEG compression. This is caused by the rough assumption that all the distortions possibly affecting the images are the

noise-like distortions.

In literature, one research group extensively develop the image-feature watermarking based quality evaluation schemes into dual purpose schemes, in which the embedded watermark is used to assess both the image quality and the strengths of the distortion applied on the image [70][71]. In [70], the watermarking based quality evaluation scheme is implemented in the DWT domain and is proposed to additionally estimate the compression strength under JPEG compression. The top left 8×8 pixel block of the original image is cut off in spatial domain and is decomposed into 512 bits as the watermark. The watermark embedding process is implemented using the method same as [68] in DWT domain. At the receiver side, the distorted watermark is extracted from the distorted watermarked image and is compared with the original watermark transmitted from the sender side to give a measure for the image quality. In this scheme, the JPEG compression strength is estimated using an exhaustive search method to find the round-off version of the distorted watermark that best matches the distorted watermark. As stated in [71], during transmission, the quality of an image is mainly affected by the channel distortions and the JPEG compression. A scheme is proposed in [71] to estimate channel conditions using an embedded watermark. In this scheme, instead of the image features, the bottom eight rows of the cover image are selected to generate the watermark. Eight dummy bits are embedded into the middle frequency DCT coefficients of the selected bottom area of the cover image using the spread spectrum method mentioned in [72], in which the eight dummy bits are repeated a number of times and is scrambled for security purpose. The watermarked part is called the dummy watermarked part. Then the first 64 bits are selected from the dummy watermarked part and are spread over the rest of the cover image. After transmission, the watermark extracted from the entire cover image using the matching filter is treated as

the reference watermark. The distorted watermark is extracted from the bottom eight rows of the distorted watermarked image. The bit error rate between the recovered reference watermark and the distorted watermark is calculated and is used to estimate the conditions of the transmission channel. The quality of the transmitted image is evaluated using the channel convolution matrix.

2.2 Watermarking based video quality evaluation

The raw video signal can be defined as a consecutive sequence of still images. Thus, it is theoretically straightforward to further develop the watermarking based image quality evaluation metrics for video quality assessment. Similar to the image quality assessment metrics, the video quality metrics can be classified as the Full-Reference, Reduced-Reference and No-Reference quality metrics. The widely used Full-Reference video quality metrics include the MSE, PSNR, VQM (Video Quality Metric) and VIF (Visual Information Fidelity) [6], etc. Same as the Full-Reference image quality metrics, the Full-Reference video quality metrics require the presence of the original video signal for the quality assessment. However, in a communication system, the transmission of the original video signals will significantly increase the transmission burden. In contrast, the most famous No-Reference video quality assessment metric is the MOS (Mean Opinion Score) subjective metric which classically expresses the video quality in 5 ranges. The higher the MOS, the better the video quality, and vice versa. The MOS subjective quality is evaluated by the human viewers and gives the most reliable quality evaluation. However, it is costly and impractical for the vast quality assessments of video signals. To solve these problems resulted by the Full-Reference metrics and MOS, the watermarking based video quality evaluation method is proposed as the Reduced-

or No-reference video quality metric in literature. Because the video signal consists of a series of frames, the video watermarking scheme can be designed very flexibly. For example, one watermark can be redundantly embedded into a number of selected frames or different watermarks can be embedded into different video frames with the same or different embedding strength(s). In this case, the quality of a watermarked video sequence can be evaluated by calculating the average degradation of the embedded watermarks. With the watermarking based quality metrics, the quality of the distorted video signals is usually evaluated in terms of PSNR or MSE.

Similar to the watermarking based image quality metrics, the watermarking based video quality evaluation schemes can be categorized into the video feature independent watermarking schemes and the video feature dependent watermarking schemes. In the video feature independent watermarking schemes, the watermark(s) is independent of the video content and is usually embedded using the semi-fragile watermarking techniques. Under distortions, the video signal and the embedded watermark degrade accordingly. The watermark needs to be known or transmitted to the receiver side. The video quality is usually assessed by comparing the distorted watermark to the original watermark. In contrast, in the video feature dependent watermarking schemes, the watermark(s) is usually some features extracted from the video signal or dependent on some part of the video content. In these schemes, the distorted watermarked video signals carry both the reference watermark and the distorted watermark. The reference watermark is blindly reconstructed at the receiver side. The video quality is evaluated by comparing the distorted watermark and the reconstructed watermark. These two types of video quality metrics will be reviewed in the following two sub-sections.

2.2.1 Schemes with video-feature-independent watermarks

In [55], a semi-fragile watermarking scheme is proposed to test the quality of the low-bit rate QCIF video sequence. A binary watermark is linearly embedded in luminance components of the video sequence in spatial domain using the adaptive watermark embedding strength which is related to the quantization parameter. To reduce the quality loss caused by the watermarking process, the binary watermark is divided into a number of sub-blocks and each sub-block is embedded into a selected video frame. Moreover, for different video frames, the positions for the watermark embedding is not overlapping. The degradation of the watermark is assessed by calculating the correct extraction rate between the extracted watermark and the original watermark which is transmitted from the sender side. In this scheme, an empirical polynomial is built up to approximate the PSNR using the watermark correct extraction rate. The quality of the distorted video signal is evaluated by calculating the approximated PSNR. The experimental results show that the approximated PSNR has a high correlation with the PSNR. This paper is tested on one 2000-frame QCIF video sequence. Though the 2000 video frames may consist of enough frame content varieties, the scheme can be better proved effective if it is tested on a few more video sequences. In contrast, a watermarking based video quality metric is implemented in DCT domain for the high bit-rate video sequences [56].

In [53], a semi-fragile watermarking scheme employing a set of noise-like watermarks is proposed to evaluate the quality of the PAL interlaced TV signals. The watermark is kept the same for every N frames. For a single video frame, the watermark is linearly embedded into the middle-high frequency DCT coefficients in the 2 interlaced fields. At the receiver side, for one video frame, the correlation between distorted video frame and each original watermark is calculated. Then, the video quality is empirically evaluated

in terms of PSNR using the computed highest correlation. The scheme is tested under MPEG-2 compression and the results show the effectiveness of the scheme. This scheme hints that it is effective to embed the watermark into the video frames in the same shot with the same embedding strength.

Another scheme is reported in [73] to evaluate video quality in terms of PSNR with the AWGN channel. This algorithm is based on the study proposed in [20][51]. The watermark is embedded in the DWT domain of the video frames using a fixed watermark embedding strength. The ideal mapping curve is generated using 120 frames from six different movies. The experimental results are less accurate compared to those presented in [20]. This is caused by the lack of adaptiveness of the watermark embedding strength assignment.

According to the research about the properties of the human visual system, the human viewers are more sensitive to the degradations or artifacts of the regions with large amount of motion. The areas with relatively small amount of motion are often regarded as the still background where the degradation in quality is often negligible to the human viewers. Farias et al. [74] proposed a watermarking based video quality assessment scheme by embedding the watermark into the 8×8 blocks involving with motion based on the optical flow information. The watermarking location is determined by the amount of motion between the corresponding blocks of two consecutive frames. For one block, once the amount of motion exceeds a pre-defined threshold, the DCT is applied to this block and the watermark is linearly embedded into the middle frequency DCT coefficients to minimize the quality degradation caused by the watermark embedding. At the receiver, the watermark is extracted from the decoded video and the degradation of the watermark is assessed by calculating the total squared error between the distorted watermark and the original watermark. The scheme is tested under MPEG-2

compression and the results show that the PSNR of the distorted video monotonically decrease with the total squared error of the watermark. In the meanwhile, the extracted watermark can reflect the temporal quality change of video contents as well.

Another type of the watermarking based quality evaluation scheme is to evaluate the QoS of the mobile communication system which includes the wireless link diagnostic, signal quality monitoring and QoS based billing service [62]. The QoS based billing service is also based on the signal quality monitoring indicating how much to charge the customer by evaluating the quality of the transmitted signal. A few of such schemes are proposed in [58][59][60][61][62], in which the watermarking based QoS evaluation is tested on the MPEG-2 or MPEG-4 video sequence using the spread spectrum method. Therefore, the evaluation of the QoS can be based on the quality evaluation of the transmitted video signal.

In literature, the watermarking based QoS evaluation scheme is first proposed in [58][59] and is tested on the MPEG-2 video signal. Inherent with the MPEG-2 standard, the watermark is embedded in the middle-frequency DCT coefficients of the raw video signal. To improve the correct extraction rate of the distorted watermark, the watermark is redundantly embedded into several selected raw video frames. After watermark embedding, the raw video sequence is compressed using MPEG-2 and the compressed video signal is finally transmitted. At the receiver side, a matched filter is used to despread the watermark and the quality of the current video frame is evaluated by computing the MSE between the original watermark and the distorted watermark. The overall quality of the whole video signal is evaluated by averaging the normalized MSEs obtained from the watermarked video frames. Further testings on the MPEG-4 video signals are addressed in [75]. For real time QoS evaluation, the watermark embedding can be proceeded in the Walsh Hadamard transform domain to increase the

computational efficiency [60][61]. To minimize the perceptual quality loss caused by the watermarking process, instead of embedding the watermark in the luminance components of the video signal, the authors in [62] proposed to embed the watermark in the T component in the YST color space. In the new YST color space, the Y component is the luminance of the video signal. The S component is achieved by mathematically averaging the U and V components in the YUV color space and indicates the color of the human skin which is an important quality concern in the 3G video phone calls. The T component is defined as the orthogonal of the YS plane and is the least important component in the YST color space. In these papers, the QoS is evaluated using MSE.

2.2.2 Schemes with video-feature-dependent watermarks

In [76], a watermarking based video quality assessment metric is proposed based on the research done in [68]. The statistical intra and inter features extracted from the video signal are used as the watermark which is robustly embedded into the low frequency components of the 3D-DCT coefficients using the index modulation techniques. To increase the robustness of watermarking, error control coding is used to enhance the robustness of the watermark embedding so that the reference watermark can be fully recovered after distortion. At the receiver side, the quality of the distorted video is assessed by evaluating the extracted watermark. This scheme is tested under various distortions, such as MPEG-2 compression, Gaussian noise addition and Gaussian blur line jittering. The quality of the distorted video signals is not quantitatively evaluated. The experimental results show that with the increasing of the distortion strength, the distortion caused to the video sequences increase monotonically.

2.3 Summary

In this chapter, various image and video watermarking based quality evaluation schemes are reviewed. Based on the dependence on the cover signal, the watermarking schemes are categorized as the signal-feature-dependent watermarking schemes and the signal-feature-independent watermarking schemes.

In the signal-feature-dependent schemes, normally, some features extracted from the original signal are embedded into the original signal. At the receiver side, the embedded features are reconstructed and are used as the reference watermark. The features extracted from the distorted signal are used as the distorted watermark. The watermark degradation is assessed by comparing the distorted watermark to the reference watermark. In such schemes, the error resilience techniques are usually utilized to minimize the distortion that may occur on the embedded reference watermark. Instead of the original signal features, using the reconstructed signal features as the reference watermark may introduce additional inaccuracy to the quality estimation. Moreover, for these schemes, it is hard to examine which kind of signal features are suitable for providing accurate quality estimation. The signal-feature-independent schemes simplify the situation. In such schemes, the watermark is independent of the original signal and needs to be known at the receiver side. The watermark degradation is evaluated using the distorted watermark and the original watermark. One challenging task to develop the signal-feature-independent watermarking based quality metric is to enable the embedded watermark to accurately reflect the quality changes of cover signal under distortions. This requires the watermark being adaptively embedded in cover signals according to the characteristics of cover signals, so that the embedded watermark degrades in a similar way as the cover signals under distortions. This is a critical part of

the whole scheme and directly affects the accuracy of quality estimation.

The reviewed watermarking based quality evaluation schemes are to measure the fidelity of the distorted signal with respect to the original signal. That is: the more similar the distorted signal compared to the original signal, the higher the fidelity of the distorted signal. In the watermarking based scheme, the use of the perceptual model is to reduce the fidelity loss caused by the watermark embedding process. A concept of the absolute image quality is introduced in [77]. Through experiments, the researchers find that the sharpness and colorfulness can greatly affect the perceptual quality of an image as well. “A reduction of sharpness or colorfulness from the original work to the distorted one corresponds to a decrease in the perceived quality” [77]. The sharpness means high contrast and can be measured by evaluating the isotropic local contrast. The colorfulness can be determined by the average chroma and the spread of the distribution of the chroma values. The sharpness and colorfulness defined in the paper are designed to be able to work with any of the existing quality metric in order to give a comprehensive perceptual quality evaluation. It will be promising to develop the watermarking based quality evaluation scheme to blindly evaluate the absolute perceptual quality of a distorted signal.

Chapter 3

The proposed framework for watermarking based quality evaluation

3.1 The proposed framework for watermarking based quality evaluation

In the watermarking based quality evaluation metrics, a semi-fragile watermark is usually embedded throughout the original image with appropriate embedding strength. The embedded watermark goes through the same distortion as the cover image. Then the image quality is evaluated by assessing the degradation of the distorted watermark. An example of the degradations of the watermark achieved under JPEG compression with different compression strengths is shown in Fig. 3.1. The numbers, from 100 to 10 with step -10, are the quality factors used in Matlab for JPEG compression. The

lower the quality factor, the higher the compression ratio. We know theoretically that, with the decreasing of the quality factor, the quality of an image decreases. In the meantime, it clearly shows in Fig. 3.1 that, with the decreasing of the quality factor, the quality of the watermark degrades as well. Thus, the degradation of the watermark can roughly reflect the quality degradation of the compressed image.

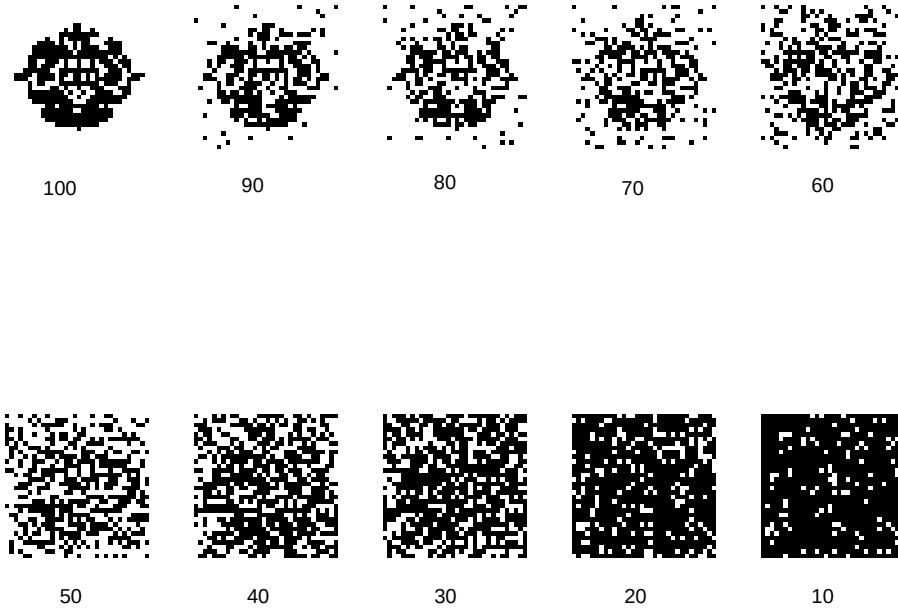


Figure 3.1: Watermark degradation under JPEG compression.

In Fig. 3.1, the numbers shown under the extracted watermarks are the quality factors used in the JPEG compression. The smaller the quality factor, the bigger the JPEG compression strength.

For the watermarking based quality evaluation applications, we propose to evaluate the degradation of the watermark using the True Detection Rate (TDR) defined in Equ. (3.1). The TDR is calculated by comparing the degraded watermark to the original watermark bit by bit. The higher the calculated TDR, the less degraded the extracted watermark, and vice versa.

$$TDR = \frac{\text{Number of correctly detected watermark bits}}{\text{Total number of watermark bits}} \quad (3.1)$$

For one image under a fixed distortion, the calculated TDR of a watermark may be quite different if we embed it with different embedding strengths. Thus, the TDR can not represent the quality of the cover image by itself. Therefore, we propose to use the watermarking based quality metric performing quality evaluation in terms of some existing Full-Reference quality metric, such as PSNR, wPSNR, JND, SSIM or VIF. In other words, the quality of the distorted image will be evaluated in terms of some quality metric like PSNR by only evaluating the calculated TDR. To this end, a link between the degradation of the watermark, TDR , and the degradation of the cover image needs to be built so that the quality of the distorted image can be estimated by mapping the calculated TDR referring to the degradation link as depicted in Equ. (3.2).

$$\hat{Q} = f(TDR) \quad (3.2)$$

where, $\hat{Q}$ is the estimated quality. $f(\bullet)$ is the degradation link mentioned above which is also called the quality mapping function. In this thesis, the quality mapping function is named the “Ideal Mapping Curve” and is experimentally generated defining the monotonous relationship between the possible TDR values and the quality values. The generation of the “Ideal Mapping Curve” will be presented in Section 3.4.

One challenging task of the watermarking based quality metric is to enable the embedded watermark to accurately reflect the quality changes of the cover images under distortions. This requires that the watermark be embedded in the cover images adaptively according to the characteristics of the cover images so that the embedded

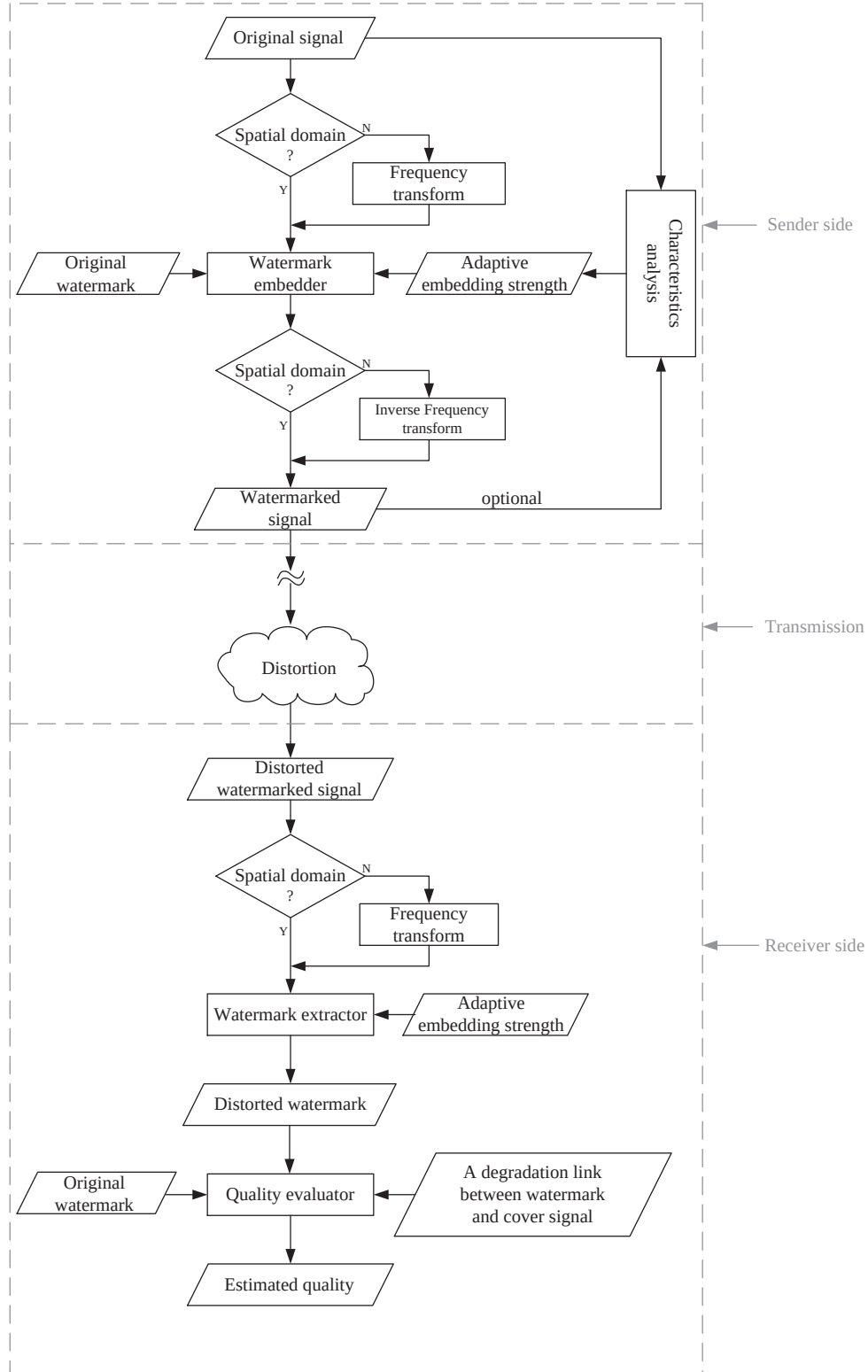


Figure 3.2: The general framework for signal-feature-independent watermarking based quality evaluation applications.

watermark degrades in a similar way as the cover image under distortions. This is a critical part of the whole scheme and directly affects the accuracy of the quality evaluation. Thus, we propose a general quality evaluation scheme based on adaptive watermarking as shown in Fig. 3.2. In the framework, the watermark embedding strength is adaptively assigned by analyzing the characteristics of the cover image. Because different images have different textures which result in different characteristics under distortions, the watermark embedding strength should be different for different images.

In this thesis, we propose to implement the watermark embedding and extraction processes in the DWT domain to take advantages of both the spatial and frequency information of the cover image. Three level DWT decomposition is applied on the cover image so that enough frequency information of the cover image can be obtained for the watermark embedding while the computational complexity can be kept relatively low. In the next two sections, we introduce the Discrete Wavelet Transform and its properties. In Section 3.5, we propose the method to evaluate the effectiveness of a watermarking based quality evaluation scheme.

3.2 Discrete Wavelet Transform (DWT)

The DWT is defined as [78]:

$$W_{\varphi}(j_0, k) = \frac{1}{\sqrt{M}} \sum_x f(x) \varphi_{j_0, k}(x) \quad (3.3)$$

$$W_{\psi}(j, k) = \frac{1}{\sqrt{M}} \sum_x f(x) \psi_{j, k}(x) \quad (3.4)$$

where, $j \geq j_0$.

The Inverse DWT (IDWT) is defined as:

$$f(x) = \frac{1}{\sqrt{M}} \sum_k W_\varphi(j_0, k) \varphi_{j_0, k}(x) + \frac{1}{\sqrt{M}} \sum_{j=j_0} \sum_k W_\psi(j, k) \psi_{j, k}(x) \quad (3.5)$$

where $f(x)$ is the host signal, $\varphi_{j_0, k}(x)$ and $\psi_{j, k}(x)$ are two basis functions of the discrete variable. Normally we let $j_0 = 0$ and select M to be a power of 2 (i.e. $M = 2^J$) so that the summations in Equ. (3.3), Equ. (3.4) and Equ. (3.5) are performed over $x = 0, 1, 2, \dots, J-1$, $j = 0, 1, 2, \dots, J-1$ and $k = 0, 1, 2, \dots, 2^j - 1$. The coefficients defined in Equ. (3.3) and Equ. (3.4) are respectively called approximation and detail coefficients.

$\varphi_{j, k}(x)$ is a member of the set of expansion functions achieved by translating and scaling a scaling function, $\varphi(x)$:

$$\varphi_{j, k}(x) = 2^{j/2} \varphi(2^j x - k) \quad (3.6)$$

$\psi_{j, k}(x)$ is a member of the set of expansion functions achieved by translating and scaling a wavelet function $\psi(x)$:

$$\psi_{j, k}(x) = 2^{j/2} \psi(2^j x - k) \quad (3.7)$$

The wavelet can be expressed as the filter bank operations. The two-scale equation is an important formula connecting the scaling function to itself at two different scales. The filter functions are defined by the two-scale equation implicitly.

$$\varphi(t) = \sum_k h_\varphi(k) \sqrt{2} \varphi(2t - k) \quad (3.8)$$

The orthonormal property for the scaling function is defined as:

$$\int_{-\infty}^{\infty} \varphi(t) \varphi(t - n) dt = \delta(n) \quad (3.9)$$

Shifting t by n in Equ. (3.8), we have

$$\varphi(t - n) = \sum_k h_\varphi(k) \sqrt{2} \varphi(2(t - n) - k) = \sum_k h_\varphi(k) \sqrt{2} \varphi(2t - (k + 2n)) \quad (3.10)$$

Let $m = k + 2n$,

$$\varphi(t - n) = \sum_m h_\varphi(m - 2n) \sqrt{2} \varphi(2t - m) = \sum_k h_\varphi(k - 2n) \sqrt{2} \varphi(2t - k) \quad (3.11)$$

From Equ. (3.8), we can see that $h_\varphi(k)$ are the expansion coefficients of $\varphi(t)$ on the basis function $\sqrt{2} \varphi(2t - k)$. Similarly, in Equ. (3.11), $h_\varphi(k - 2n)$ are the expansion coefficients of $\varphi(t - n)$ on the same basis function. Since we have the assumption in Equ. (3.9) and the basis functions are orthonormal, it is intuitive that:

$$\sum_k h_\varphi(k) h_\varphi(k - 2n) = \delta(n) \quad (3.12)$$

Here, we derive the Double shift orthogonality property of the filter function. If $n = 0$, we get $\sum_k h_\varphi^2(k) = \delta(0) = 1$. The same deviation is valid for the wavelet function, $\psi(t)$, and the filter $h_\psi(n)$:

$$\psi(t) = \sum_n h_\psi(n) \sqrt{2} \psi(2t - n) \quad (3.13)$$

To design the wavelet filters $h(n)$ including $h_\varphi(n)$ and $h_\psi(n)$, the following properties of the filters must be observed:

1. Normalization: $\sum_k h^2(k) = 1$.
2. Double shift orthogonality: $\sum_k h(k)h(k - 2n) = 0, n \neq 0$.
3. Low pass filter $h_\varphi(n)$: $\sum_k (-1)^k h_\varphi(k) = 0$.
4. High pass filter $h_\psi(n)$: $h_\psi(k) = (-1)^k h_\varphi(N - k)$.

For example, the filter functions of the length 2 Haar wavelet are $h_\varphi = [1/\sqrt{2}, 1/\sqrt{2}]$, and $h_\psi = [1/\sqrt{2}, -1/\sqrt{2}]$.

The *db*-2 length 4 wavelet filters have the following filter bank ($\alpha = \frac{\pi}{3}$):

$$h_\varphi(0) = \frac{1 - \cos\alpha + \sin\alpha}{2\sqrt{2}} \quad (3.14)$$

$$h_\varphi(1) = \frac{1 + \cos\alpha + \sin\alpha}{2\sqrt{2}} \quad (3.15)$$

$$h_\varphi(2) = \frac{1 + \cos\alpha - \sin\alpha}{2\sqrt{2}} \quad (3.16)$$

$$h_\varphi(3) = \frac{1 - \cos\alpha - \sin\alpha}{2\sqrt{2}} \quad (3.17)$$

And $h_\psi(0) = h_\varphi(3), h_\psi(1) = -h_\varphi(2), h_\psi(2) = h_\varphi(1), h_\psi(3) = -h_\varphi(0)$, which are obtained by solving the above properties of wavelet filters.

Since the DWT can be formulated as a filtering operation with two related FIR filters, low-pass filter h_φ and high-pass filter h_ψ . Both $W_\varphi(j, k)$ and $W_\psi(j, k)$, the scale j approximation and the detail coefficients, can be computed by convolving $W_\varphi(j+1, k)$. The scale $j+1$ approximation coefficients, with the time-reversed scaling and wavelet vectors, $h_\varphi(-n)$ and $h_\psi(-n)$, and

$$W_\psi(j, k) = h_\psi(-n) * W_\varphi(j + 1, n)|_{n=2k, k \geq 0} \quad (3.18)$$

$$W_\varphi(j, k) = h_\varphi(-n) * W_\varphi(j + 1, n)|_{n=2k, k \geq 0} \quad (3.19)$$

Sub-sampling the results by 2, expressed in Equ. (3.18) and Equ. (3.19) and illustrated in Fig. 3.3.

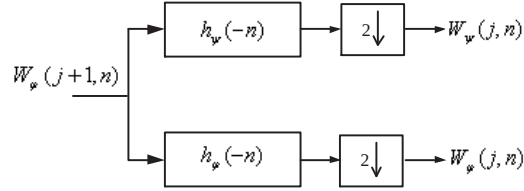


Figure 3.3: A DWT analysis bank.

The filter bank in Fig. 3.3 can be iterated to implement multi-resolution analysis. The IDWT can be implemented by up-sampling and synthesis filtering. The one-dimensional DWT and IDWT can be extended to two-dimensional.

Wavelet algorithms are recursive and the smoothed data d_i becomes the input for the next step of the wavelet transform. A widely used example is the Haar wavelet filters whose low pass filter is $H_\varphi(\omega) = \frac{1}{2} + \frac{1}{2}e^{-j\omega}$ and high pass filter is $H_\psi(\omega) = \frac{1}{2} - \frac{1}{2}e^{-j\omega}$.

To ensure the IDWT and DWT relationship, the following orthogonality condition is necessary:

$$|H_\varphi(\omega)|^2 + |H_\psi(\omega)|^2 = 1 \quad (3.20)$$

As stated in [79]: “Orthonormal transforms are of interest because they can be used to re-express a time series in such a way that we can easily reconstruct the series

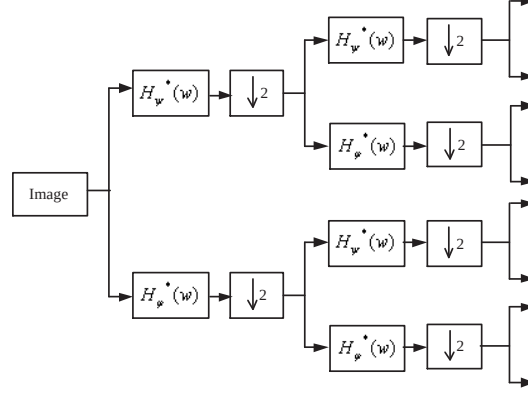


Figure 3.4: The DWT down-sampling.

from its transform.” In other words, we can reconstruct the signal inversely from its decomposed wavelet coefficients without loss of information.

We can easily extend the wavelet to the two dimensional image processing as shown in Fig. 3.4. There are lots of other wavelet filters such as the Daubechies wavelet transform, which have their own wavelet and scaling functions.

If we apply the 3-level DWT decomposition on a image, the pyramid structure of DWT decomposed image can be shown as Fig. 3.5.

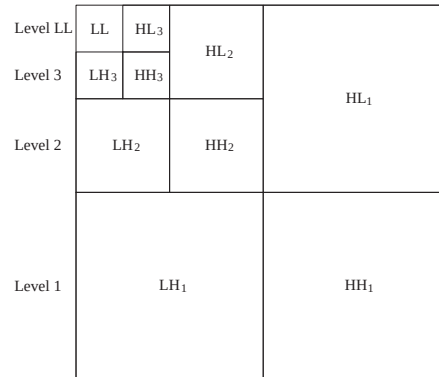


Figure 3.5: The DWT decomposed blocks.

In Fig. 3.5, LL is the approximation coefficient subband. The other 9 blocks are

the detail coefficient subband. In this thesis, we denote subband HL1, LH1 and HH1 as the subbands in the first DWT level; HL2, LH2 and HH2 as the subbands in the second DWT level; HL3, LH3 and HH3 as the subbands in the third DWT level. As the approximation coefficient subband, LL is denoted as the subband in the DWT level LL to distinguish it with the other 9 detail coefficient subbands. These denotations for the 10 DWT decomposed subbands will be used throughout the thesis.

3.3 The proposed method to evaluate the image degradation characteristics

As shown in Fig. 3.2, the watermark embedding strength is adaptively assigned by analyzing the quality degradation characteristics of the cover image under distortions. Thus, we are expecting that the adaptive watermark embedding strength can bring similar degradation characteristics to the watermark as the cover image. Theoretically, the accuracy of the quality evaluation can be greatly improved using this strategy. In the 3-level DWT domain, the watermark embedding strength can be greatly affected by embedding different numbers of watermark bits in different DWT levels. We define the watermark embedding strength as $A_{wb} = [a_1, a_2, a_3, a_{LL}]$, where a_i indicates the number of watermark bits that will be embedded into each selected subband on DWT level i and $i \in [1, 2, 3, LL]$. Thus, one way to approximate the quality degradation characteristics is that, for the images with different characteristics, A_{wb} is different. In this section, we will analyze the image characteristics in terms of PSNR under JPEG compression. Similar analysis can be made for other signals (video or audio signals) in terms of other quality metrics (such as wPSNR and JND, SSIM or VIF) under other distortions (such as JPEG2000 compression, Gaussian noise distortion and low-pass

filtering).

Two examples are illustrated in Fig. 3.6. After applying 3-level DWT on the original image Lena and Baboon in Fig. 3.6 (a) and (c) respectively, we achieve the wavelet decomposed subbands and put them in the pyramid structure as shown in Fig. 3.6 (b) and (d). By observing the image Lena and Baboon, we can have a general impression on their texture characteristics. The image Lena has a girl head in the center and some bland background around. It can be assumed that the image Lena has a wide variety of frequency components. The image Baboon has a large monkey head with lots of hair and beard. So it can be assumed that the most dominant frequency components in image Baboon are high frequency components. These assumptions can be illustrated in Fig. 3.6, where Baboon has much more components in the detail coefficient subbands than Lena. Under JPEG compression with quality factors varying from 100 to 10 with a step of -10, the quality degradation of image Lena and Baboon are illustrated using the solid curves in Fig. 3.7 and Fig. 3.8, respectively. The solid curves clearly show that image Baboon has a quicker quality drop than image Lena. Therefore, for the images with rich high frequency components, we need to make the watermark more vulnerable to reflect the quick drop of quality. For the images with dominant middle frequency components, we embed the watermark with a moderate strength to make it reflect the moderate quality change of the distorted images. On the contrary, we embed the watermark with stronger embedding strength in the images with dominant low frequency components.

Recall Equ. (3.18) and Equ. (3.19), the detail coefficient blocks in the 2^{nd} or 3^{rd} DWT level are obtained by decomposing the approximation coefficient block in their previous DWT level. Thus, the detail coefficients in the 1^{st} DWT level are the relatively higher frequency coefficients comparing to those in the 2^{nd} DWT level. Similarly, the

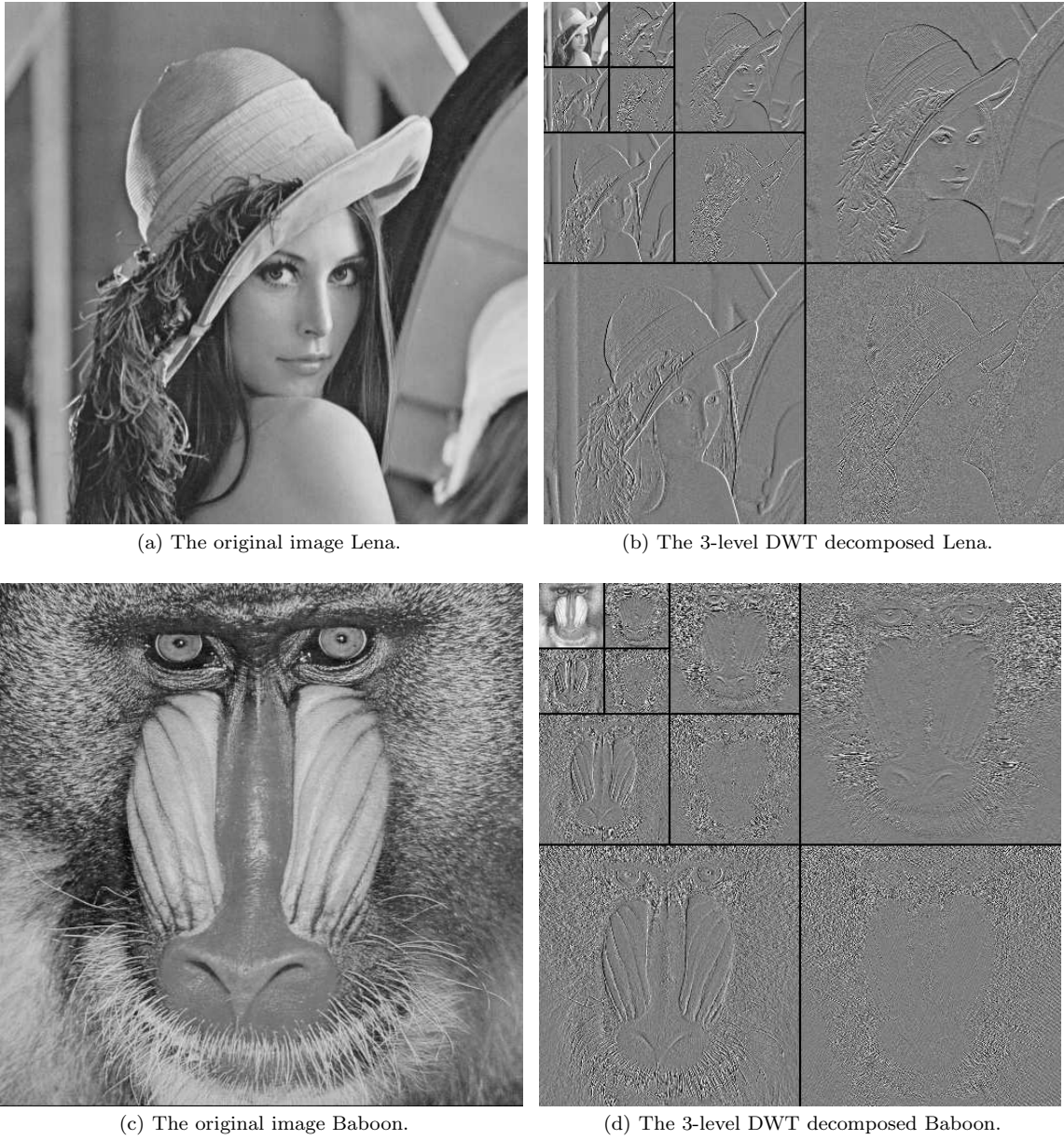


Figure 3.6: The original Lena and Baboon with their wavelet decomposed images.

detail coefficients in the 2^{nd} DWT level are the higher frequency coefficients comparing to those in the 3^{rd} level. Usually, the higher frequency DWT subbands are more sensitive to distortions, and vice versa. Based on this analysis, under the same distortion, the DWT decomposed subbands in different DWT levels will behave differently and the quality degradations of these subbands contribute more or less to the quality loss of the cover image. Thus, we propose to embed more watermark bits in the DWT subbands that contribute more to the quality loss of the cover image and less watermark bits in the other subbands.

A small experiment is implemented to test the quality degradation characteristics of each the DWT decomposed subbands under JPEG compression. The quality degradation characteristics of one DWT decomposed subband can be reflected by evaluating the quality changes of the same subband distorted using a set of compression strengths. To test the degraded quality of subband ch under quality factor 50, we compress the cover image using quality factor 50 and, then, apply 3-level DWT to the compressed image and achieve the distorted block ch' . The degraded quality of the subband ch' is calculated in terms of PSNR comparing to the subband ch . The quality factors, $[10 : 10 : 100]$, are used in this small experiment. Based on the experimental results, the three subbands in the same DWT level have similar characteristics. Thus, the degradation characteristics of one DWT level is achieved by averaging the degradation characteristics of the three detail subbands in the same DWT level.

The quality degradation characteristics of different DWT subbands tested on image Lena and image Baboon are shown in Fig. 3.7 and Fig. 3.8, respectively. The quality degradation characteristics of the whole image is shown in the figures as well. The “performance of the subband in Level LL” indicates the quality degradation characteristics the approximation subband and its contribution to the quality degradation of

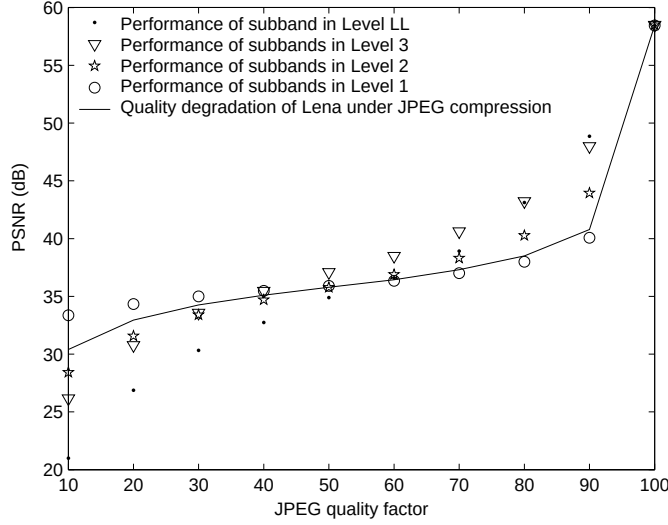


Figure 3.7: The degradation characteristics of image Lena.

the cover image. The “performance of the subbands in Level i ” indicates the average quality degradation characteristics of the three detail coefficient subbands in DWT level i under JPEG compression, where $i \in [1, 2, 3]$.

It clearly shows in Fig. 3.7 that, when the quality factor is large, the JPEG compression affects the high frequency components mostly. With the quality factor becoming smaller and smaller, the middle frequency and low frequency components are starting to be affected by the JPEG compression. Thus, the watermark bits should be embedded throughout the four frequency subbands.

Similarly, in Fig. 3.8, the detail blocks in the DWT level 1 contribute the most to the quality loss of image Baboon, which indicates that the JPEG compression with quality factors varying from 100 to 10 with a step of -10 hardly affects the frequency components in the DWT level 2, 3, and LL. In this case, we can embed the entire watermark in the coefficients of the DWT level 1 and achieve the most fragile watermark.

Therefore, the watermark embedding strength, A_{wb} , assigned to image Lena should

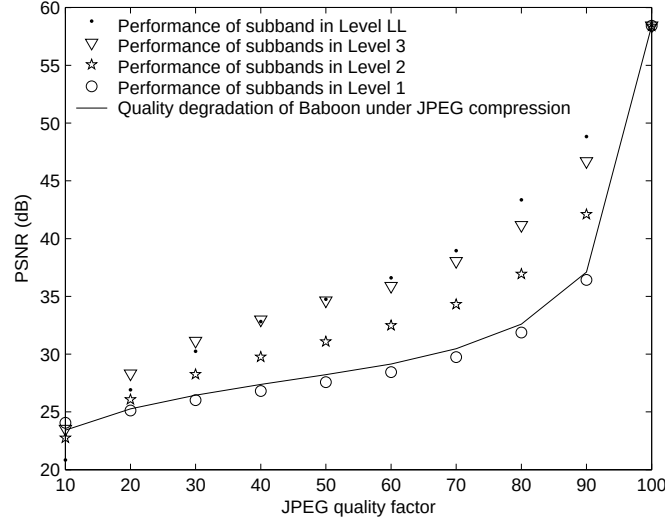


Figure 3.8: The degradation characteristics of image Baboon.

be stronger than that of image Baboon. By doing so, we can control the degradation characteristics of the watermark referring to the degradation characteristics of the cover images. With the analysis presented above, it is experimentally implementable to assign the binary watermark a similar degradation characteristics as the cover image. Therefore, the watermark will degrade in a similar way as the cover image and it is theoretically reasonable to use the degradation of the watermark accurately evaluate the quality degradation of the cover image.

3.4 “Ideal Mapping Curve” – the degradation relationship between the watermark and images

As proposed in Section 3.1, the quality of a distorted image is evaluated by assessing the degradation of the embedded watermark. Therefore, it is necessary to build up a

degradation relationship between the watermark and the images so that, once the TDR is calculated, the quality of the distorted image can be quantitatively evaluated by mapping the TDR referring to the degradation relationship. In this case, it is expected to experimentally relate the TDR with possible quality values of the distorted images. To this end, we propose to express the degradation relationship using a set of data with coordinates $(TDR, Quality)$. The *Quality* in the coordinates is the possible quality value and is calculated between the distorted image and the original image in terms of a desired quality metric. As mentioned previously, the watermarking based quality metric can only conduct quality evaluation in terms of the existing quality metrics, such as PSNR, wPSNR, JND, SSIM or VIF. Thus, for each desired quality metric, one degradation relationship needs to be built up. For example, if we want to evaluate image quality in terms of JND under JPEG compression, one TDR-JND-JPEG degradation relationship needs to be built up. Similarly, if we want to assess image quality in terms of SSIM under low-pass filtering, one TDR-SSIM-filtering degradation relationship needs to be built up. This strategy aims to reduce the quality evaluation error.

In this thesis, we name the degradation relationship between the watermark and the images as the “Ideal Mapping Curve”. This concept is first proposed with our watermarking and quantization based image quality evaluation scheme in literature [50][52]. The “Ideal Mapping Curve” is pre-generated. Whenever a quality evaluation needs to be conducted, the desired “Ideal Mapping Curve” is retrieved to assess signal quality. Thus, the generation of the “Ideal Mapping Curve” does not reduce the computational efficiency of the watermarking based quality metric.

The “Ideal Mapping Curve” is mathematically expressed using a TDR-Quality curve. To accommodate more degradation characteristics of different textured images, the “Ideal Mapping Curve” is usually tested on a number of different textured images

using the user-defined watermark embedding strength(s). The watermark embedding strength(s) can be a fix value for all the test images or different values for different images, which is decided by the way that the watermark embedder works. A typical “Ideal Mapping Curve” generator is shown in Fig. 3.9. In the figure, N is the number

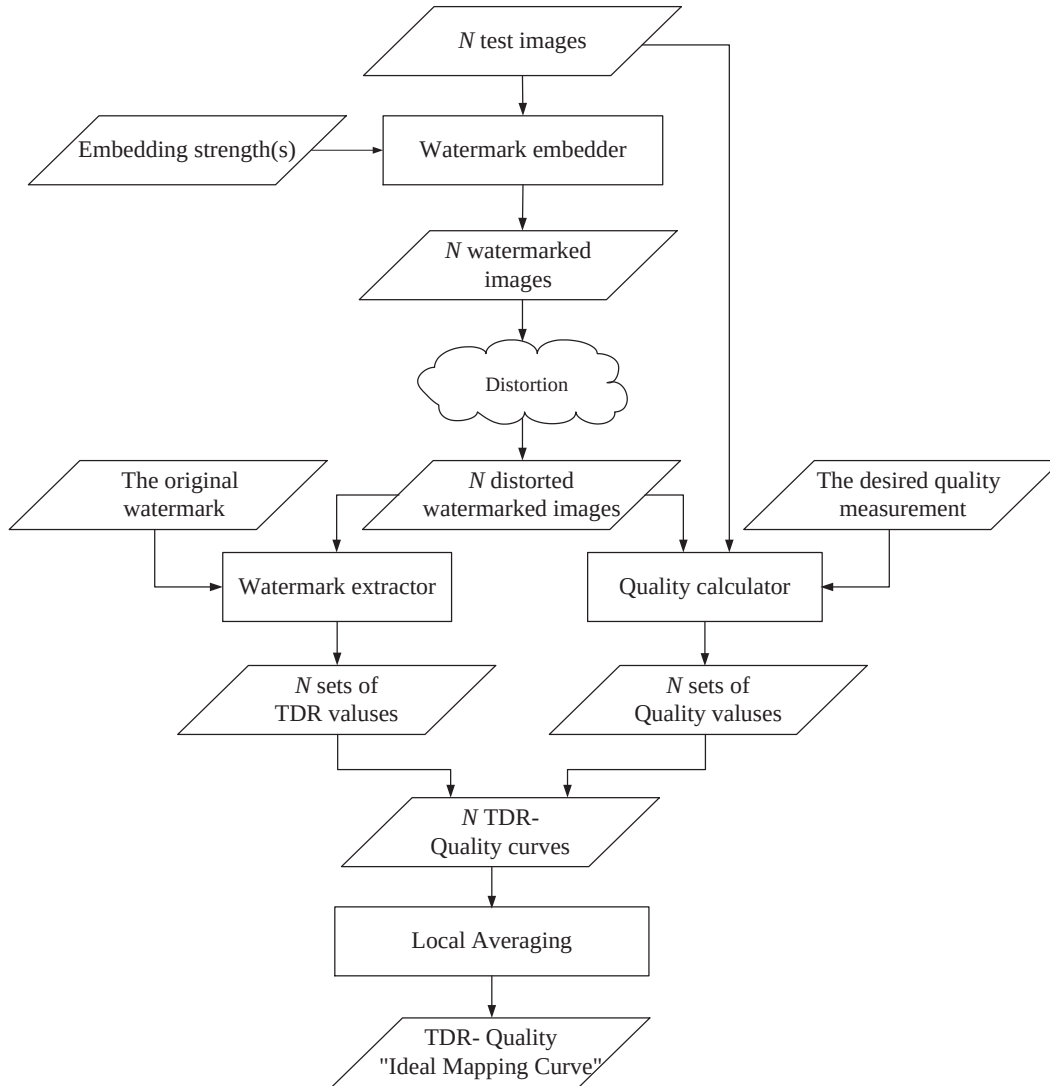


Figure 3.9: The “Ideal Mapping Curve” generator.

of images used in the testing. If N images are involved in the testing, as a result, we will

obtain N sets of TDR-Quality curves. Then, the “Ideal Mapping Curve” is generated by locally averaging the N sets of TDR-Quality curves. The local averaging consists of the following steps:

1. Respectively distort the N test images using the specified distortion strengths and calculate the Quality values and TDR values under each distortion strength. Find the range of the possible Quality values.
2. List all the desired target Quality values in the Quality range found in step 1.
3. Compare each of the calculated Quality values with all the target Quality values. By finding the closest target Quality, all the calculated Quality values and their corresponding TDR values are categorized into different groups.
4. Average the TDR values in each group.
5. Generate the “Ideal Mapping Curve” by relating the target Quality values to the mean TDR values.

The same local averaging rule can be used in terms of different quality metrics under different distortions. The “Ideal Mapping Curve” is a general concept and it can be used with the watermarking schemes that are different from the ones proposed in this thesis. With the “Ideal Mapping Curve”, the original signal is unnecessary to be present for the quality evaluation at the receiver side.

The generation of the “Ideal Mapping Curve” can be extended for the video signals using the same steps presented above. In this case, video frames are used instead of images.

3.5 The proposed method to evaluate the effectiveness of the watermarking based quality evaluation schemes

For a watermarking based quality evaluation scheme, we propose to assess its effectiveness by evaluating the correlation of the estimated quality and the calculated quality. The calculated quality is achieved by computing the quality of the distorted watermarked image comparing to the watermarked image or original image using some existing quality metric, such as PSNR, wPSNR, JND, SSIM or VIF. The estimated quality is obtained using the watermarking based quality evaluation scheme in terms of the same quality metric. The more accurate the evaluated quality comparing to the calculated quality, the more effective the watermarking based quality evaluation scheme, and vice versa.

In this thesis, we evaluate the effectiveness of the watermarking based quality evaluation scheme using the Mean Absolute Error (MAE), the Pearson correlation coefficient and the RMSE.

The MAE (Mean Absolute Error) is defined in Equ. (3.21).

$$\begin{aligned}
 MAE &= \frac{\sum_{i=1}^N \sum_{j=1}^k |X(i, j) - Y(i, j)|}{N \cdot k} \\
 &= \frac{\sum_{i=1}^N \sum_{j=1}^k |\Delta Q(i, j)|}{N \cdot k}
 \end{aligned} \tag{3.21}$$

where, N is the total number of images tested in the experiments; i is the current number of image; k is the total number of distortion strengths used to distort the watermarked images; j is the current number of the distortion strength; $X(i, j)$ and

$Y(i, j)$ are respectively the calculated quality and the evaluated quality of the i^{th} image distorted using the j^{th} distortion strength; ΔQ is the quality evaluation error. The smaller the calculated MAE, the more accurate the quality evaluation.

The Pearson correlation coefficient [80] between the calculated quality and the evaluated quality is defined in Equ. (3.22). $Corr_p$ presents the Pearson correlation coefficient; the X and Y are the evaluated quality and the calculated quality, respectively; n is the total number of X or Y that involved in the comparison, the S_X and S_Y are the sample standard deviation of X and Y , respectively. The Pearson correlation has a range: $|Corr_p| \leq 1$. If $Corr_p = 1$, X and Y has a perfect positive linear relationship. If $Corr_p = -1$, X and Y has a perfect negative linear relationship. If $Corr_p = 0$, X and Y are independent. When $Corr_p$ is between ± 1 , it indicates the degree of the linear relationship between X and Y .

$$\begin{aligned}
 Corr_p &= \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{(n-1)S_X S_Y} \\
 &= \frac{\sum_{i=1}^n X_i Y_i - \frac{\sum_{i=1}^n X_i \sum_{i=1}^n Y_i}{n}}{\sqrt{\left(\sum_{i=1}^n X_i^2 - \frac{(\sum_{i=1}^n X_i)^2}{n}\right) \left(\sum_{i=1}^n Y_i^2 - \frac{(\sum_{i=1}^n Y_i)^2}{n}\right)}} \quad (3.22)
 \end{aligned}$$

In our experiments, we are expecting $Corr_p \in [0, 1]$. When $Corr_p = 1$, we achieve the highest accuracy of the quality evaluation. When $Corr_p = 0$, we get the lowest accuracy of the quality evaluation. The higher the calculated Pearson correlation, the more accurate the quality evaluation. We deem the proposed scheme has high accuracy if the calculated Pearson correlation is higher than 0.9 and has good accuracy if the calculated Pearson correlation locates between 0.8 and 0.9.

The definition of RMSE is presented in Section 1.1.1.1.

3.6 Summary

In this chapter, we proposed a general framework for the signal-feature-independent watermarking based quality evaluation schemes. This framework can be used to develop the watermarking based image quality metrics, video quality metrics and audio quality metrics. A binary watermark is proposed to be embedded in the DWT domain of the cover signal. The embedded watermark goes through the same distortion as the cover signal. The degradation of the watermark is proposed to be evaluated using the TDR in Equ. (3.1). In Section 3.4, the definition of the degradation relationship between the watermark and cover images are presented as an example. Similar degradation relationships can be generated for video and audio signals. The quality of the distorted signal can be evaluated by mapping the calculated TDR referring to the degradation relationship. Associated with the general framework, the method to evaluate the effectiveness of the watermarking based quality evaluation scheme is reported in Section 3.5.

Another important factor reported in this chapter is that the watermark is proposed to be embedded in the cover signals with adaptive embedding strengths. For different textured cover signals, the watermark embedding strengths are different. The adaptive watermark embedding strengths make the embedded watermark able to degrade accordingly with the cover image and then help minimize the quality evaluation error. In Section 3.3, details are presented to illustrate how we evaluate the degradation characteristics of images under JPEG compression. Similar evaluations can be made for other signals under other distortions.

Chapter 4

Image quality evaluation using quantization based watermarking

With the analysis made in Chapter 3, an adaptive watermarking and quantization based image quality estimation scheme is proposed in this chapter.

4.1 The quantization based watermark embedding process

Fig. 4.1 shows the quantization based watermark embedding process which consists of the following steps:

1. Embed the watermark into the original image.
 - (a) Apply 3-level DWT to the original image and we get 10 DWT decomposed subbands and the DWT levels as numbered in Fig. 4.2.

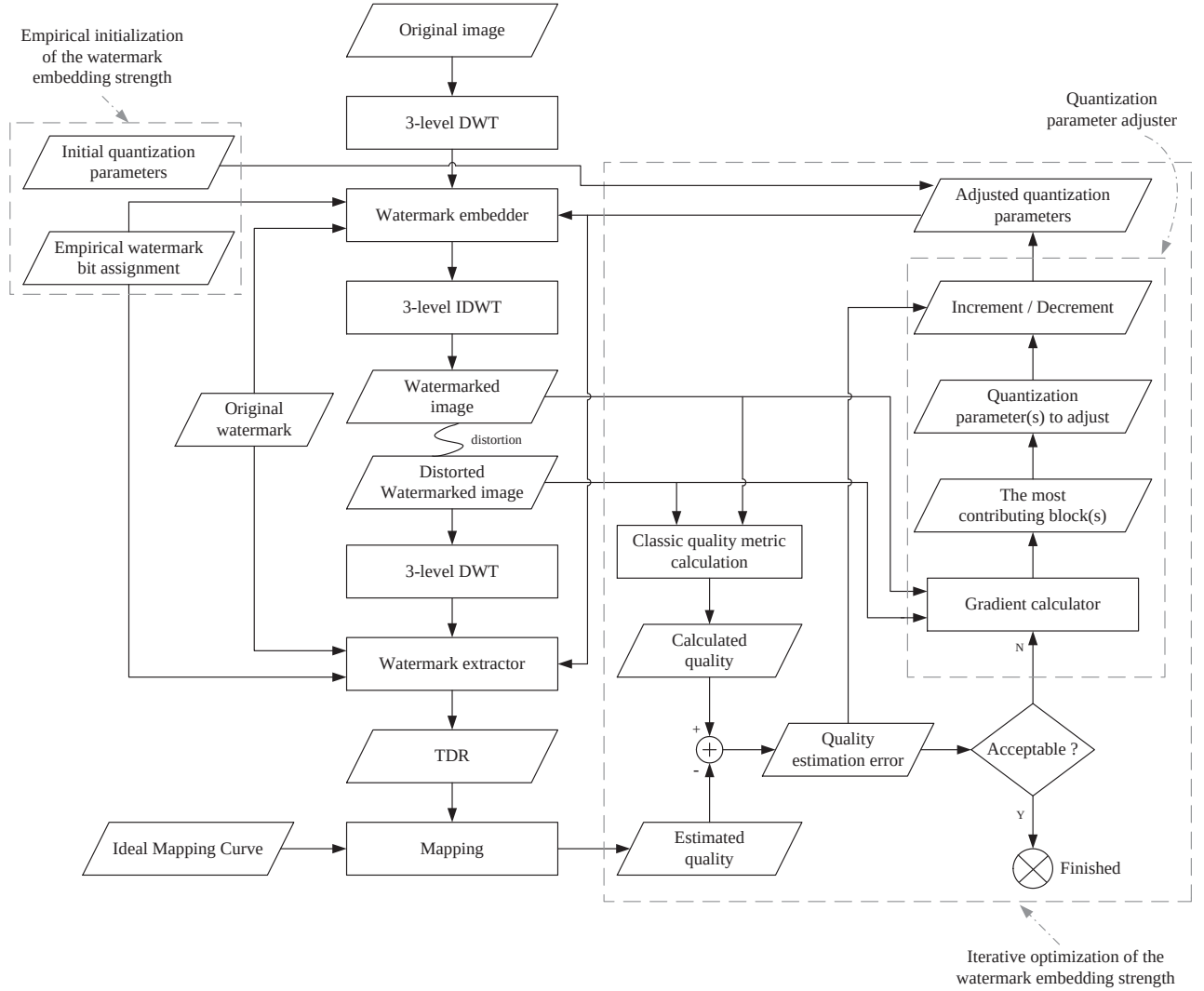


Figure 4.1: The quantization based image watermark embedding process.

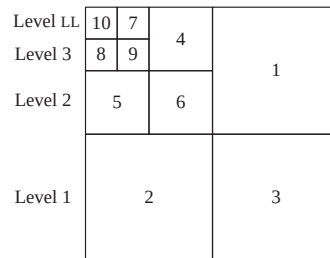


Figure 4.2: Numeration of the 10 DWT decomposed subbands.

- (b) Initially embed the watermark into the selected DWT coefficients using the empirically initialized parameters.
 - (c) Apply inverse 3-level DWT to the watermarked DWT coefficients.
- 2. Obtain the watermarked image.
- 3. Attack the watermarked image to simulate some possible distortions in the transmission procedure.
- 4. Obtain the distorted watermarked image.
- 5. Extract the distorted watermark from the distorted watermarked image.
 - (a) Apply 3-level DWT to the distorted watermarked image.
 - (b) Extract the distorted watermark.
- 6. Evaluate the quality of the distorted image and prepare to minimize the quality evaluation error.
 - (a) Evaluate the degradation of the watermark by calculating the True Detection Rate (TDR) as defined in Equ. (3.1).
 - (b) Evaluate image quality by mapping the calculated TDR onto the “Ideal Mapping Curve”. Note that the quality of the distorted watermarked image comparing to the watermarked image is calculated as well to minimize the quality evaluation error.
- 7. Improve of the quality evaluation accuracy by minimizing the quality evaluation error.

- (a) Calculate the quality evaluation error between the evaluated quality and the calculated quality.
 - (b) Evaluate the quality evaluation error. If it is acceptable, go to Step 8. Otherwise, adjust the quantization parameters according to the quality evaluation error and repeat the above steps with the adjusted quantization parameters.
8. Output the watermarked image and the adjusted quantization parameters.

In the watermark embedding process, a quantization-based method is used to embed watermark into the selected DWT coefficients and this method will be presented in Section 4.1.1.

In the proposed quantization based quality evaluation scheme, the watermark embedding strength is controlled by two factors:

1. The empirical watermark bit assignment which is defined as $A_{wb} = [a_1, a_2, a_3, a_{LL}]$. As described in Section 3.3, a_i indicates the number of watermark bits embedded into each selected subband on DWT level i , where $i \in [1, 2, 3, LL]$. The pyramid structure of the DWT decomposed image is shown in Fig. 4.2. In this scheme, the watermark bit assignment is set up in the empirical initialization step and is fixed for all images.
2. The quantization parameters which is defined as $Qp = [q_1, q_2, q_3, q_4, q_5, q_6, q_7, q_8, q_9, q_{10}]$. q_j is the quantization parameter used for watermark embedding in the DWT decomposed subband j , where $j \in [1, 2, 3, \dots, 10]$. The 10 DWT decomposed subband is numerated in Fig. 4.2. The 10 quantization parameters are empirically initialized in the empirical initialization step and are iteratively optimized to achieve the highest quality evaluation accuracy.

Because the watermark bit assignment is empirically fixed, once the quantization parameters are changed, the overall watermark embedding strength is changed. The quantization parameters are iteratively optimized by testing the degradation characteristics of the cover image under the specific distortions. The watermark embedding strength that results in the minimal quality evaluation error is finally assigned to the cover image. In this case, the optimized watermark embedding strength is categorized as the adaptive watermark embedding strength. Under the same distortion, different textured images have different degradation characteristics. Thus, the adaptive watermark embedding strengths assigned to different images are different. The empirical initialization of the watermark embedding strength and the iterative optimization of the watermark embedding strength will be presented in Section 4.1.2 and Section 4.1.3, respectively.

In this scheme, the link between the degradation of the watermark and the degradation of the cover image is built up and is named the “Ideal Mapping Curve”. The “Ideal Mapping Curve” makes it feasible to quantitatively evaluate the quality of the distorted image by only assessing the degradation of the watermark. The details of the “Ideal Mapping Curve” will be presented in Section 4.2.2.

4.1.1 Quantization method used for watermark embedding

Fig. 4.3 shows the details of the quantization method used for watermark embedding.

Prior to the watermark embedding, the DWT coefficients in one DWT decomposed subband is first sorted in descending order. Then, every DWT coefficient is quantized using Equ. (4.1). By doing so, each coefficient is assigned a binary 1 or 0. The binary bit associated with DWT coefficient is defined as $Q(e)$.

To embed a watermark bit into a target DWT coefficient, the $Q(e)$ associated with

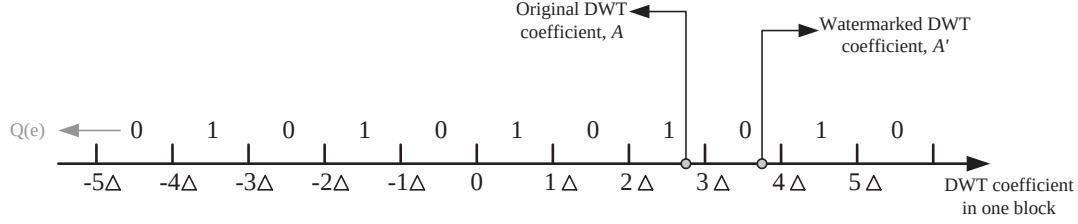


Figure 4.3: The quantization method used for watermark embedding. $\frac{1}{\Delta}$ is the quantization parameter.

the target DWT coefficient is compared with the watermark bit. If they are equal, we do not make any change to the target DWT coefficient. Otherwise, the target DWT coefficient is modified by adding Δ in order to make $Q(e)$ equal to the watermark bit.

$$Q(e) = \begin{cases} 1, & \text{if } \left\lfloor \frac{DWT \text{ coefficient}}{Quantization \text{ parameter}} \right\rfloor \text{ is even} \\ 0, & \text{if } \left\lfloor \frac{DWT \text{ coefficient}}{Quantization \text{ parameter}} \right\rfloor \text{ is odd} \end{cases} \quad (4.1)$$

An example of the watermark embedding is shown in Fig. 4.3. We want to embed a watermark bit 0 into the DWT coefficient A whose $Q(e)$ is 1 . Apparently, the watermark bit is not equal to $Q(e)$. Then, we add Δ to A to change A to A' . The $Q(e)$ of modified DWT coefficient, A' , is 0 and equals to the watermark bit. Thus, the watermark embedding for coefficient A is accomplished and the modified coefficient A' is the watermarked DWT coefficient.

In this scheme, the redundancy of the watermark embedding is 50, which means that each watermark bit is embedded into 50 selected DWT coefficients. In one DWT decomposed subband, the DWT coefficients are sorted in descending order. The first 50 biggest DWT coefficients are used to embed the first watermark bit and the next 50 biggest coefficients are used to embed the second watermark bit and so on. This

strategy is to improve the correct extraction rate of the distorted watermark.

Moreover, Fig. 4.3 also indicates the tolerance of the watermarked coefficients to distortions. If the quantization parameter is very big which results in the Δ is very small, a small variation of the watermarked coefficient has bigger chance to result in the wrong extraction of the watermark bit. In this case, with the fixed watermark bit assignment, the bigger the quantization parameters, the weaker the watermark embedding strength, and vice versa.

4.1.2 Empirical initialization of the watermark embedding strength

Nowadays, PSNR is still the most commonly used metric in the image and video quality evaluation and the JPEG compression is one of the most frequently appeared distortion to digital images. The quality metric PSNR and JPEG compression with quality factors varying from 100 to 10 with a step of -10 are used to find the appropriate initial watermark embedding strength which includes the initial watermark bit assignment and the initial quantization parameters.

The following three criteria are used to choose the initial watermark embedding strength:

1. The embedded watermark should be invisible. With the initial watermark embedding strength, we expect that the PSNR of the watermarked images are higher than 40 dB.
2. With the increasing of the distortion strength, the calculated TDR decreases monotonically with the PSNR values of the distorted images, where TDR is the True Detection Rate of the distorted watermark and is defined in Equ. (3.1).

3. The embedded watermark should be semi-fragile. Under JPEG compression with the quality factor decreasing from 100 to 10, the degraded quality of the compressed watermarked images comparing to the watermarked images usually varies from 55 dB to 20 dB in PSNR. Thus, we expect that the TDR calculated using the initial watermark embedding strength can vary in the possibly biggest range, i.e. from 1 to 0.2, to make the quality evaluation possible. In other words, if TDR varies in a very small range, i.e. from 1 to 0.7, then a small variation of the TDR may cause very big quality evaluation error. Note, because the quality of the distorted images will not be 0 dB in PSNR in this thesis, the lower limit of the calculated TDR is not expected to be 0.

Once the initial watermark embedding strength is selected, the watermark bit assignment, $A_{wb} = [a_1, a_2, a_3, a_{LL}]$, is fixed for all the images and the initial quantization parameters, Qp , will start being optimized.

Based on the research done in [51], the initial watermark bit assignment is set as $A_{wb} = [20, 256, 49, 49]$. The initial quantization parameters are set as $Qp = [80, 80, 80, 80, 80, 80, 80, 80, 80, 80, 80]$ and tested using the initial watermark bit assignment with the quantization parameters varying from 10 to 200 with a step of 10. The parameters that make the proposed scheme best meet the criteria listed above are selected.

With the initial watermark embedding strength, 10 different textured images are tested under JPEG compression. The degradation characteristics of the watermark achieved using the initial watermark embedding strength under JPEG compression is summarized in Fig. 4.4. In the figure, the bars indicate the possible value range of TDR tested using the 10 images. The dot curve is the mean TDR values computed under each quality factor. With the decreasing of the quality factor, the TDR decreases

monotonically.

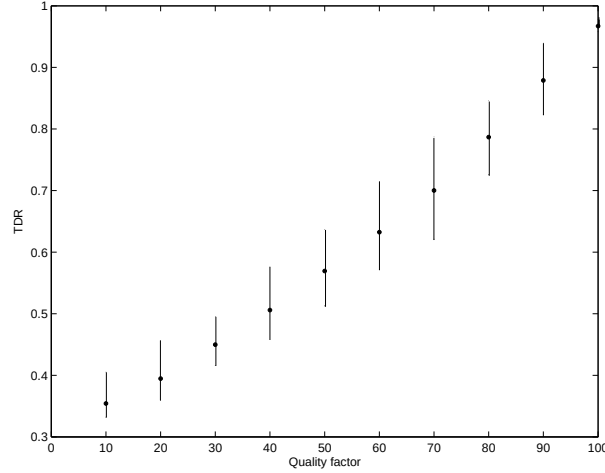


Figure 4.4: The degradation characteristics of the watermark achieved using the initial watermark embedding strength.

On the other hand, the watermark is respectively embedded into image Lena and image Baboon using the initial watermark embedding strength. The watermarked images are compressed using JPEG compression with quality factors varying from 100 to 10. The PSNR between the compressed watermarked images and the watermarked images is calculated as well. The TDR-PSNR curves of image Lena and image Baboon are illustrated in Fig. 4.5. One problem raises from this figure: when TDR is 0.62, the PSNR of Lena is 37 dB and the PSNR of Baboon is 33 dB. It is impossible to use a fixed TDR value to evaluate two different quality values. Therefore, it is desirable if the TDR-PSNR curves can be made as convergent as possible. In the proposed scheme, this goal is accomplished by adjusting the vulnerability of the embedded watermark which is controlled by the watermark embedding strength.

Based on the analysis made above, we use the iterative process shown in Fig. 4.1 to optimize the watermark embedding strength by minimizing the quality evaluation

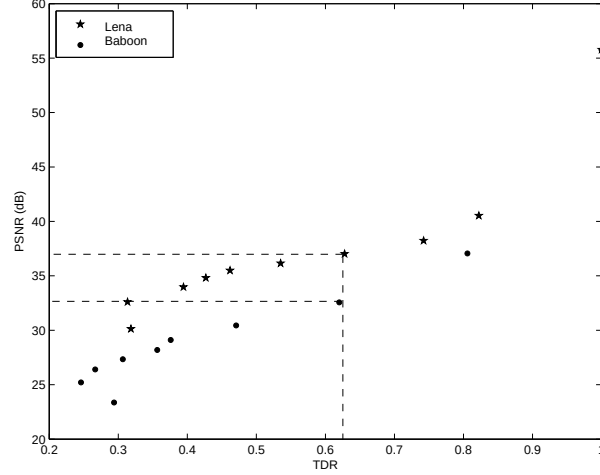


Figure 4.5: The TDR-PSNR curves tested using the initial watermark embedding strength.

error. By doing this, the optimized watermark embedding strengths for different images will be different. The details of the embedding strength optimization will be presented in the next section.

4.1.3 Iterative optimization of the watermark embedding strength

Because the watermark bit assignment is fixed in the proposed scheme, the iterative optimization of the watermark embedding strength is to optimize the quantization parameters. It is to finely improve the accuracy of the quality evaluation by minimizing the quality evaluation error. As shown in Fig. 4.6, The iterative process consists of two components: the quality evaluation error calculator and the quantization parameter adjuster.

The quality evaluation error calculator is to calculate the difference between the quality evaluated using the proposed scheme and the quality calculated using the clas-

sic quality metric, i.e. PSNR or Watson JND. The evaluated quality is obtained by mapping the calculated TDR referring to the “Ideal Mapping Curve”. If the quality evaluation error is larger than a threshold (i.e. 4.5 dB or 4.5 JND for the summation of the quality evaluation error under 9 quality factors from 100 to 20 with a step of -10), the quantization parameter adjuster will automatically adjust the quantization parameter(s) associated with the DWT decomposition subbands that contribute the most to the current quality evaluation error. In Fig. 4.6, the function $f(\bullet)$ is used to empirically estimate the increment or decrement for the quantization parameters that need to be adjusted by evaluating the quality evaluation error.

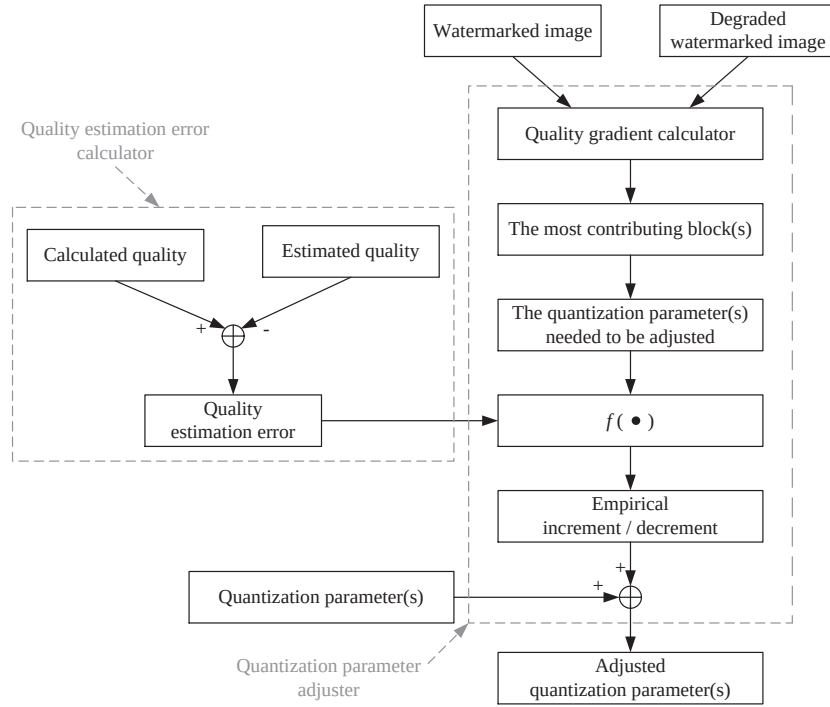


Figure 4.6: The optimization of the quantization parameters.

The contributions of the DWT decomposed subbands are mathematically evaluated using the quality gradient calculation. This procedure is referred as the quality gradient

calculation because it is similar to the gradient computation and a gradient vector points to the direction of the greatest rate of change. The larger the calculated gradient,

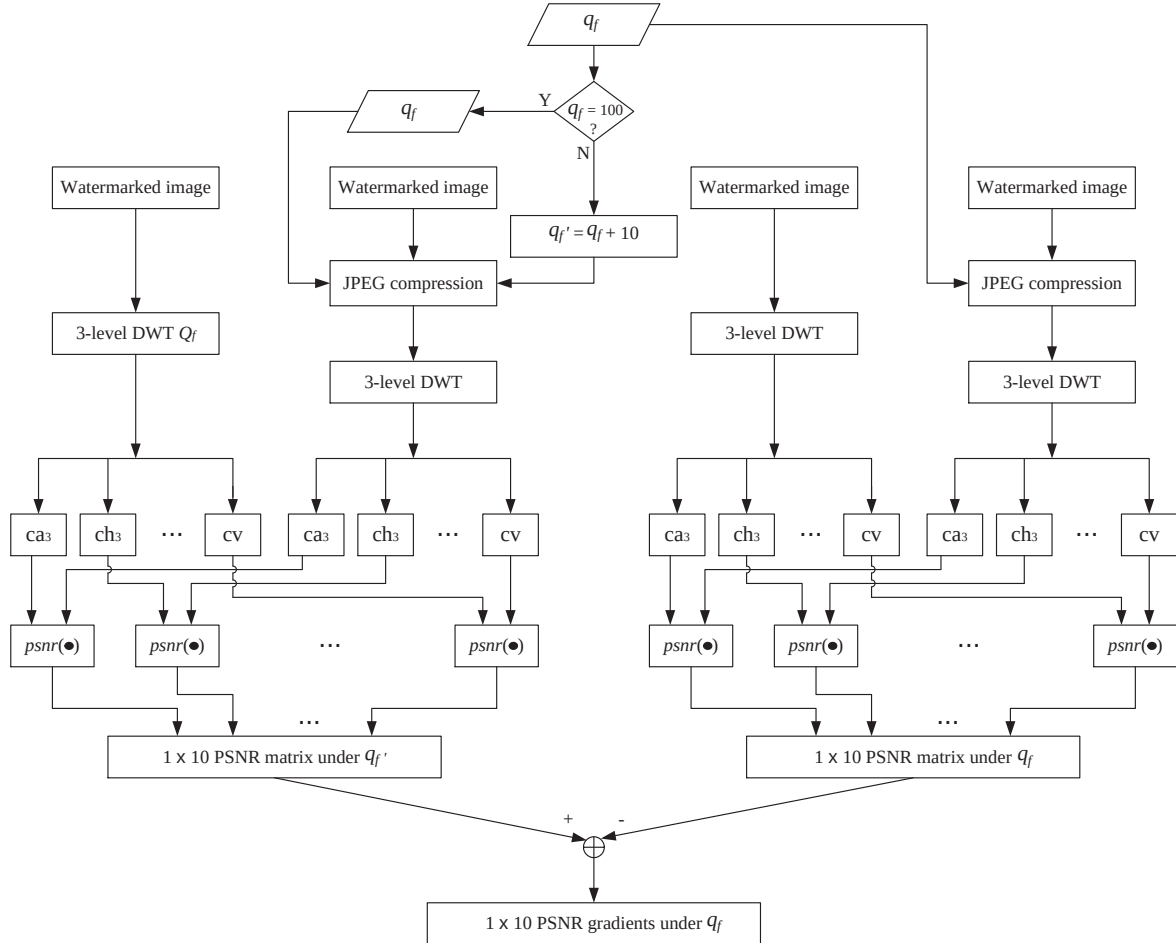


Figure 4.7: The implementation details of the quality gradient calculator.

the more sensitive the subband to distortion. The subband with the largest quality gradient is the one whose quantization parameter has the largest effect on changing the vulnerability of the watermark. Thus, by finding the maximal quality gradient under the current quality evaluation error, we can locate the quantization parameter that needs to be adjusted. Fig. 4.7 shows the implementation details of the quality gradient calculator used for estimating image quality in terms of PSNR under JPEG

compression. It can be also used for other quality metrics, such as wPSNR, Watson JND or SSIM by change the $psnr(\bullet)$ to $wpsnr(\bullet)$, $jnd(\bullet)$ or $ssim(\bullet)$. To maximize the optimization effect, the quantization parameters of the subbands located in the same DWT level as the most contributing subband are adjusted as well.

Fifteen loops are set up for the iterative process. The quantization parameters that result in the minimal quality evaluation error, and the corresponding watermarked image will be output. In all, the iterative optimization of the quantization parameters consists of the following steps:

1. Initialize the number of the iterative loop as 0. Set $Quan$ equals to the 10 initial quantization parameters, Qp .
2. Increase the number of the current iterative loop with 1.
3. Embed the watermark into the cover image using the watermark bit assignment A_{wb} and the quantization parameters $Quan$. Obtain the watermarked image.
4. Compress the watermarked image using JPEG with quality factors $[100 : -10 : 20]$, respectively.
5. Calculate the quality evaluation error, $\Delta Q(q_f)$, between the calculated quality, $Q_C(q_f)$, and the evaluated quality, $Q_E(q_f)$, under the JPEG quality factor q_f which varies from 100 to 20 with a step of -10:

$$\Delta Q(q_f) = Q_C(q_f) - Q_E(q_f) \quad (4.2)$$

Thus, totally 9 quality evaluation errors are computed under the 9 quality factors. Store the 9 quality evaluation errors in array $QError_{1 \times 9}$.

6. For each q_f , calculate the quality gradients of the 10 DWT subbands and find the subband that contributes the most to the current quality evaluation error using the following steps:

- (a) Apply 3-level DWT to the undistorted watermarked image $I^w(k)$ to get the 10 wavelet decomposed subbands $I_i^w(K)$.

$$I^w(k) \longrightarrow I_i^w(K), \quad (i = 1, \dots, 10) \quad (4.3)$$

where, k and K respectively denote the coefficients of the image in spatial domain and in the DWT domain; while i is the number of the wavelet decomposed subband.

- (b) Apply 3-level DWT to the compressed watermarked image $I^{cw}(k, q_f)$ to obtain the corresponding wavelet decomposed subbands $I_i^{cw}(K, q_f)$. The compressed watermarked image is obtained by compressing the watermarked image with a quality factor q_f .

$$I^{cw}(k, q_f) \longrightarrow I_i^{cw}(K, q_f), \quad (i = 1, \dots, 10) \quad (4.4)$$

- (c) Next, calculate the PSNR values between the corresponding DWT decomposed subbands of the watermarked image and the compressed watermarked image. For example, the PSNR between the i^{th} subband of the undistorted watermarked image, $I_i^w(K)$, and the i^{th} subband of the compressed watermarked image, $I_i^{cw}(K, q_f)$, can be calculated as:

$$V_{PSNR}(i, q_f) = f_{PSNR}(I_i^w(K), I_i^{cw}(K, q_f)) \quad (4.5)$$

where, $i = 1, 2, \dots, 10$. V_{PSNR} is the PSNR value calculated between the corresponding DWT subbands using the function f_{PSNR} .

- (d) Calculate the PSNR gradients $G(i, q_f)$ under q_f for all the 10 subbands using Equ. (4.6).

$$G(i, q_f) = \begin{cases} |V_{PSNR}(i, q_f+10) - V_{PSNR}(i, q_f)|, & \text{if } q_f < 100 \\ 0, & \text{if } q_f = 100 \end{cases} \quad (4.6)$$

- (e) For quality factor q_f , find the most influential subband j whose PSNR gradient is the largest among the 10 calculated gradients and find the DWT level $\ell(q_f)$ that the subband j locates in referring to Fig. 4.2.

$$G(j, q_f) = \max\{G(i, q_f)\} \quad (i = 1, \dots, 10) \quad (4.7)$$

Thus, for the 9 quality factors, we achieve 9 most contributing subbands and their corresponding DWT levels. Store these DWT levels in array $\mathfrak{L}_{1 \times 9}$. Recall the array $QError_{1 \times 9}$ generated in Step 5 and $\mathfrak{L}_{1 \times 9}$ generated in Step 6e. Each $\ell(q_f) \in \mathfrak{L}_{1 \times 9}$ corresponds to one $\Delta Q(q_f) \in QError_{1 \times 9}$.

7. Among the 9 quality evaluation errors, respectively calculate the times that the DWT level 1, 2, 3 and LL appear in $\mathfrak{L}_{1 \times 9}$. The most frequently appeared level in $\mathfrak{L}_{1 \times 9}$ is the DWT level that contribute the most to the quality loss of the cover image under JPEG compression and we denote it as ℓ . From $QError_{1 \times 9}$, find the maximal quality evaluation error, ΔQ , that corresponds to ℓ in $\mathfrak{L}_{1 \times 9}$. In the experimental implementation, the quantization parameter(s) associated with all the subband(s) in DWT level ℓ are empirically adjusted by evaluating ΔQ so that the optimization effect of the quantization parameters can be maximized.

8. Empirically estimate the increment or decrement, ΔQ_{uan_ℓ} , for the quantization parameter(s) in level ℓ :

$$\Delta Q_{uan_\ell} = \begin{cases} +10 & +2.0 \leq \Delta Q \\ +5 & +1.5 \leq \Delta Q < +2.0 \\ +3 & +1.0 \leq \Delta Q < +1.5 \\ +1 & +0.3 \leq \Delta Q < +1.0 \\ 0 & -0.3 \leq \Delta Q < +0.3 \\ -1 & -1.0 \leq \Delta Q < -0.3 \\ -3 & -1.5 \leq \Delta Q < -1.0 \\ -5 & -2.0 \leq \Delta Q < -1.5 \\ -10 & \Delta Q < -2.0 \end{cases} \quad (4.8)$$

9. Empirically adjust the adjusted quantization parameter for subband j .

$$Q_{uan'_{j \in \ell}} = Q_{uan_{j \in \ell}} + \Delta Q_{uan_\ell} \quad (4.9)$$

where $Q_{uan_{j \in \ell}}$ is the quantization parameter for subband j in the DWT level ℓ and $Q_{uan'_{j \in \ell}}$ is the adjusted quantization parameter.

10. Update the 10 quantization parameters:

$$Q_{uan'} = [Q_{uan_1}, \dots, Q_{uan'_{j \in \ell}}, \dots, Q_{uan_{10}}] \quad (4.10)$$

11. If the number of current iterative loop is less than 15, $Q_{uan} = Q_{uan'}$. Repeat the quantization parameter optimization from Step 2 using the adjusted quantization parameters, Q_{uan} . If the number of the current iterative loop equals to 15, output the optimized quantization parameters, $Q_{uan'}$ that result in the minimal mean

absolute error of the quality evaluation.

To this end, the watermark embedding strength is completely optimized and the highest quality evaluation accuracy is achieved. Since the watermark embedding strength is optimized by evaluating the quality degradation characteristics of the cover images under distortion, the optimized watermark embedding strength can be categorized as the adaptive watermark embedding strength.

4.2 The quantization based watermark extraction and quality evaluation

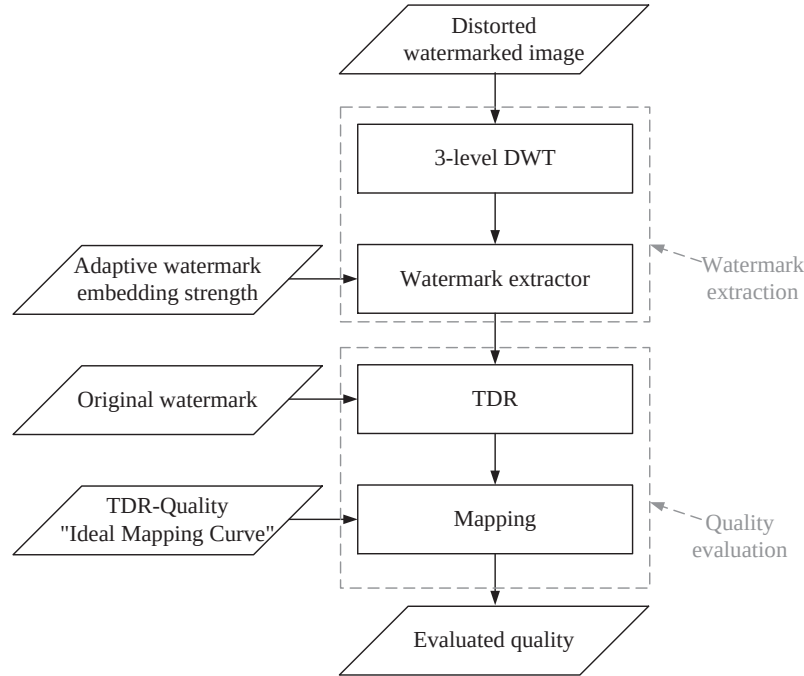


Figure 4.8: The quantization based watermark extraction and quality evaluation process.

The proposed watermark extraction and quality evaluation scheme is shown in Fig.

4.8. The adaptive watermark embedding strengths achieved at the sender side need to be transmitted to the receiver side for the use of watermark extraction. The True Detection Rate (TDR) of the distorted watermark is defined in Equ. (3.1) to assess the degradation of the watermark. As found in the experiments, the calculated TDR decreases monotonically with the increasing of the distortion strengths [51].

4.2.1 The watermark extraction

In the quantization based watermark extraction scheme, the distorted coefficients in one DWT decomposed subband are sorted in descending order. The biggest 50 coefficients are used to extract the first watermark bit. The next 50 biggest coefficients are used to extract the second watermark bit and so on.

To extract a watermark bit, the $Q(e)$ associated with the 50 distorted watermarked DWT coefficients will be first calculated using Equ. (4.1). Because the DWT coefficients of the distorted image may be changed during the transmission, the newly calculated $Q(e)$ may be different from those calculated using the original DWT coefficients. Equ. (4.11) shows the details of the watermark bit extraction.

$$w_e = \begin{cases} 1, & \text{if } N_1 \geq N_0 + \text{const} \\ 0, & \text{if } N_0 \geq N_1 + \text{const} \\ 2, & \text{otherwise} \end{cases} \quad (4.11)$$

where, w_e is the current watermark bit to extract; N_1 is the number of the extracted 1s; N_0 is the number of the extracted 0s; and $N_1 + N_2 = 50$; const is an empirical parameter and equals to 8 [81]. So, if the extracted 1s are 8 bits more than the extracted 0s, the current watermark bit is extracted as 1. On the contrary, if the extracted 0s are 8 bits more than the extracted 1s, the current watermark bit is extracted as 0. Otherwise,

the current watermark bit is extracted as a wrong bit.

4.2.2 The generation of “Ideal Mapping Curve”

As presented in Section 3.4, the “Ideal Mapping Curve” is mathematically expressed using a TDR-Quality curve and experimentally defines the mapping relationship between the calculated TDR and the degraded quality of the distorted images. It is generated using the generator shown in Fig. 3.9 by testing 20 different watermarked images, in which the watermark is embedded using the initial watermark embedding strength. In the figure, the proposed quantization based watermark embedder and extractor are used for the watermarking processes. As presented in Section 4.1.2, the initial watermark embedding strength is fixed for all the test images. The “Ideal Mapping Curve” can be generated using more images as well. However, it is found in the experiments that it does not make significant difference in the resulted “Ideal Mapping Curve” by testing more images. In the proposed quantization based quality evaluation scheme, the Quality can be in terms of any desired quality metric, such as PSNR, wPSNR, Watson JND or SSIM. For a specific quality metric under one distortion, one “Ideal Mapping Curve” needs to be generated.

An example is shown in Fig. 4.9 to illustrate how we generate the “Ideal Mapping Curve” for estimating image quality in terms of Watson JND under JPEG compression. The grey dashed curves are the 20 TDR-JND curves obtained by testing the 20 watermarked images. Each image is compressed with JPEG quality factors [100, 90, 80, ..., 20]. Therefore, there are nine points on each grey TDR-JND curve and each point represents a correspondence between TDR and the calculated JND. In the figure, the blue star curve is the TDR-JND “Ideal Mapping Curve” which is computed using the local averaging presented in Section 3.4.

In this proposed scheme, the “Ideal Mapping Curve” is used as the reference for the optimization of the quantization parameters. Once the tested PSNR-TDR curve of any image is adjusted to converge to the “Ideal Mapping Curve”, the image quality after compression with an arbitrary quality factor can be evaluated with high accuracy.

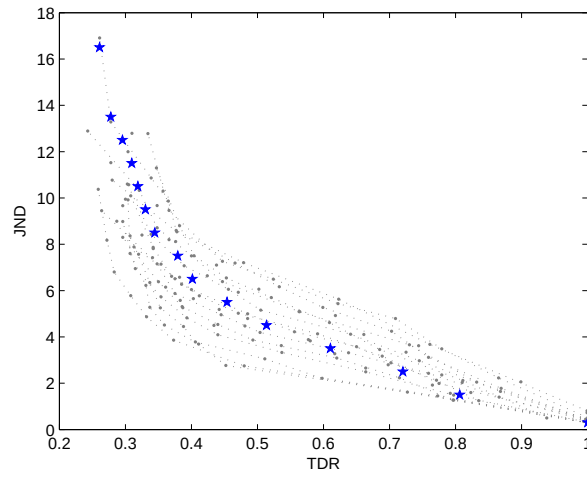


Figure 4.9: Generation of the “Ideal Mapping Curve” using a fixed watermark embedding strength in the quantization based scheme.

The principle of the quantization parameter optimization is shown in Fig. 4.10. Suppose that the TDR-PSNR “Ideal Mapping Curve” is generated as the solid curve shown in the figure. The point (A, QC) is obtained by testing an image using a specific quality factor under JPEG compression, where A is the calculated TDR and QC is the calculated quality of the compressed watermarked image. However, by mapping A referring to the “Ideal Mapping Curve”, the quality of the compressed image is evaluated as QE . In this case, the quality evaluation error, $\Delta Q = QC - QE$, is resulted. To minimize the quality evaluation error, we optimize the watermark embedding strength to adjust the calculated TDR from A to B . By doing so, the quality of the distorted image can be accurately evaluated as QC and the quality evaluation error is reduced

to 0.

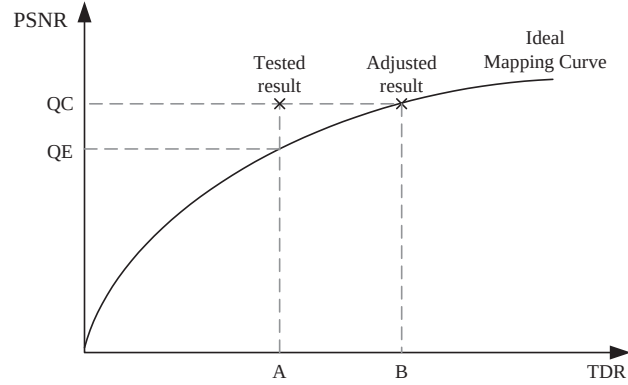


Figure 4.10: The principle of the quantization parameter optimization.

The optimization of the quantization parameter is further illustrated on multiple tested TDR-Quality points as shown in Fig. 4.11. In the figure, the tested curve is obtained by testing an image under JPEG compression with different quality factors. The shaded area between the tested curve and the “Ideal Mapping Curve” indicates the quality evaluation error. Thus, the optimization of the watermark embedding strength is to minimize the shaded area. If the shaded area is eliminated, the theoretically highest accuracy of the quality evaluation is achieved. In other words, we expect to converge the tested TDR-Quality curves to the “Ideal Mapping Curve”. The more convergent the tested curves to the “Ideal Mapping Curve”, the higher the quality evaluation accuracy.

The advantages of the “Ideal Mapping Curve” in the proposed quantization based quality evaluation scheme are listed as follows:

1. The “Ideal Mapping Curve” makes it possible to evaluate image quality using the degradation of the distorted watermark, TDR.

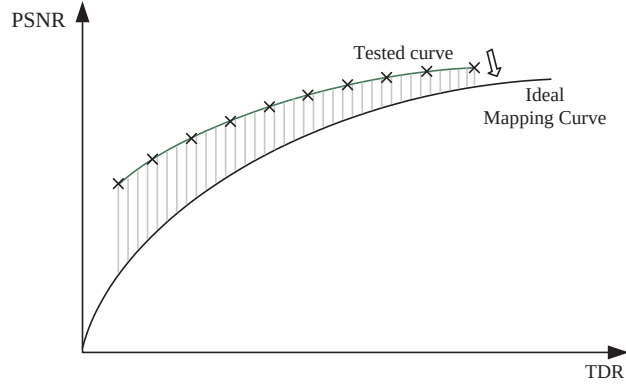


Figure 4.11: An example of the quantization parameter optimization on multiple tested quality points.

2. The “Ideal Mapping Curve” provides the most important guidance to improve the accuracy of the quality evaluation. The quality evaluation accuracy is maximized by minimizing the quality evaluation error. The quality evaluation error is minimized by converging the TDR-Quality curves tested on the cover images to the “Ideal Mapping Curve”.
3. The “Ideal Mapping Curve” makes the optimization of the watermark embedding strength faster and easier. As mentioned previously, the “Ideal Mapping Curve” accommodates the degradation characteristics of 20 different textured images by statistically averaging the 20 tested TDR-Quality curves. Thus, the “Ideal Mapping Curve” consists of a set of statistical mean value points and is able to make the needed adjustment less.

4.2.3 The watermarking based quality evaluation

The quality of the distorted image is evaluated by mapping the calculated TDR to a quality value by referring to the “Ideal Mapping Curve”. The value of the calculated

TDR could possibly lie between two neighboring TDR values on the “Ideal Mapping Curve” as illustrated in Fig. 4.12. Through experiments, it is found that, by increasing of distortion degrees, the calculated TDR value decrease monotonically [51]. In this case, linear interpolation is used to compute quality value. An example of the quality evaluation is presented in the following.

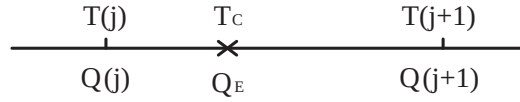


Figure 4.12: Linear interpolation.

In Fig. 4.12, $(T(j), Q(j))$ and $(T(j + 1), Q(j + 1))$ are two adjacent points on the “Ideal Mapping Curve”, where $T(\bullet)$ is the TDR value and $Q(\bullet)$ is the quality value in terms of PSNR, wPSNR, JND or SSIM. T_C is the calculated TDR of any image and Q_E is the evaluated quality which can be obtained using Equ. (4.12).

$$Q_E = Q(j) + \frac{T_C - T(j)}{|T(j + 1) - T(j)|} \times |Q(j + 1) - Q(j)| \quad (4.12)$$

4.3 Experimental results and accuracy evaluation

4.3.1 Summary of experiments

The proposed adaptive watermarking and quantization based image quality evaluation scheme is tested under JPEG compression in terms of PSNR, wPSNR, Watson JND and SSIM, respectively. Four sets of experiments are implemented. The quantization parameters of each image are optimized under 9 quality factors (from 100 to 20 with a step size of -10). To demonstrate the effectiveness of the iterative optimization process,

the proposed scheme is tested with a number of quality factors that are different from those used in the quantization parameter optimization.

In each set of experiment, 100 images are tested. For the quality evaluation purpose, we compress each image with quality factors from 100 to 20 with a step size of -5 . The degraded quality of the distorted image is evaluated by mapping the calculated TDR referring to the “Ideal Mapping Curve”. Therefore, a total of 1700 quality points are tested in each set of experiment. The experimental results show a high correlation between the evaluated quality values and the calculated quality values, which indicates the high accuracy of the proposed scheme.

The wPSNR model discussed in [1] and the SSIM metric presented in [4] are used for the image quality calculation in terms of wPSNR, and SSIM, respectively. The software Dctune2.0 [82] is used for the image quality calculation in terms of Watson JND.

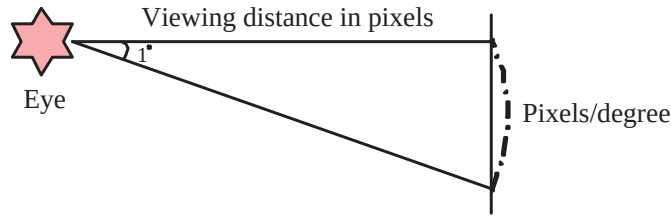


Figure 4.13: The default parameters of Dctune2.0.

The default parameters of Dctune2.0 shown in Fig. 4.13 are:

1. The resolution of the computer is defaulted as the standard screen resolution 72 dpi.
2. Both the horizontal and vertical pixels density are 32 pixels/degree.

3. The viewing distance from the human eye to the screen is calculated as 25.333 inches.

As presented in Section 3.5, the Mean Absolute Error (MAE) and Pearson correlation coefficient are used to evaluate the accuracy of the experimental results, respectively.

4.3.2 The original images and the original watermark

The original grey images in our image library are shown in Appendix A. Each image is 512×512 in size and the image textures include portraits, plants, animals, animations, sceneries, buildings and crowd. For each set of experiment, the first 100 images in our image library are tested.

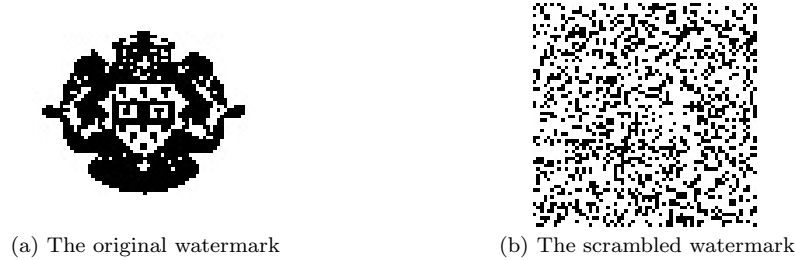


Figure 4.14: The 32×32 original watermark.

The original binary watermark is shown in Fig. 4.14(a) and is 32×32 in size. For the purpose of security, the watermark to be embedded is scrambled and is shown in Fig. 4.14 (b).

As mentioned in Section 4.1.2, the initial watermark embedding strength is set up as:

1. The initial watermark bit assignment $A_{wb} = [0.0195, 0.25, 0.0479, 0.0479]$.

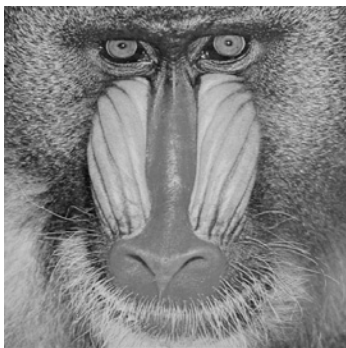
2. The initial quantization parameters $Qp = [80, 80, 80, 80, 80, 80, 80, 80, 80, 80]$.

The watermark embedding effect is illustrated in Fig. 4.15. Fig. 4.15 (a), (c) and (e) are three original images with the most high frequency components in image Baboon and the least high frequency components in image White flower. Fig. 4.15 (b), (d) and (f) are the watermarked images obtained with the optimized watermark embedding strengths which are listed with the corresponding watermarked images. As analyzed in Section 4.1.1, the bigger the quantization parameters, the weaker the watermark embedding strength, and vice versa. Thus, image Baboon has the weakest watermark embedding strength and the quality of the watermarked image is 42.8346 dB in PSNR. The image White flower has the strongest watermark embedding strength and the quality of the watermarked image is 38.4972 dB in PSNR. These results experimentally proved the theoretical analysis made in Section 3.3.

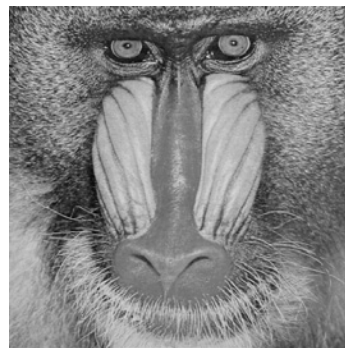
4.3.3 Experimental results

In developing the adaptive watermarking and quantization based image quality evaluation scheme, we have more concerns about the feasibility of using the watermarking based metric to evaluate image quality in terms of the existing quality metrics. In the current situation, only the quality degradation caused by the JPEG compression is considered in the accuracy evaluation.

In the experiments, the “Ideal Mapping Curve” generated for the image quality evaluation in terms of PSNR, wPSNR, Watson JND and SSIM under JPEG compression are shown in Fig. 4.16 (a)-(d), respectively. All these “Ideal Mapping Curve” are generated by testing the first 20 images in the image library under JPEG compression with quality factors varying from 100 to 20 with a step of -10. For image compression,



(a) The original image Baboon.



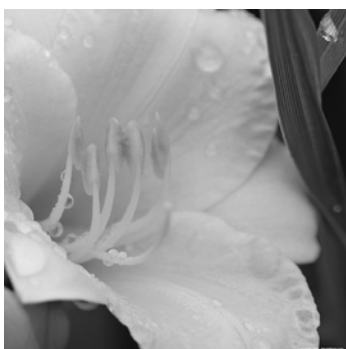
(b) The watermarked image Baboon. PSNR = 42.8346 dB. The optimized quantization parameters $Qp = [90, 90, 90, 110, 110, 110, 80, 80, 80, 80]$.



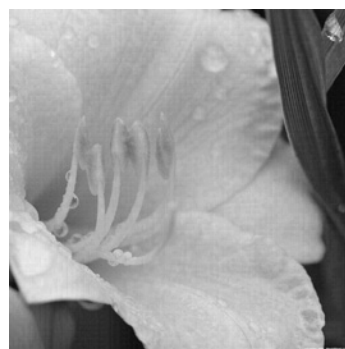
(c) The original image Barbara.



(d) The watermarked image Barbara. PSNR = 40.5614 dB. The optimized quantization parameters $Qp = [65, 65, 65, 65, 65, 65, 60, 60, 60, 60]$.



(e) The original image White flower.



(f) The watermarked image White flower. PSNR = 38.4972 dB. The optimized quantization parameters $Qp = [15, 15, 15, 19, 19, 19, 24, 24, 24, 20]$.

Figure 4.15: Illustration of the effect of watermark embedding.

normally the smaller the JPEG quality factor, the stronger the compression strength and the worse the image quality after compression. Consequently, the calculated quality will be lower in terms of PSNR, wPSNR, or SSIM and will be higher in terms of JND.

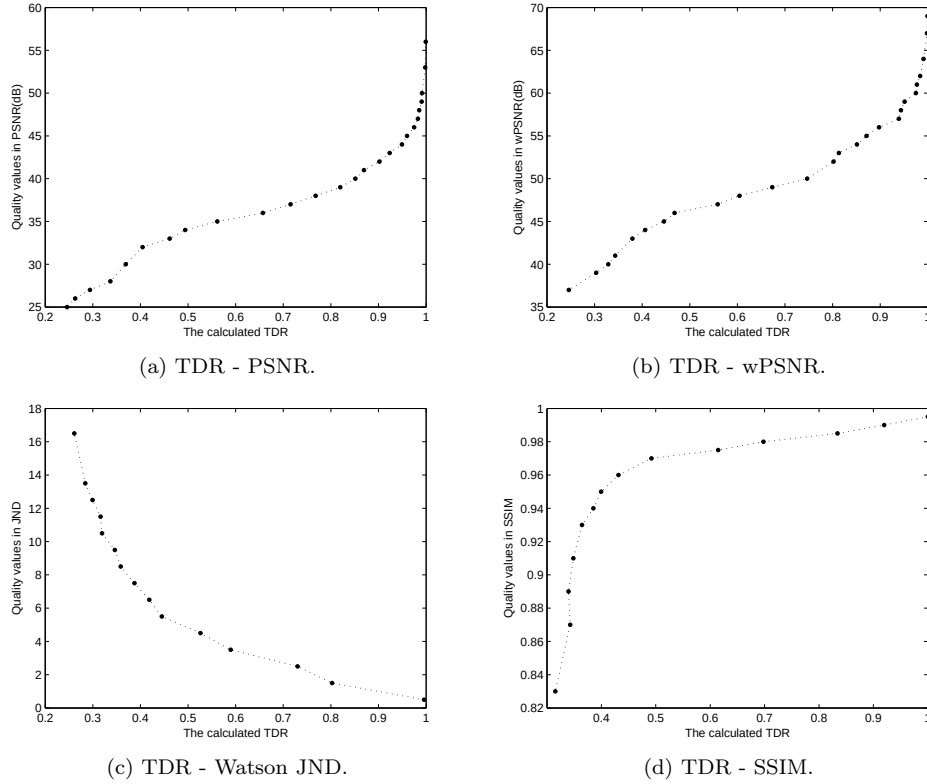


Figure 4.16: The “Ideal mapping curves” generated for image quality evaluation in terms of PSNR, wPSNR, Watson JND and SSIM under JPEG compression.

In Fig. 4.16, the horizontal axis is the TDR values calculated using the distorted watermark and the original watermark. The vertical axis is the quality values calculated using the distorted watermarked images comparing to the undistorted watermarked images. Thus, in Fig. 4.16 (a), the quality values under the quality factor of 100 are higher than those presented in Fig. 4.15.

As shown in Fig. 4.16 (a) and (b), when the quality factor is 100, the quality value is 56 dB in PSNR and 69 dB in wPSNR. With the increasing of the compression strength,

the “Ideal Mapping Curves” generated for quality evaluation in terms of PSNR and wPSNR decrease monotonically. In Fig. 4.16 (c) and (d), when the calculated TDR is smaller than 0.4, the slope of the “Ideal Mapping Curve” becomes much higher. In Fig. 4.16 (c), when the TDR drops from 0.4 to 0.25, the quality value increases from 6 JND to 17 JND. In Fig. 4.16 (d), when the TDR drops from 0.43 to 0.3154, the SSIM value decreases from 0.96 to 0.83. As shown in the figure, when the quality factors varies from 100 to 20, the quality value in SSIM only changes from 1 to 0.83, which indicates that the drop from 0.96 to 0.83 is quite big. In this case, a small variation in TDR will result in relatively bigger evaluation error in JND or SSIM. Thus, with these analysis, it can be empirically predicted that, when quality is lower than 6 JND or 0.96 in SSIM, the accuracy of the experimental results will be relatively lower. In Fig. 4.16 (c), with the increasing of the compression strength, the TDR values drops monotonically and the quality value in JND increases monotonically. In Fig. 4.16 (d), with the increasing of the compression strength, the TDR and SSIM values drop monotonically.

The experimental results of image quality evaluation in terms of PSNR, wPSNR, JND and SSIM are shown in Fig. 4.17 and Fig. 4.18. Fig. 4.17 and Fig. 4.18 (a), (c) are the results tested using the initial watermark embedding strength. The experimental results shown in Fig. 4.17 and Fig. 4.18 (b), (d) are tested using the optimized watermark embedding strengths. In the figures, the x -axis is the evaluated quality achieved using the proposed quantization based quality evaluation scheme. The y -axis is the quality values calculated for the compressed watermarked images comparing to the undistorted watermarked images in terms of PSNR, wPSNR, JND or SSIM. The points in the figures are the experimental results and are called the quality points. The solid line with a 45-degree angle is the matching line which indicates the accuracy of the quality evaluation. The more convergent the quality points to the solid line, the higher

the accuracy of the quality evaluation, and vice versa. If all of the quality points fall on the solid line, the evaluated quality equals to the calculated quality and the highest accuracy is achieved.

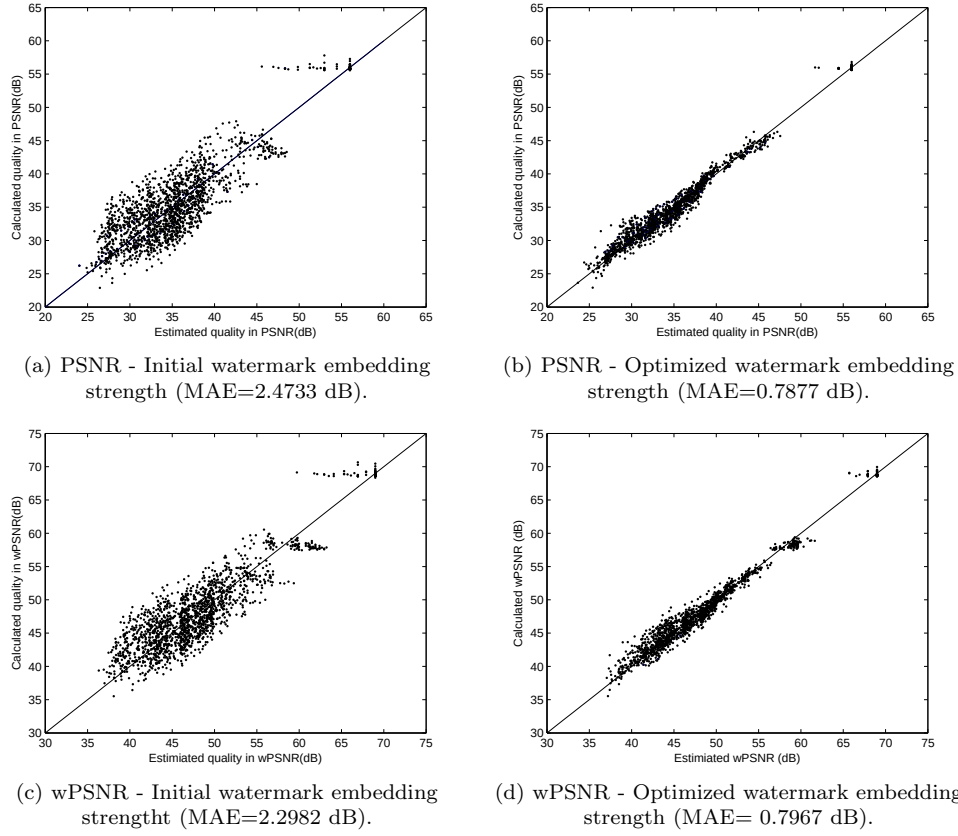


Figure 4.17: Image quality evaluation in terms of PSNR and wPSNR achieved using both the initial watermark embedding strength and the optimized watermark embedding strengths.

Fig. 4.17 and Fig. 4.18 also show the optimization effects of the watermark embedding strength. The quality points achieved using the optimized watermark embedding strengths are much more convergent to the matching line than those tested using the initial watermark embedding strength. In Fig. 4.18 (b) and (d), when the evaluated quality is higher than 8 JND or lower than 0.96 SSIM, the quality points become more

scattering. This experimentally proved the theoretical analysis made about the slope change of the “Ideal Mapping Curves” in Fig. 4.16 (c) and (d).

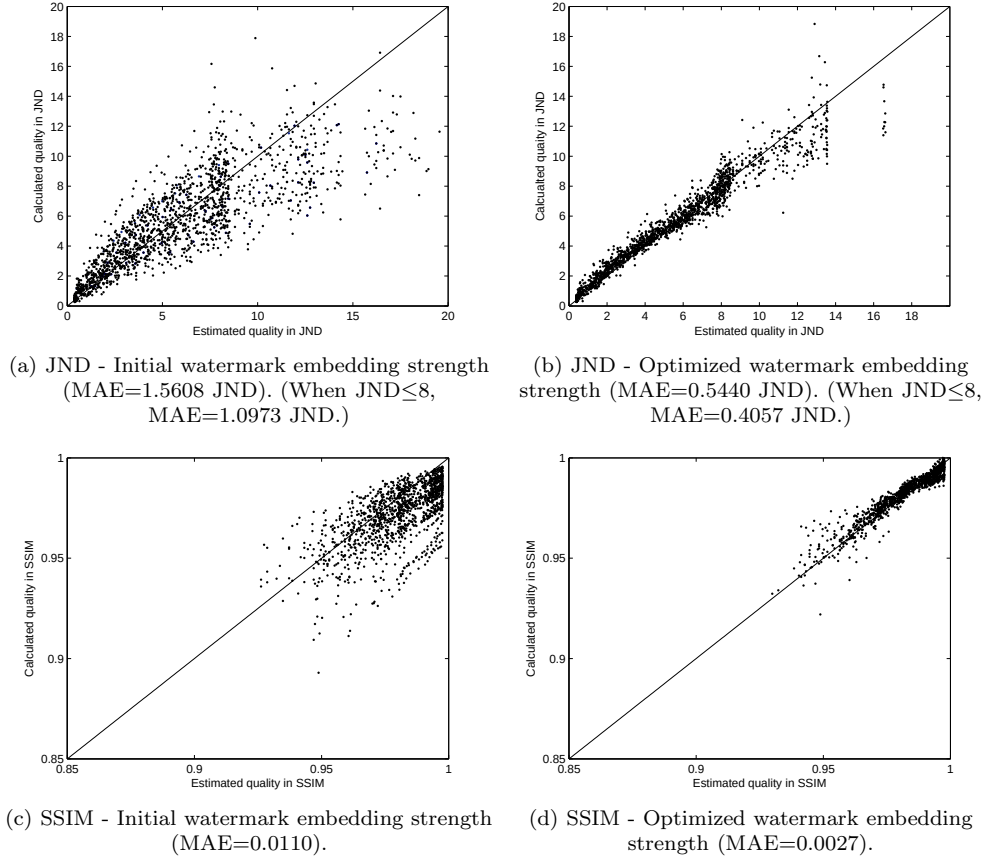


Figure 4.18: Experimental results of image quality evaluation in terms of Watson JND and SSIM achieved using both the initial watermark embedding strength and the optimized watermark embedding strengths.

The accuracy of the experimental results tested with the initial watermark embedding strength and the optimized watermark embedding strengths is evaluated in both MAE using Equ. (3.21) and Pearson correlation using Equ. (3.22). The evaluated accuracy is listed in Table 4.1 and Table 4.2, respectively.

The evaluated accuracy clearly shows that, with the iterative optimization process, the MAEs and the Pearson correlations are greatly improved, which indicates that

Table 4.1: Summary of the quality evaluation accuracy in MAE.

	Initial watermark embedding strength	Optimized watermark embedding strength
PSNR (dB)	2.4733	0.7877
wPSNR (dB)	2.2982	0.7967
JND	1.5608	0.5440
SSIM	0.0110	0.0027

the watermark embedding strength is adaptively assigned to the cover images referring to the quality degradation characteristics of the cover images. Therefore, it can be concluded that, with the optimized watermark embedding strengths, the proposed adaptive watermarking and quantization based image quality evaluation scheme can accurately evaluate image quality in terms of PSNR, wPSNR, JND and SSIM under JPEG compression without accessing the original images.

Table 4.2: Summary of the quality evaluation accuracy in Pearson correlation coefficient.

	Initial watermark embedding strength	Optimized watermark embedding strength
PSNR (dB)	0.8887	0.9872
wPSNR (dB)	0.9151	0.9876
JND	0.8273	0.9714
SSIM	0.6582	0.9583

4.3.4 Advantages of the proposed quantization based scheme

With the experimental results, the advantages of the proposed quantization based image quality evaluation scheme can be summarized as follows:

1. The proposed scheme proved that it is theoretically practical and experimentally feasible to develop a watermarking based quality evaluation scheme.

2. The introduction of the “Ideal Mapping Curve” makes it possible to relate the degradation of the binary watermark to the degradation of the cover image and created a new way to assign adaptive watermark embedding strength to a cover image referring to its degradation characteristics under distortions.
3. The proposed scheme can be used to evaluate image quality in terms of PSNR, wPSNR, JND and SSIM with high accuracy.
4. It formed important and solid basis for our current and future research work.

4.4 Summary

In this chapter, we present the image quality evaluation scheme using quantization based watermarking. This scheme is implemented based on the framework proposed in Chapter 3. The watermark is embedded with the adaptive embedding strength which is iteratively optimized by assessing the degradation characteristics of the cover image under JPEG compression. For different images, the watermark embedding strengths are different. With the adaptive watermark embedding strength, the embedded watermark has similar degradation characteristics as the cover image. The “Ideal Mapping Curve” makes it feasible to quantitatively evaluate the image quality by assessing the degradation of the watermark. With the optimized watermark embedding strength, the accuracy of the quality evaluation is maximized.

The quantization based scheme works with high accuracy under JPEG compression. The relatively high computational complexity caused by the iterative process gives us the inspiration to further develop a more efficient watermarking based quality evaluation scheme according to the proposed framework. With this motivation, the quality

evaluation scheme using the SPIHT tree structure and HVS based watermarking is proposed in the following chapters.

Chapter 5

Image quality evaluation using SPIHT tree structure and HVS based watermarking

5.1 The motivation of using tree structure

As presented in Chapter 4, the adaptive watermarking and quantization based quality evaluation scheme can accurately evaluate image quality under JPEG compression using adaptive watermark embedding strengths which are iteratively optimized by assessing the degradation characteristics of the images. In our research, a couple limitations of the quantization based quality evaluation scheme are found and summarized as follows:

1. The computational efficiency. The iterative process used to optimize the watermark embedding strength is set up as 15 loops and it needs an average of 12 to 13 recursive loops to find the optimal watermark embedding strengths for the test images by iteratively minimizing the quality evaluation error. This iterative

process provides high accuracy to the quality evaluation. However, it introduced relatively high computational complexity to the quantization based scheme which makes it less suitable for some applications like video quality evaluation.

2. The quality degradation caused by the watermark embedding process. As shown in Fig. 4.15, the images with different frequency distributions are assigned different watermark embedding strengths. The more high frequency components contained in the image, the weaker the watermark is embedded, and vice versa. The quality of the watermarked images in Fig. 4.15 are varying from 42.8 dB to 38.5 dB in PSNR. The quality of the 100 watermarked images is about 40 dB in average.

In our researches, it is desirable to keep the accuracy of the quality evaluation achieved in Section 4.3.3 while improving the computational efficiency and it is desirable to greatly reduce the quality degradation caused by the watermark embedding process. Furthermore, it is expected that the watermarking based quality evaluation scheme can work under more distortions.

In light of the discussions above, we propose the adaptive watermarking and tree structure based quality evaluation scheme in this chapter. In Chapter 4, the DWT domain is proved to be a good choice for the watermarking based quality evaluation by taking advantages of both the spatial and frequency information of the cover image. Three level DWT decomposition is applied to the cover image so that enough frequency information of the cover image can be obtained while the computational complexity can be kept relatively low. Recently, the Set Partitioning in Hierarchical Trees (SPIHT) has become one of the most popular image and video coding method because of its efficiency which is accomplished by exploiting the inherent similarities across the sub-

bands in the wavelet decomposed image. The DWT and SPIHT together provide a good summarization of image local region characteristics which is important for adaptive watermark embedding. In the proposed tree structure based quality evaluation scheme, all the correlated DWT coefficients across the subbands are grouped together using the SPIHT tree structure. The DWT decomposed image is further decomposed into a set of bitplane images as shown in Fig. 5.1. In this case, each DWT coefficient is

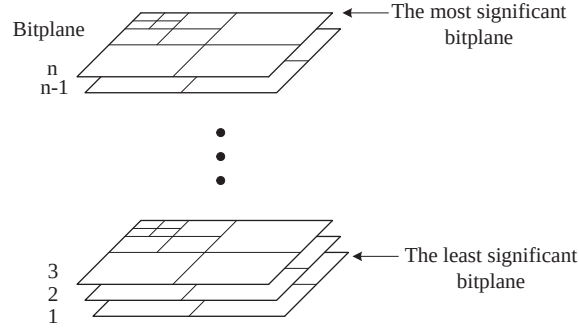


Figure 5.1: Illustration of the bitplane decomposed DWT image.

decomposed into a sequence of binary bits. The binary watermark bits are embedded into the selected bitplanes of the selected DWT coefficients of the selected trees. The HVS masking is used to guide the bitplane selection. As found in Section 3.3, the higher frequency DWT subbands are more sensitive to distortions, and vice versa. The less significant bitplanes are more sensitive to distortions, and vice versa. Therefore, in the proposed tree structure based scheme, the robustness of the watermark is controlled by two factors:

1. The percentages of the watermark bits embedded into the three DWT levels, respectively. We name this factor as the empirical watermark bit assignment, A_{wb} .

2. The selection of the bitplanes for the watermark embedding.

Thus, for different textured images, the watermark embedding strengths are different. In one image, for different selected trees, the watermark embedding strengths are different.

In the proposed schemes, the pixels are denoted using the Matlab coordinates.

5.2 The tree structure based watermark embedding process

The tree structure and HVS based image watermark embedding process is shown in Fig. 5.2. The proposed scheme includes the watermark embedding procedure, the watermark pre-processing procedure and the image pre-analysis procedure. The watermark embedding procedure consists of the following three steps:

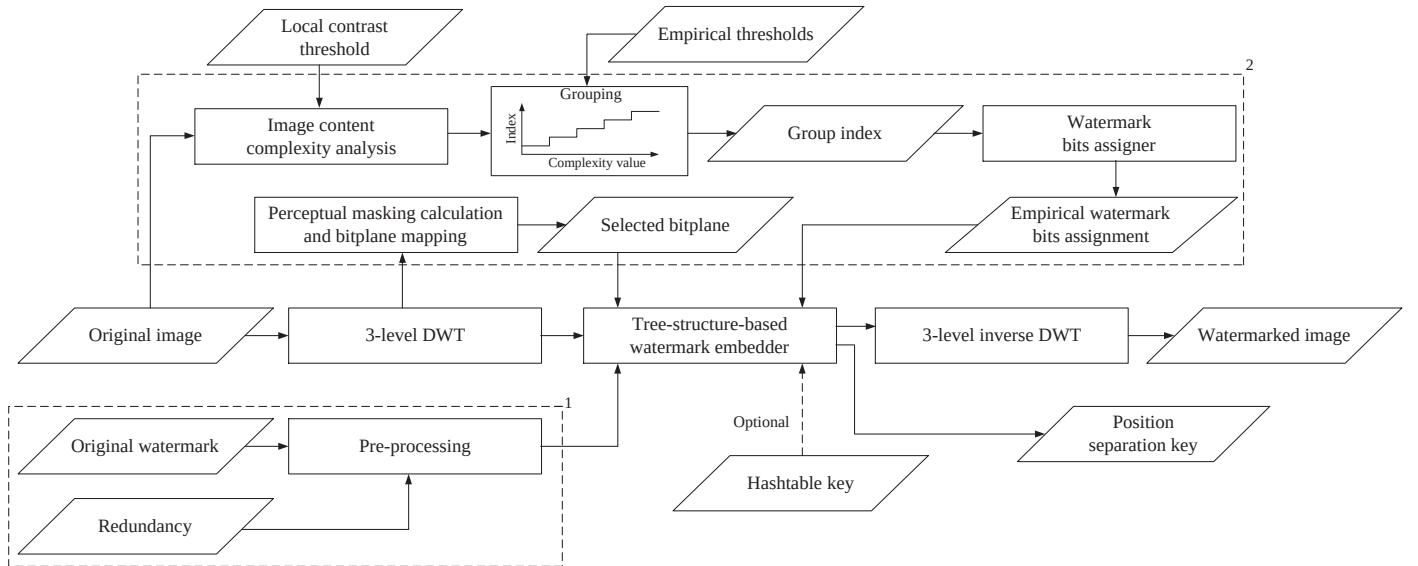


Figure 5.2: The SPIHT tree structure and HVS based image watermark embedding process.

1. Apply 3-level Discrete Wavelet Transform to the original image to achieve the DWT decomposed image. The 3-level DWT decomposed blocks are denoted as those shown in Fig. 3.5.
2. Embed the watermark with the adaptive embedding strength using the tree structure based watermark embedder which will be presented in Section 5.4 in detail. The output of the watermark embedder is the watermarked DWT image.
3. Apply 3-level inverse DWT to the watermarked DWT image to achieve the watermark image.

The watermark pre-processing procedure is shown in the grey-dashed block with number 1 on its upper right corner in Fig 5.2. In this step, the original watermark is re-organized column by column into a sequence. We denote the length of the original watermark sequence as len . Then the original watermark sequence is repeated *Redundancy*-1 times to achieve the redundant watermark sequence which is $Redundancy \times len$ bits long. Finally, the watermark sequence is scrambled and input into the watermark embedder. The introduction of the redundancy aims to increase the probability of the correct watermark bit extraction at the receiver side. The scrambling of the watermark sequence is to increase the security of the watermark embedding. In the proposed scheme, we set $Redundancy = 3$.

The image pre-analysis procedure is shown in the grey-dashed block with number 2 on its upper right corner, which includes the analysis of the original image and the analysis of the DWT decomposed image. This procedure is used to estimate the watermark embedding strength and will be presented in Section 5.4 in detail.

The position separation key is used to locate the positions for watermark embedding. The Hashtable key is an optional input and can be used to secure the watermark

embedding.

5.3 The tree structure based watermark embedder

The tree structure based watermark embedder is designed to embed the binary watermark bits into the selected bitplanes of the selected DWT coefficients of the selected trees. In the DWT decomposed image, all the correlated DWT coefficients across the three detail subbands and the LL subband are grouped into trees. In the proposed scheme, the tree structure based watermark embedder has three functions: forming the tree structure, selecting the trees and the DWT coefficients for the watermark embedding, and embed the binary watermark bits into the selected bitplanes of the selected coefficients. All these functions will be presented in the following three subsections, respectively.

5.3.1 The formation of the tree structure

The tree structure is formed by categorizing the DWT coefficients with inherent similarities across all the subbands. These correlated coefficients build up the parent-descendants relationship and form a tree [83][84]. The descendants of a coefficient include all its children, grandchildren, great-grandchildren, etc. An example of the tree formation is shown in Fig. 5.3. In the LL block, $1/4$ DWT coefficients with both odd horizontal and odd vertical coordinates have no descendants. Only the coefficients with at least one even coordinate have descendants and we define these coefficients as the roots of the trees. For the DWT coefficients with odd horizontal and even vertical coordinates, all their descendants are located in the HL_ℓ detail blocks, where $\ell = \{1, 2, 3\}$. For the coefficients with even horizontal and even vertical coordinates, all their descen-

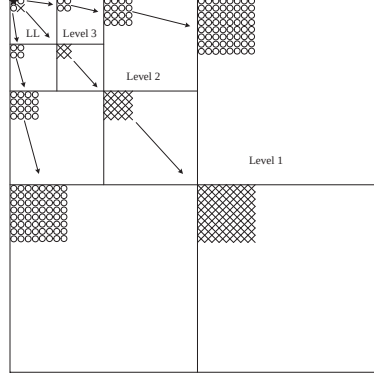


Figure 5.3: Formation of the tree structure.

dants are in the HH_ℓ detail blocks. Similarly, all the descendants for the coefficients with even horizontal and odd vertical coordinates are in the LH_ℓ blocks. Each of these roots has $(2^{(Le-\ell+1)})^2$ descendants in the corresponding detail block in level ℓ , where Le is the total number of levels of the DWT decomposition. In our implementation, $Le = 3$.

A single-root tree is illustrated in Fig. 5.4 (a). As presented above, 3/4 of the DWT coefficients in the LL block have descendants. Therefore, 3072 single-root trees can be formed in a 512×512 image. In this case, the watermark bits can be embedded into the selected single-root trees with different embedding strengths. The number of the selected single-root trees can be evenly distributed in the three orientations, HL_ℓ , HH_ℓ and LH_ℓ , or more in some orientation(s) and less in the other(s). Moreover, the single-root trees can be further categorized into bigger groups to form bigger trees. As shown in Fig. 5.4(b), the DWT image is divided into twelve multiple-roots trees and each multiple-root tree consists of 256 single-root trees. Thus, we can choose to embed the watermark in the multiple-roots trees as well. With the formation of the tree structure, the locations for the watermark embedding can be more flexibly chosen. In this thesis, the watermark is embedded into the single-root trees evenly selected from the three

orientations.

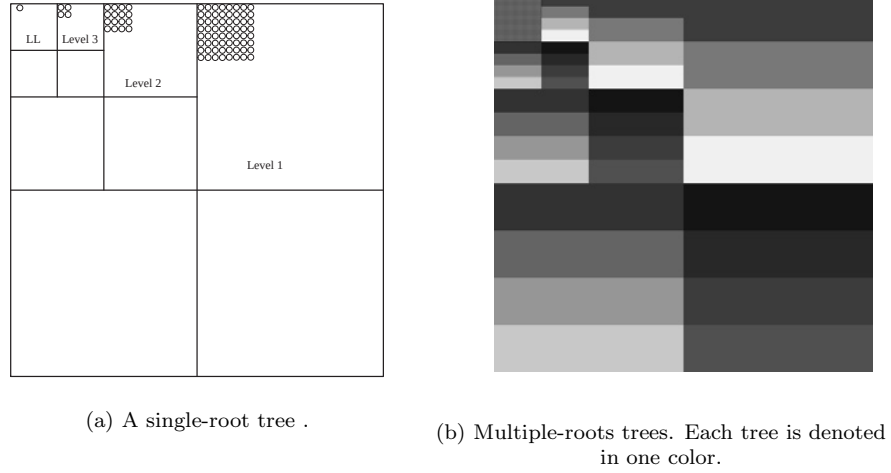


Figure 5.4: Illustration of the formed trees.

5.3.2 The selection of trees and DWT coefficients

For the applications of the watermarking based quality evaluation, the watermark is desired to be embedded throughout the cover image so that, even the watermarked image is locally tampered, the extracted watermark can still reflect the quality degradation of the cover image. According to the length of the watermark sequence, the single-root trees for watermark embedding are chosen using the the position separation key.

In the proposed tree structure based scheme, to keep the embedded watermark invisible and limit the quality degradation caused by the watermark embedding, the watermark bits are not embedded into the LL block of the DWT decomposed image and the watermark bits are not embedded into the bitplanes higher than 5. The watermark bits assignment to the DWT levels is denoted as $A_{wb} = [a_1, a_2, a_3]$, where a_1 , a_2 and a_3 are the number of watermark bits to be embedded in the DWT level 1, 2 and 3 in every

single-root tree. With A_{wb} , it becomes more flexible to choose the DWT levels that we prefer to embed more watermark bits. For watermark embedding, the redundant watermark sequence is divided into w_{segs} segments as depicted in Equ. (5.1).

$$w_{segs} = \left\lfloor \frac{Redundancy \times len}{\sum A_{wb}} \right\rfloor = \left\lfloor \frac{Redundancy \times len}{a_1 + a_2 + a_3} \right\rfloor \quad (5.1)$$

where, len is the length of the watermark sequence. If we denote the numbers of rows and columns of the DWT decomposed image as M and N , there will be totally w_{segs} trees out of $\frac{M}{2^3} \cdot \frac{N}{2^3} \cdot \frac{3}{4}$ trees selected for the watermarking processes. Here, we define another parameter $T_{per} = [T_{HL}, T_{HH}, T_{LH}]$ to control the percentage of trees selected from the three orientations. With T_{per} , we can choose to embed the watermark into the three orientations evenly or by focusing more on one or two orientations.

To serve the purpose of quality evaluation while keeping the quality degradation of the cover image caused by the watermark embedding as low as possible, two strategies are used to select the single-root trees:

1. The trees selected from the three orientations are positionally non-overlapping.

For example, in Fig. 5.5 (a), we use the square blocks in HL_3 , HH_3 and LH_3 to indicate the tree positions available in the three orientations for selection. The trees A , B and C shown in Fig. 5.5 (a) are all carrying the spatial information of the block 0 in LL . Thus, if we embed watermark bits in all the trees A , B and C , it will result in bigger quality degradation in the top-left corner of the watermarked image. Therefore, in the tree selection, if tree A is chosen for the watermark embedding, trees B and C will not be selected. In this case, for an image with $\frac{M}{2^3} \cdot \frac{N}{2^3} \cdot \frac{3}{4}$ trees, there are $T_{NP} = \frac{M}{2^3} \cdot \frac{N}{2^3} \cdot \frac{3}{4} \cdot \frac{1}{3}$ non-overlapping tree positions available for selection. In Fig. 5.5 (b), the 'X' marked positions represent the trees selected

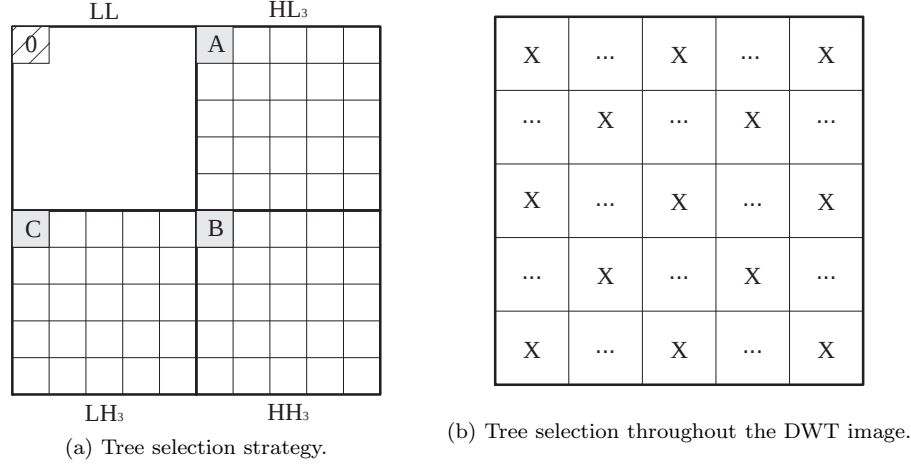


Figure 5.5: Illustration of the tree selection.

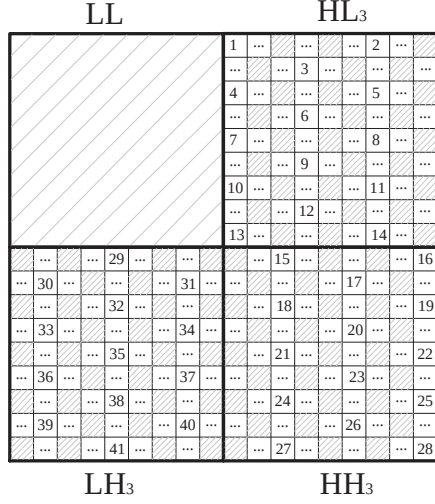
from the three DWT orientations for the watermark embedding. The ‘...’ marked positions represent the separation between any two selected tree positions. In our implementation, we use uniform separation defined as Equ. (5.2).

$$N_{sep} = \left\lfloor \frac{T_{NP}}{w_{segs}} \right\rfloor - 1 \quad (5.2)$$

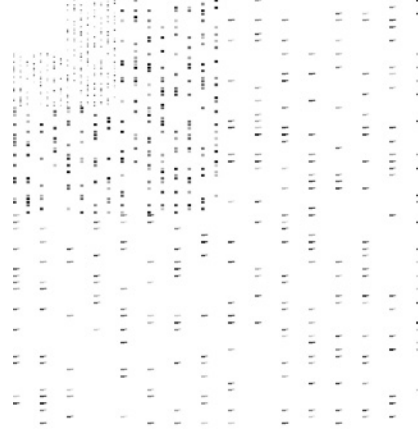
The calculated N_{sep} will be output as the position separation key and will be transmitted to the receiver side. Here, we set $w_{segs} \leq T_{NP}$.

2. The trees are selected throughout the DWT decomposed image.

The w_{seg} selected trees are further distributed into the three orientations referring to T_{per} . In the proposed scheme, we set $T_{per} = [\frac{1}{3}, \frac{1}{3}, \frac{1}{3}]$, which means that the trees are evenly selected from the three orientations as illustrated in Fig. 5.6 (a). In this figure, the selected trees are numerically ordered using $1, 2, \dots, w_{seg}$. As presented in Equ. (5.1), the watermark sequence is divided into w_{seg} segments. Thus, the watermark will be embedded into the selected trees segment by segment



(a) Graphical illustration.



(b) Experimental illustration.

Figure 5.6: The tree selection from the three orientations.

following the order number of the selected tree position.

After the tree selection, we start to embed the watermark bit segments into the selected single-root trees referring to A_{wb} . On the ℓ^{th} level, the watermark bits are embedded into the DWT coefficients one by one until the number of $A_{wb}(\ell)$ is reached, where $\ell = \{1, 2, 3\}$.

An experimental example of the tree and DWT coefficients selection is shown in Fig. 5.6 (b). The position separation key is 3. All the selected DWT coefficients are marked as dark points. The darker points means that the less significant bitplanes are selected, and vice versa. The bitplane selection will be presented in Section 5.4.2.

5.3.3 The watermark embedding

The binary watermark bits are embedded into the selected bitplanes of the selected DWT coefficients. Here, we denote the watermark bit as w , the DWT coefficient bit on

the selected bitpalne as c and the watermarked DWT coefficient bit as c_w . Then, the watermark bit will be embedded using the following equation:

$$c_w = \begin{cases} c, & \text{if } c = w \\ w, & \text{if } c \neq w \end{cases} \quad (5.3)$$

5.3.4 Security of the watermarking process

In the proposed tree structure based scheme, associated with the function of signal quality evaluation, an optional security function is designed to meet some special requirements of users to securely transmit signals over internet. This security function is controlled by the HashTable Key as shown in Fig. 5.2.

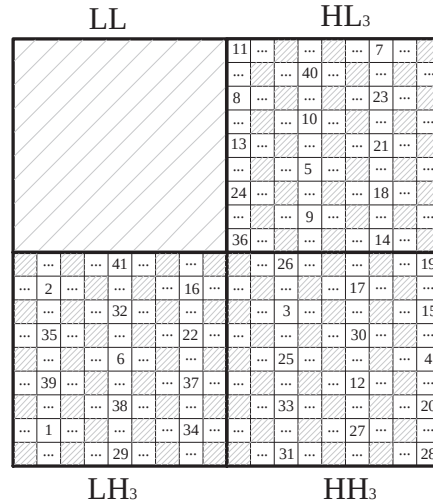


Figure 5.7: Illustration of the secure reordering of the selected trees in Fig. 5.6 (a).

The HashTable key is also a position key for the tree selection and is used to reorder the selected trees in Fig. 5.6 (a). An illustration of the secure reordering of the selected trees using the HashTable Key, 0, is shown in Fig. 5.7. In this case, the HashTable

Key is also needed to be sent to the receiver side and the lack of the HashTable key will result in a failure of watermark extraction.

5.4 The estimation of the adaptive watermark embedding strength

As mentioned in Section 5.2, the watermark embedding strength is controlled by the empirical watermark bits assignment and the bitplane selection. These two factors will be presented in Section 5.4.1 and Section 5.4.2, respectively.

5.4.1 The empirical watermark bit assignment

The empirical watermark bit assignment is to assign watermark bits to the 3 DWT levels in a selected single-root tree based on the analysis of the image complexity. It is denoted as $A_{wb} = [a_1, a_2, a_3]$ and will be the same for all the selected single-root trees in one image and may be different for different images. The watermark bits will be empirically assigned to an image in the following steps:

1. Analyze the content complexity of the cover image and calculate a complexity index.
2. Categorize the test images into different groups according to their complexity indices.
3. Assign watermark bits to the 3 DWT levels of the cover image.

All the details of the watermark bit assignment will be presented in Section 5.4.1.1 and Section 5.4.1.2.

5.4.1.1 The analysis of the content complexity

The quad-tree decomposition based complexity analysis is used in the proposed scheme for a better match with the DWT [85]. For the gray scale images, the intensity difference, V_{int} , is used to verify whether a further quad-tree decomposition is needed. Here, we define an intensity difference threshold as T_{int} . If $V_{int} > T_{int}$, the image or current block will be decomposed into 4 sub-blocks until $V_{int} \leq T_{int}$ or the size of the sub-block reaches 1.

Each quad-tree decomposition is recorded as a decomposing node. The depth of the quad-tree decomposition is denoted as the level of the decomposition. The content complexity of the cover image is assessed using the following equation [85]:

$$complexity = \sum_{i=1}^n (N_i \times 2^i) \quad (5.4)$$

where, $i \in [1, n]$ is the current quad-tree decomposition level; n is the highest decomposition level; N_i is the number of quad-tree decomposition nodes on level i . Then, the calculated complexity values of all the test images in our image library are normalized. In the proposed tree structure based scheme, the normalized content complexity value is used as the complexity index which locates in $[0,1]$.

The image content complexity analysis evaluates how much details that an image contains [86][87]. A higher complexity value indicates that the image is more complex and the image contains more detail information. Comparing to a less complex image, the quality of a more complex image degrades faster against the same distortion [20]. For this case, to reflect the quality degradation of the cover image, we need to embed more watermark bits into the lower DWT levels of a more complex image. For a less complex image, we consider to embed more watermark bits into the higher DWT levels.

Three experimental examples are shown in Fig. 5.8. These quad-tree decomposed images are achieved using the thresholds $T_{int} = 0.17$, where the maximum intensity value of the cover image is not bigger than 1. In this case, the brighter the quad-tree decomposed image, the more complex the cover image. The complexity indices are listed with the quad-tree decomposed images.

As shown in Fig. 5.8, image (a) is much less complex than the image (c) and (e). According to our experimental observations, the quality degradation for image (a) caused by some distortion will be slower than that of image (c) and (e). Thus, we need to embed more watermark bits in the lower frequency subbands for image (a) and more watermark bits in the higher frequency subbands for image (e).

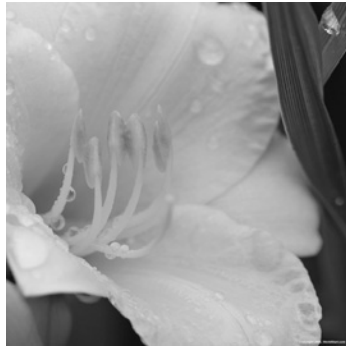
5.4.1.2 The watermark bits assignments

With the complexity indices, the watermark bits are empirically assigned to the images using the following steps:

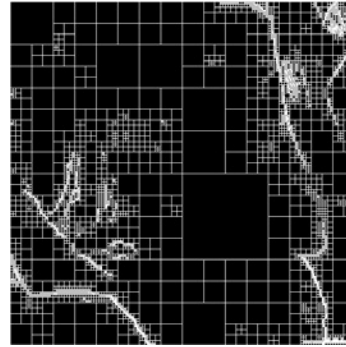
1. The complexity indices are divided into 6 groups using Equ. (5.5). One integer index is associated with each group.

$$Gindex = \begin{cases} 1, & Vc > t_1 \\ 2, & t_1 \geq Vc > t_2 \\ 3, & t_2 \geq Vc > t_3 \\ 4, & t_3 \geq Vc > t_4 \\ 5, & t_4 \geq Vc > t_5 \\ 6, & t_5 \geq Vc > 0 \end{cases} \quad (5.5)$$

where, $Gindex$ is the group index; Vc is the complexity index. t_1, t_2, t_3, t_4, t_5 and



(a) Original image.



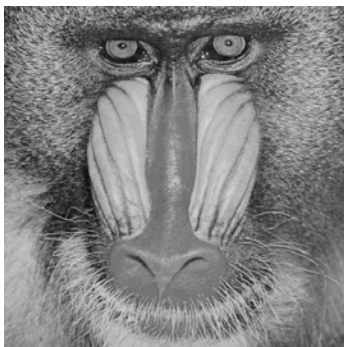
(b) Quad-tree decomposed image.
Complexity = 0.1224.



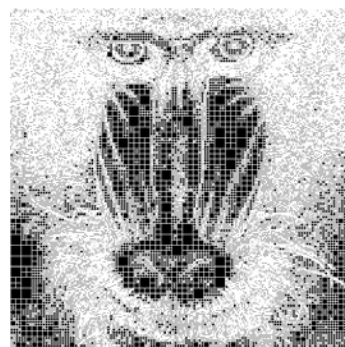
(c) Original image.



(d) Quad-tree decomposed image.
Complexity = 0.5558.



(e) Original image.



(f) Quad-tree decomposed image. Complexity
= 0.8147.

Figure 5.8: Illustration of the image complexity evaluation.

t_6 are the empirical grouping thresholds. These thresholds may be different for different distortions.

2. With the group indices, the watermark bits are assigned to the images using Equ. (5.6).

$$A_{wb} = \begin{cases} [27 \ 0 \ 0], & Gindex = 1 \\ [19 \ 7 \ 1], & Gindex = 2 \\ [13 \ 12 \ 2], & Gindex = 3 \\ [8 \ 15 \ 4], & Gindex = 4 \\ [1 \ 16 \ 4], & Gindex = 5 \\ [0 \ 8 \ 4], & Gindex = 6 \end{cases} \quad (5.6)$$

and,

$$\sum A_{wb} = \begin{cases} 27, & \text{when } Gindex \in \{1, 2, 3, 4\} \\ 21, & \text{when } Gindex = 5 \\ 12, & \text{when } Gindex = 6 \end{cases} \quad (5.7)$$

Recall that, in Equ. (5.1), $w_{segs} = \left\lfloor \frac{Rlen}{\sum A_{wb}} \right\rfloor$ trees are selected throughout the DWT decomposed image, where $Rlen$ is the length of the redundant watermark sequence; w_{segs} is the number of single-root trees selected for the watermark embedding. Thus, for images with different complexity, the number of selected trees and the position separation key (N_{sep}) may be different. For example, to evaluate image quality in terms of PSNR under JPEG compression, for the image in Fig. 5.8 (a), $Gindex = 6$, $A_{wb} = [0, 8, 4]$, $\sum A_{wb} = 12$, and $w_{segs} = \left\lfloor \frac{Rlen}{12} \right\rfloor$. For the image in Fig. 5.8 (c), $Gindex = 3$, $A_{wb} = [13, 12, 2]$, $\sum A_{wb} = 27$, and

$w_{segs} = \lfloor \frac{Rlen}{27} \rfloor$. Recall that the position separation security key $N_{sep} = \lfloor \frac{T_{NP}}{w_{segs}} \rfloor$, where T_{NP} is the total non-overlapping tree positions and is a fixed number for all the images. Thus, for different images, N_{sep} may be different. Therefore, it is necessary to transmit $Gindex$ and N_{sep} to the receiver side.

5.4.2 The HVS and bitplane based perceptual masking

The other factor that controls the watermark embedding strength is the bitplane selection for the watermark embedding. By decomposing the DWT image into a set of binary bitplanes, each DWT coefficient is decomposed into a binary bit sequence. The less significant bits are more sensitive to distortions than the more significant bits. If we embed the whole watermark in the least significant bitplane of the DWT image, we achieve the most fragile watermark and the least quality degradation to the cover image, and vice versa. In the proposed scheme, the bitplane selection is guided by the analysis of the HVS masking effects on the DWT decomposed image. This strategy can effectively reduce the quality degradation caused by the watermarking processes.

The bitplane selection includes two steps: calculating the HVS masks and mapping the calculated perceptual masks to bitplane indices. The achieved bitplane indices decide which bitplanes of the selected DWT coefficients to embed the watermark bits. The HVS masking calculation and the mapping to the bitplane indices are presented in details in the following two subsections.

5.4.2.1 The HVS masking

The HVS masking is used in the proposed scheme to better balance the invisibility and the robustness of the embedded watermark. One HVS mask is generated for one DWT decomposed detail block. Totally, nine HVS masks are generated for one image in the

watermark embedding process. To accommodate the proposed tree structure based watermarking scheme, every HVS mask is mapped into bitplane indices referring to the distribution of the HVS mask. The mapping relationship between the coefficients of the HVS masks and the bitplane indices is defined based on experiments. Then, the binary watermark bits are embedded into the selected bitplanes referring to the achieved bitplane indices. The luminance values of the DWT coefficients in the LL block are used in the HVS masking calculation.

The HVS masking presented in [63] is used in the proposed scheme. As researched in [63][88][89], four factors, as listed in the following, are greatly affecting the behavior of the HVS:

1. Band sensitivity or frequency masking: Intensity variations are less visible in high resolution subbands and are also less visible in the diagonally decomposed blocks, HH_ℓ . This factor is expressed using Equ. (5.8).

$$M_F(\ell, \theta) = M_1(\theta) \cdot M_2(\ell) \quad (5.8)$$

where,

$$M_1(\theta) = \begin{cases} \sqrt{2}, & \text{if } \theta = 2 \\ 1, & \text{o.w.} \end{cases}$$

$$M_2(\ell) = \begin{cases} 1, & \text{if } \ell = 1 \\ 0.32, & \text{if } \ell = 2 \\ 0.16, & \text{if } \ell = 3 \end{cases}$$

and,

$$\theta = \begin{cases} 1, & \text{for HL blocks} \\ 2, & \text{for HH blocks} \\ 3, & \text{for LH blocks} \end{cases}$$

2. Background luminance: Intensity variations are less visible over the brighter and darker areas. The luminance masking is denoted as M_L .

$$\begin{aligned} M_L(\ell, i, j) &= 1 + I(\ell, i, j) \\ &= \begin{cases} 2 - \frac{1}{256} I_{LL} \left(\left\lceil \frac{i}{2^{Le-\ell}} \right\rceil, \left\lceil \frac{j}{2^{Le-\ell}} \right\rceil \right), & \text{if } I(\ell, i, j) < 0.5 \\ 1 + \frac{1}{256} I_{LL} \left(\left\lceil \frac{i}{2^{Le-\ell}} \right\rceil, \left\lceil \frac{j}{2^{Le-\ell}} \right\rceil \right), & \text{o.w.} \end{cases} \end{aligned} \quad (5.9)$$

3. Spatial masking or edge proximity: The eyes are more sensitive to noise addition near edges or contours of images. This factor, M_E , is evaluated using the empirically scaled local energy of the DWT coefficients in all detail subbands.

$$M_E(\ell, i, j) = \sum_{k=0}^{Le-\ell} \rho \sum_{\theta=1}^3 \sum_{x=0}^1 \sum_{y=0}^1 \left[I_{k+\ell}^{\theta} \left(x + \left\lceil \frac{i}{2^k} \right\rceil, y + \left\lceil \frac{j}{2^k} \right\rceil \right) \right]^2 \quad (5.10)$$

where, ρ is a weighting parameter and the suggested value for ρ is presented in the following equation [63].

$$\rho = \begin{cases} \frac{1}{4}, & \text{if } k = 0 \\ \frac{1}{16^k}, & \text{o.w.} \end{cases}$$

4. Texture sensitivity: Intensity variations in highly textured areas are less visible than those in the flat-filed areas of images. This masking factor, M_T , is estimated using the local variance of the corresponding DWT coefficients in the LL subband.

$$M_T(\ell, i, j) = var \left\{ I_{LL} \left(x + \left\lceil \frac{i}{2^{L_e - \ell}} \right\rceil, y + \left\lceil \frac{j}{2^{L_e - \ell}} \right\rceil \right) \right\}_{x=\{0,1\}; y=\{0,1\}} \quad (5.11)$$

The HVS mask is achieved by polling the four factors listed above and is computed using Equ. (5.12) [63].

$$M_{HVS}(\ell, \theta, i, j) = \alpha \cdot M_F(\ell, \theta) \cdot M_L(\ell, i, j) \cdot M_E(\ell, i, j)^\beta \cdot M_T(\ell, i, j)^\gamma \quad (5.12)$$

where M_{HVS} denotes the coefficients in one HVS mask; $\alpha = \frac{1}{2}$ is a scaling parameter, which implies that disturbs having values lower than half of $M_F \cdot M_L \cdot M_E^\beta \cdot M_T^\gamma$ are assumed invisible. The suggested value for β and γ is 0.2 [63]. In Equ. (5.8) to Equ. (5.12), (i, j) are the coordinates of the current pixel in the calculation. For computational convenience, we use the relative coordinates of a single DWT block. In this case, $i \in [1, \frac{M}{2^\ell}]$ and $j \in [1, \frac{N}{2^\ell}]$ (M and N are the number of rows and columns of an image respectively). $I(\ell, i, j)$ is the luminance value of the pixel, (i, j) , in one detail block in level ℓ ; $\ell \in \{1, 2, \dots, L_e\}$ indicates the current DWT level in the HVS masking calculation and $L_e = 3$ is the maximum level applied in the DWT decomposition in this thesis; $I_{LL}(\lceil \frac{i}{2^{L_e - \ell}} \rceil, \lceil \frac{j}{2^{L_e - \ell}} \rceil)$ is the luminance value of pixel $(\lceil \frac{i}{2^{L_e - \ell}} \rceil, \lceil \frac{j}{2^{L_e - \ell}} \rceil)$ in the approximation subband which corresponds to the pixel (i, j) in a detail block on level ℓ .

As proposed in Section 5.2, the binary watermark bits are not embedded in the approximation subband so that the invisibility of the embedded watermark can be further improved. Therefore, the HVS masks are only calculated for the nine detail



(a) The original image of **Barbara**.

(b) The nine HVS masks generated for the nine DWT detail blocks.

Figure 5.9: Illustration of the generated HVS masks.

blocks, with one mask for each detail block. The generated HVS masks for image Barbara are shown in Fig. 5.9 (b).

5.4.2.2 Mapping from the HVS mask to the bitplane indices

To use the HVS masking in the watermark embedding process, a mapping relationship from the coefficients of the HVS mask to the bitplane indices is experimentally defined using the multiple-bands-thresholding method. Considering that different HVS masks have different distributions, the thresholds used in the mapping procedure should be able to change with the shape of the distribution of the HVS mask. To further limit the quality degradation caused by the watermarking processes, only the bitplanes from 1 to 5 are used for the watermark embedding. The thresholds calculation and the mapping procedure are listed as follows:

1. The HVS mask is first normalized using its maximum coefficient so that the

mapping relationship can be consistent with different HVS masks.

2. The coefficients of the HVS masks are sorted in an ascending order and the sorted sequence is denoted as *Sortmask*.
3. The thresholds are calculated using Equ. (5.13).

$$T_n(\ell, \theta) = \text{Sortmask} \left(\left\lfloor \frac{n \cdot N_{\text{mask}}(\ell, \theta)}{5} \right\rfloor \right) \quad (5.13)$$

where, $n \in [1, 2, 3, 4]$, N_{mask} is the total number of the coefficients in one HVS mask, the denominator indicates that 5 bitplanes are used for the watermark embedding.

4. The mapping is conducted using Equ. (5.14).

$$I_{bp}(\ell, \theta, i, j) = \begin{cases} 1, & v(\ell, \theta, i, j) \leq T_1(\ell, \theta) \\ 2, & T_1(\ell, \theta) < v(\ell, \theta, i, j) \leq T_2(\ell, \theta) \\ 3, & T_2(\ell, \theta) < v(\ell, \theta, i, j) \leq T_3(\ell, \theta) \\ 4, & T_3(\ell, \theta) < v(\ell, \theta, i, j) \leq T_4(\ell, \theta) \\ 5, & T_4(\ell, \theta) < v(\ell, \theta, i, j) \leq 1 \end{cases} \quad (5.14)$$

where, (i, j) are the coordinates of a selected DWT coefficient. $I_{bp}(\ell, \theta, i, j)$ means the bitplane index achieved for the pixel located at (i, j) on DWT level ℓ with orientation θ . v is the value of the HVS mask coefficient.

Therefore, each DWT coefficient in the selected single-root trees has its own $I_{bp}(\ell, \theta, i, j)$. At the receiver side, the strategy presented above will also be applied on the distorted image to locate the bitplanes for the watermark extraction. To increase the probability of correct extraction of the watermark bits, in every single-root

tree on the DWT level ℓ at orientation θ , all the calculated $I_{bp}(\ell, \theta, i, j)$ values are averaged. In other words, in each selected tree, for all the selected coefficients, the watermark bits will be embedded on the same bitplane. Thus, the bitplane indices are updated as Equ. (5.16).

$$\begin{aligned} (i, j) &= (i_{tree} + i_r, j_{tree} + j_r) \\ &= \left(\left\lceil \frac{i}{2^{Le-\ell+1}} \right\rceil + rem\left(i - 1, 2^{Le-\ell+1}\right), \left\lceil \frac{j}{2^{Le-\ell+1}} \right\rceil + rem\left(j - 1, 2^{Le-\ell+1}\right) \right) \end{aligned} \quad (5.15)$$

$$I_{bp}(\ell, \theta, i_{tree}, j_{tree}) = \left\lfloor \frac{\sum_{i_r=0}^{2^{(Le-\ell+1)}-1} \sum_{j_r=0}^{2^{(Le-\ell+1)}-1} I_{bp}(\ell, \theta, i, j)}{(2^{(Le-\ell+1)})^2} \right\rfloor \quad (5.16)$$

where $I_{bp}(\ell, \theta, i_{tree}, j_{tree})$ is the averaged bitplane of the DWT coefficients located in a specific tree on level ℓ and orientation θ .

Two examples of the thresholds selection are shown in Fig. 5.10. The computed thresholds for Fig. 5.10 (c) are $[T_1, T_2, T_3, T_4] = [0.0706, 0.1153, 0.1774, 0.2659]$. The thresholds for Fig. 5.10 (f) are $[T_1, T_2, T_3, T_4] = [0.1073, 0.1347, 0.1664, 0.2201]$. The histograms are generated using Matlab, which categorizes the coefficients of the HVS masks into 10 equally spaced bins and leaves a narrow space between every two bins for clearer view. Because of the way that the histogram is displayed, an extra explanation needs to be made for Fig. 5.10 (c): the number of the coefficients located between $[0, 0.05]$ is 6648, which is not exactly the same as shown in the figure. The figures are only used to illustrate how the thresholds are selected.

With the calculated thresholds, the coefficients of the HVS mask which correspond to the selected DWT coefficients are mapped to the bitplane indices. In this case, the bitplanes of the selected coefficients for the watermark embedding are located.

As expressed in Equ. (5.13), the thresholds for the mapping from HVS mask to bitplane indices are calculated based on sorting the coefficients in the HVS mask. The sorting step can be removed and the thresholds can be alternatively computed by examining the statistics of the HVS mask. An example is depicted in Equ. (5.17).

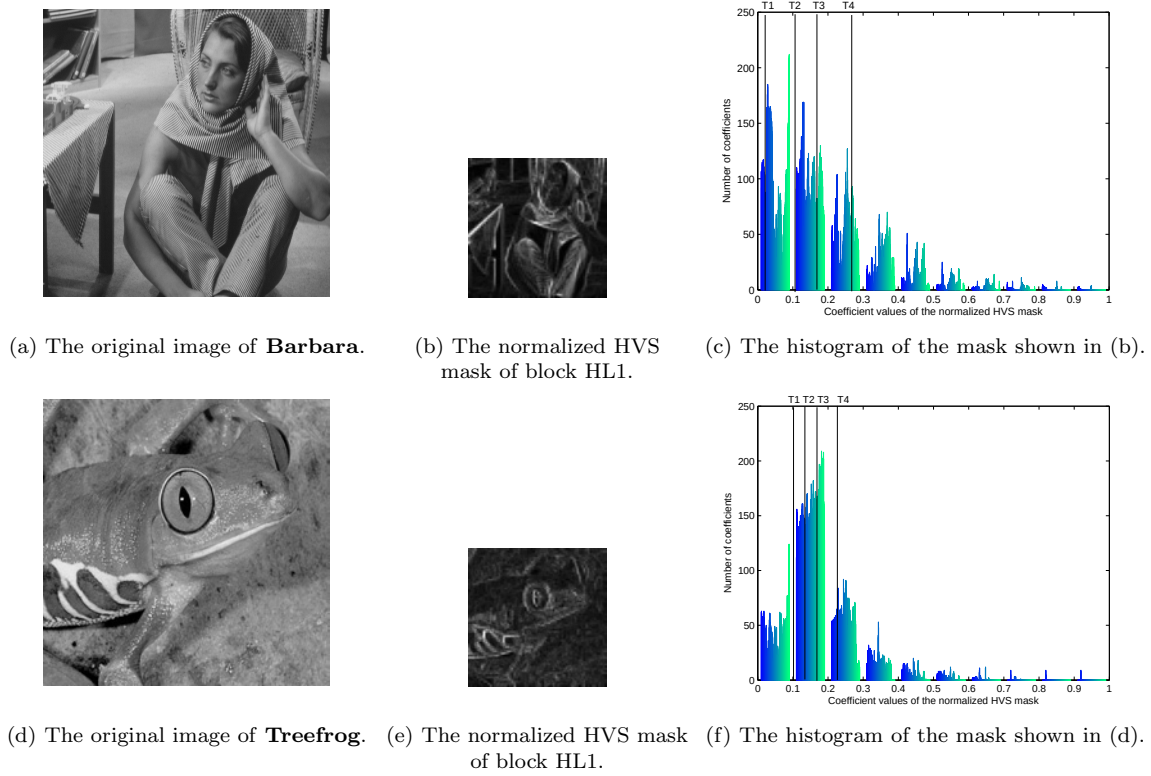


Figure 5.10: Illustration of the mapping thresholds selection.

$$T_n(\ell, \theta) = \min(M_{HVS}(\ell, \theta)) + n \cdot \frac{\max(M_{HVS}(\ell, \theta)) - \min(M_{HVS}(\ell, \theta))}{5} \quad (5.17)$$

where, $n \in [1, 2, 3, 4]$, $M_{HVS}(\ell, \theta)$ denotes the coefficients in the HVS mask on DWT level ℓ at orientation θ . The $\max((M_{HVS}(\ell, \theta)))$ and $\min((M_{HVS}(\ell, \theta)))$ respectively are the maximum and minimum coefficients in the HVS mask on DWT level ℓ at orientation

θ . The removal of the sorting step helps reduce the computational complexity of the proposed scheme.

5.5 The tree structure based watermark extraction and quality evaluation

The proposed watermark extraction and quality evaluation scheme is shown in Fig. 5.11, where the position separation key is used to locate the watermarked DWT coefficients.

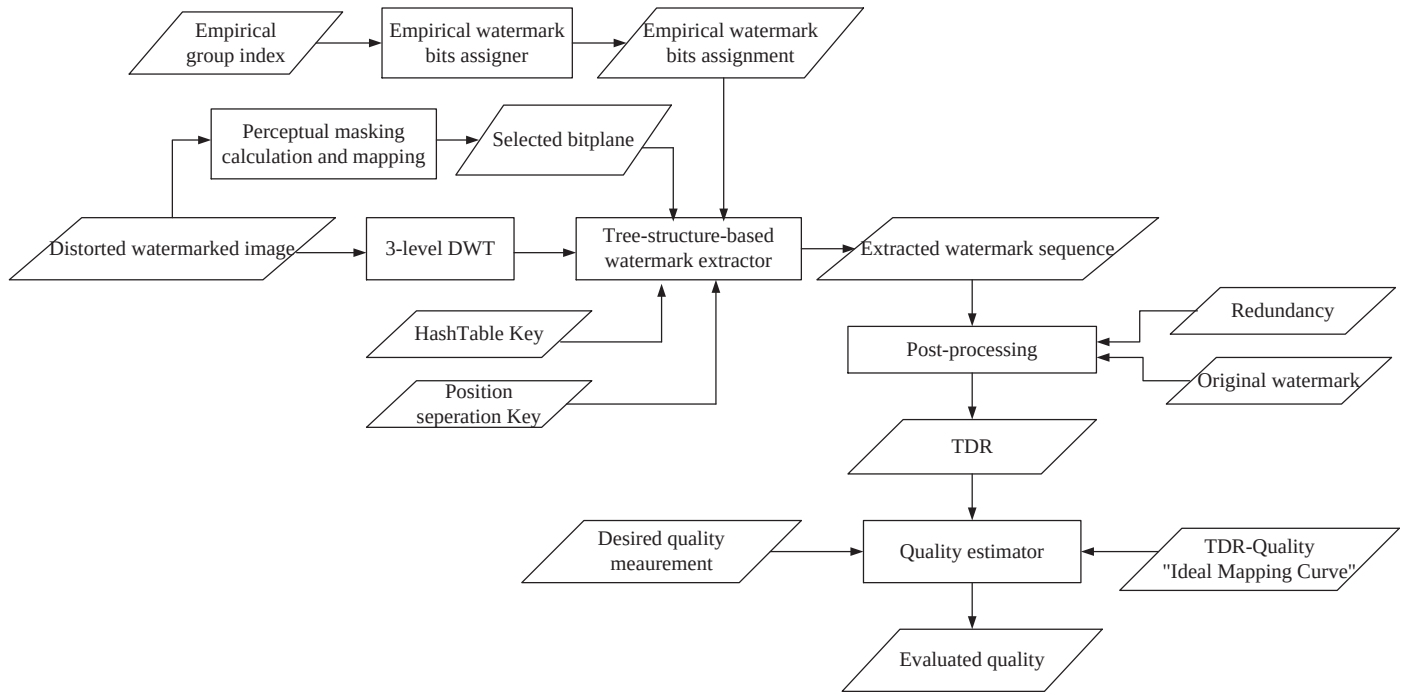


Figure 5.11: The SPIHT tree structure and HVS based watermark extraction and quality evaluation process.

The watermark bit assignment is retrieved using the image group index transmitted from the sender side. The bitplane indices for watermark extraction are obtained by calculating the HVS masks of the distorted watermarked image. As presented in Section 5.4.2.2, in one tree, the bitplane indices for all the DWT coefficients on each DWT level

are averaged. This strategy effectively reduces the watermark extraction error caused by the bitplane selection in the watermark extraction scheme.

Recall that *Redundancy* = 3, in the watermark post-processing procedure, the extracted redundant watermark sequence is recovered to three distorted watermarks. Then, the three distorted watermarks are compared bit by bit and the distorted watermark is extracted using Equ. (5.18).

$$w_e(i, j) = \begin{cases} 1, & N_1 \geq N_0 \\ 0, & N_1 < N_0 \end{cases} \quad (5.18)$$

where, $w_e(i, j)$ is the extracted watermark bit with coordinates (i, j) ; N_1 is the number of extracted 1s and N_0 is the number of extracted 0s.

Then, the extracted watermark is compared with the original watermark bit by bit and the True Detection Rates (TDR) is calculated using Equ. (5.19).

$$TDR = \frac{\text{Number of correctly detected watermark bits}}{\text{Total number of watermark bits}} \quad (5.19)$$

With Equ. (3.2), the quality of the distorted image is evaluated by mapping the calculated TDR to a quality value by referring to the “Ideal Mapping Curve”. The value of the calculated TDR could possibly lie between two neighboring TDR values on the “Ideal Mapping Curve” as shown in Fig. 5.12. Through the experiments, it is found that, by increasing of distortion strength, the calculated TDR values decrease monotonously [51]. In this case, linear interpolation is used to estimate the image quality. An example of the quality evaluation is presented in the following.

In Fig. 5.12, $(T(j), Q(j))$ and $(T(j + 1), Q(j + 1))$ are two adjacent points on the “Ideal Mapping Curve”, where $T(\bullet)$ is the TDR value and $Q(\bullet)$ is the quality value

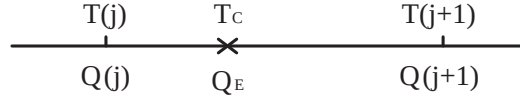


Figure 5.12: Linear interpolation.

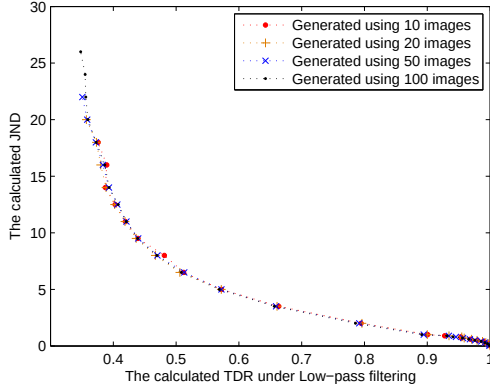
in terms of PSNR, wPSNR, Watson JND, SSIM or VIF. T_C is the calculated TDR of any distorted image and Q_E is the evaluated quality which can be obtained using Equ. (5.20).

$$Q_E = Q(j) + \frac{T_C - T(j)}{|T(j+1) - T(j)|} \times |Q(j+1) - Q(j)| \quad (5.20)$$

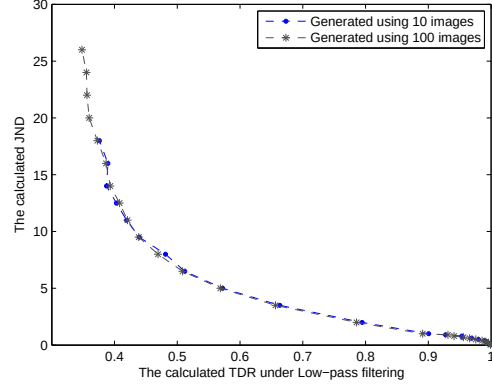
5.6 The generation of the “Ideal Mapping Curve”

In the SPIHT tree structure and HVS based quality evaluation scheme, the “Ideal Mapping Curve” is generated using the “Ideal Mapping Curve” generator presented in Section 3.4 to define the mapping relationship between the TDR and the possible quality values of the distorted images. The TDR is the True Detection Rate of the extracted watermark and is defined in Equ. (5.19). Because the proposed tree structure can work in estimating image quality in terms of different quality metrics such as PSNR, wPSNR, Watson JND, SSIM, or VIF to achieve the highest accuracy, one “Ideal Mapping Curve” is generated in terms of one quality metric under one distortion.

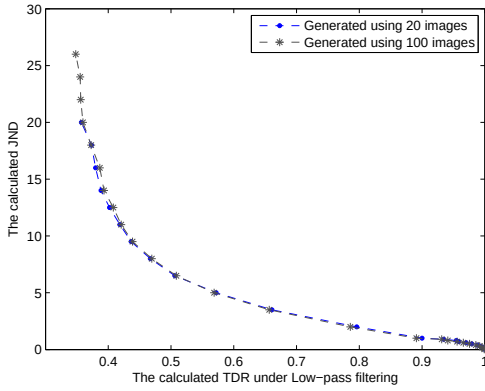
With the “Ideal Mapping Curve” generator shown in Fig. 3.9, to generate the “Ideal Mapping Curve” in the tree structure based scheme, the tree structure based watermark embedder and watermark extractor are used. In the meanwhile, a binary watermark is embedded in the N test images using adaptive watermark embedding strength estimated in Section 5.4. The “Ideal Mapping Curve” is generated by locally averaging the TDR-Quality curves tested using the N images.



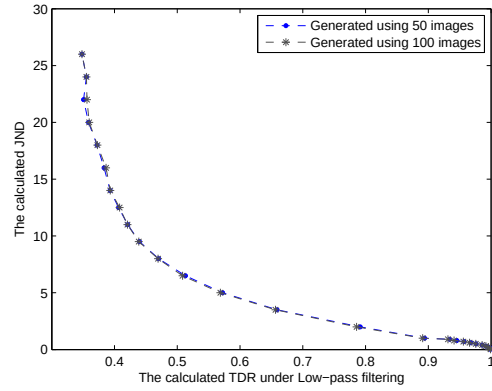
(a) The “Ideal Mapping Curves” generated using different numbers of images.



(b) The “Ideal Mapping Curve” generated using 10 images and 100 images.



(c) The “Ideal Mapping Curve” generated using 20 images and 100 images.



(d) The “Ideal Mapping Curve” generated using 50 images and 100 images.

Figure 5.13: Comparisons of the “Ideal Mapping Curve” generated using different number of test images.

An example of “Ideal Mapping Curves” generated respectively using 10, 20, 50 and 100 different natural images in terms of JND under Gaussian low-pass filtering are shown in Fig. 5.13 (a). From this figure, we can see that the four generated “Ideal Mapping Curves” are nearly identical. As mentioned previously, the adaptive watermark embedding strength is assigned by analyzing the quality degradation characteristics of the cover image. With the adaptive watermark embedding strength, we expect the TDR-Quality curves tested using the N images as convergent as possible. The more

convergent the tested TDR-Quality curves, the more accurate the quality estimation. The similar convergence of the TDR-Quality curves tested respectively using 10, 20, 50 and 100 images results in the four generated “Ideal Mapping Curves” nearly identical as shown in Fig. 5.13 (a), which also indicates the flexibility of choosing the number of images for the “Ideal Mapping Curve” generation. With the proposed tree structure based scheme, we suggest to use at least 10 natural images for generating the “Ideal Mapping Curves” to make it statistically meaningful. In our implementation, we use 50 natural images selected from our image library to generate the “Ideal Mapping Curves”. The 50 test images include different types of characteristics, such as portrait, plants, animals, animation, scenery, buildings and crowd. The scope of our image library is presented in Section 6.1 and all the original images in the image library are shown in Fig. A-1 in Appendix.

The “Ideal Mapping Curves” generated using 10, 20 and 50 images are respectively compared with the “Ideal Mapping Curve” generated using 100 images as shown in Fig. 5.13 (b) - (d) for additional observations.

Two examples of the “Ideal Mapping Curve” generation in terms of Watson JND and PSNR under Gaussian low-pass filtering are shown in Fig. 5.14. The gray dots are the TDR-Quality curves calculated using the 50 test images. The 50 TDR-Quality curves are computed when the filter variance $Var_{filter} = [0.1 : 0.1 : 0.3, 0.31 : 0.01 : 0.5, 0.55 : 0.1 : 1.5]$. When $Var_{filter} = 1.5$, the quality of the low-pass filtered image is about 25 dB in PSNR. The “Ideal Mapping Curve” is denoted using black dot dashed curve and is calculated by locally averaging the 50 gray dot curves. The rule of the local averaging has been presented in Section 3.4 in detail and is used for the “Ideal Mapping Curve” generation under other distortions, such as JPEG2000 compression, JPEG compression and Gaussian noise distortion.

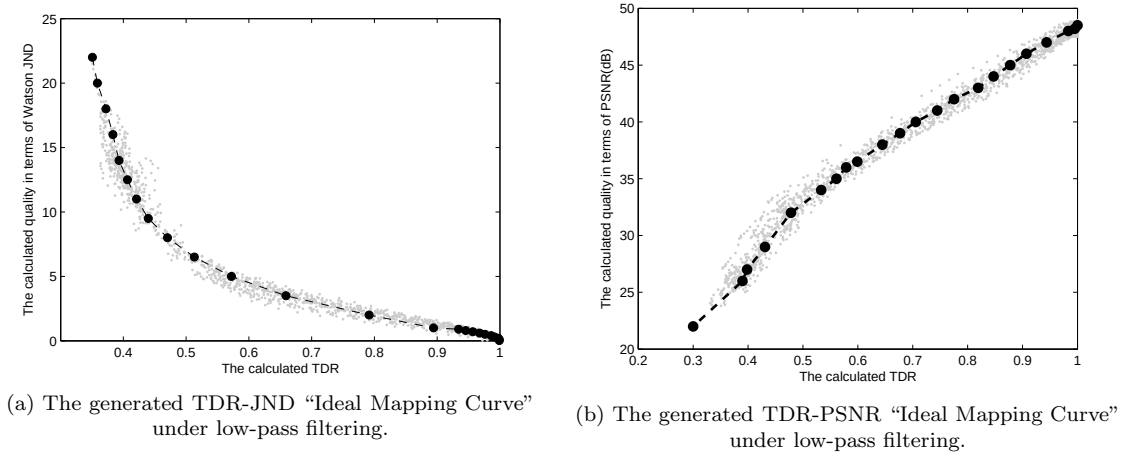


Figure 5.14: Generation of the “Ideal Mapping Curve” using adaptive watermark embedding strengths in the tree structure based scheme.

At the end of this section, an extra comparison is made between Fig. 4.9 and Fig. 5.14. In these two figures, the “Ideal Mapping Curves” are both generated by averaging the tested TDR-Quality curves. Thus, the “Ideal Mapping Curve” carries statistical information of the quality degradation characteristics of images. In the watermarking based quality evaluation scheme, the convergence of the tested TDR-Quality curves directly affects the accuracy of the quality evaluation. As shown in Fig. 5.14, once the watermark is embedded with the watermark embedding strengths adaptive to the quality degradation characteristics of the cover images, it has the ability to reflect the image quality changes and the tested TDR-Quality curves are convergent. Based on the experiments, the better the TDR changes reflecting the image quality changes, the more convergent the tested TDR-Quality curves, and the smaller the quality evaluation error. In Fig. 4.9, the tested TDR-Quality curves are obtained using a fixed watermark embedding strength without considerations of the image quality degradation characteristics, which results in the tested TDR-Quality curves divergent. Thus, to maximize the

quality evaluation accuracy in the quantization based scheme, the tested TDR-Quality curves in Fig. 4.9 are made convergent by iteratively optimizing the watermark embedding strengths referring to the image quality degradation characteristics. By this way, the watermark embedded with the optimized embedding strength is able to reflect the image quality changes. In the tree structure based scheme, the accuracy of the quality evaluation is maximized by pre-estimating the quality degradation characteristics of the cover image. On the contrary, in the quantization based scheme, the accuracy of the image quality evaluation is maximized by experimentally post-evaluating the quality degradation characteristics of the cover images at the sender side.

In conclusion, in the quantization based scheme and the tree structure based scheme, the high quality evaluation accuracy is achieved using the same rules in different ways. With the involvement of the iterative optimization process, the quantization based scheme has relatively low computational efficiency. Based on the pre-estimation of the image degradation characteristics, the tree structure based scheme has much higher computational efficiency.

5.7 Summary

In this chapter, a quality evaluation scheme using SPIHT tree structure and HVS based watermarking is presented according to the proposed framework. This scheme is developed aiming to provide accurate quality evaluation and high computational efficiency.

The tree structure based scheme is implemented in the DWT domain. The DWT and SPIHT together provide a good summarization of local region characteristics of the cover image. All the correlated DWT coefficients across the subbands are categorized

into SPIHT trees. The watermark is embedded in the selected SPIHT trees of the cover image with adaptive watermark embedding strength which is obtained by pre-estimating the image degradation characteristics. The HVS perceptual masking is used to guide the watermark embedding. The experimental results will be presented in the following chapter.

Chapter 6

Experimental results of tree structure based image quality evaluation

The SPIHT tree structure and HVS based quality evaluation scheme will be evaluated in this chapter.

The data need to be transmitted to the receiver side are:

1. The original binary watermark which is 48×48 pixels and 2304 bits. There is no need to transmit the watermark if both the sender and receiver side use the same watermark.
2. The 2 position security keys. If the Hashtable key is fixed at both the sender and receiver side, only the position separation key needs to be transmitted.
3. The empirical group index for the watermark bits assignment.

6.1 The original images and the original watermark

There are 150 gray images in our image library including all the reference images from Cornell A57 image database, IVC image database, TID2008 image database, and CSIQ image database. Besides these image databases, our image library also includes computer generated images and more natural images. All of these images are 512×512 in size containing different textures, such as, portraits, plants, animals, animations, sceneries, buildings and crowd and these images are shown in Fig. A-1 in Appendix A. The first 50 images in the image library are used to generate the “Ideal Mapping Curve” and other empirical parameters. To avoid losing generality, these images are not used to test the proposed scheme. The 51st to the 150th images are used to evaluate the accuracy of the proposed scheme.

The original watermark used in the experiments is 48×48 in size and is shown in Fig. 6.1. The binary watermark is randomly generated and there is no special requirements for it. The original watermark can be changed to any size or pattern.



Figure 6.1: The original watermark.

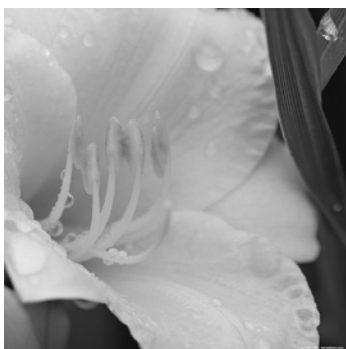
Fig. 6.2 shows the effects of the watermark embedding under Gaussian low-pass filtering. Fig. 6.2 (a), (c) and (e) are the original images with different complexity and their watermarked images are shown in Fig. (b), (d) and (f), respectively. Recall that

$A_{wb} = [a_1, a_2, a_3]$ is the watermark bits assignment which assigns a_1 bits of watermark to the DWT level 1 in a single-root tree; a_2 bits of watermark to the DWT level 2 in a single-root tree and a_3 bits of watermark to the DWT level 3 in a single-root tree. As shown in the figure, the higher the complexity value, the more complex the image. The complexity value calculated for the image White flower is 0.1224. The A_{wb} associated with it is $[0, 8, 4]$ which indicates that the image, White flower, needs a strong watermark embedding strength to reflect the degradation caused by the distortion. The calculated complexity value for image Barbara is 0.5558. It has $A_{wb} = [13, 12, 2]$, which indicates that the image needs a moderate watermark embedding strength. Among the three original images, the image Baboon shown in Fig. 6.2(e) has the highest complexity which is computed as 0.8147 and it needs the weakest embedding strength. That is because image Baboon has vast of details which will be distorted quickly under distortions. Thus, we need to embed most of the watermark bits in the details to reflect the quality loss. In the figure, the PSNR values of all the watermarked image are more than 47.5 dB, which proves that the quality degradation caused by the watermark embedding process is very limited.

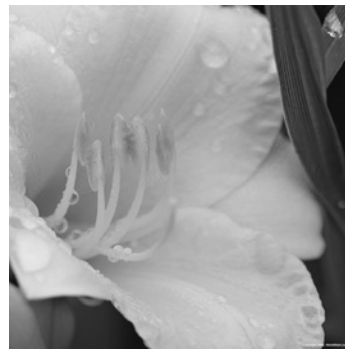
6.2 Summary of the experiments

The proposed adaptive watermarking and tree structure based quality metric is tested under JPEG compression, JPEG2000 compression, Gaussian low-pass filtering and Gaussian noise pollution. Under each distortion, the watermarked image is distorted using different distortion strengths.

In the meanwhile, under each distortion, the quality of the distorted images are evaluated in terms of PSNR, wPSNR, Watson JND, SSIM and VIF, respectively. Therefore,



(a) Original image White flower.
Complexity = 0.1224.



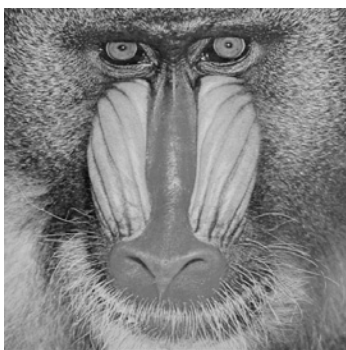
(b) The watermarked image. Watermark bits assignment: $A_{wb} = [0, 5, 4]$. PSNR = 47.6549 dB.



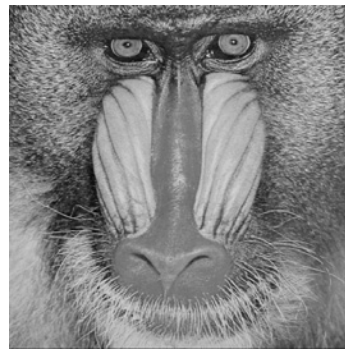
(c) Original image Barbara.
Complexity = 0.5558.



(d) The watermarked image. Watermark bits assignment: $A_{wb} = [13, 12, 2]$. PSNR = 48.0903 dB.



(e) Original image Baboon.
Complexity = 0.8147.



(f) The watermarked image. Watermark bits assignment: $A_{wb} = [27, 0, 0]$. PSNR = 48.2801 dB.

Figure 6.2: Illustration of the effects of watermark embedding.

totally 20 sets of experiments are implemented for the tree structure based image quality evaluation. In each set of the experiment, 100 images are tested. To evaluate the performance of the proposed algorithm, the quality of the distorted images comparing to the original images are calculated as well. Then, as presented in Section 3.5, the evaluated quality values and their corresponding calculated quality values are compared and the quality evaluation accuracy is assessed using MAE (Mean Absolute Error) and Pearson correlation coefficient, respectively.

The wPSNR model discussed in [1], the software Dctune2.0 [82], the SSIM metric presented in [4], and the VIF metric reported in [5] are used for the image quality calculation in terms of wPSNR, Watson JND, SSIM and VIF respectively.

6.2.1 Choosing the range for the distortion strength

The watermarked images are distorted using a set of distortion strengths varying in a selected range, $ds \in [ls, hs]$, where ds is the distortion strength; ls and hs respectively are the lowest distortion strength and the highest distortion strength that is used in the experiments. We choose the lowest available distortion strength that can be used in Matlab as ls . For example, under JPEG compression, we set the quality factor, 100, as ls . With ls , the least quality loss caused by distortion will be resulted. The hs is selected by evaluating the averaged quality of the degraded watermarked images, $\overline{Q_{wimg}|_{hs}}$, distorted using a few interested hs values. $Q_{wimg} = PSNR(wimg, orgimg)$, where the $wimg$ is the watermarked image and $orgimg$ is the original image. If the $\overline{Q_{wimg}|_{hs}} \in [24.50, 25.50]$ dB in PSNR, the hs will be selected. On the other hand, when the quality of the distorted image is below 30 dB in PSNR, the image quality is poor. The lower the image quality, the less meaningful to focus on its quality evaluation accuracy. Thus, we choose another highest boundary for ds : hs' if $\overline{Q_{wimg}|_{hs'}} \in [29.50, 30.50]$ dB in

PSNR. The 150 images in the image library are used to choose ls , hs and hs' . The range boundaries for the distortion strengths are listed in the following for the four distortions used in the experiments:

JPEG: the distortion strength is the compression strength and is denoted using “Quality Factor” in Matlab.

$$ls = 100. \overline{Q_{wimg}|_{ls}} = 47.7092 \text{ dB in PSNR.}$$

$$hs = 5. \overline{Q_{wimg}|_{hs}} = 25.0870 \text{ dB in PSNR.}$$

$$hs' = 20. \overline{Q_{wimg}|_{hs'}} = 30.2940 \text{ dB in PSNR.}$$

JPEG2000: the distortion strength is the compression rate.

$$ls = 1. \overline{Q_{wimg}|_{ls}} = 48.0957 \text{ dB in PSNR.}$$

$$hs = 0.01. \overline{Q_{wimg}|_{hs}} = 24.4588 \text{ dB in PSNR.}$$

$$hs' = 0.05. \overline{Q_{wimg}|_{hs'}} = 30.1790 \text{ dB in PSNR.}$$

Gaussian low-pass filtering: The distortion strength is controlled by the standard deviation of Gaussian filter.

$$ls = 0.1. \overline{Q_{wimg}|_{ls}} = 48.3144 \text{ dB in PSNR.}$$

$$hs = 1.5. \overline{Q_{wimg}|_{hs}} = 27.7597 \text{ dB in PSNR.}$$

$$hs' = 1. \overline{Q_{wimg}|_{hs'}} = 29.7629 \text{ dB in PSNR.}$$

Gaussian noise: The distortion strength is the standard deviation of the Gaussian noise:

$$ls = 0. \overline{Q_{wimg}|_{ls}} = 48.1919 \text{ dB in PSNR.}$$

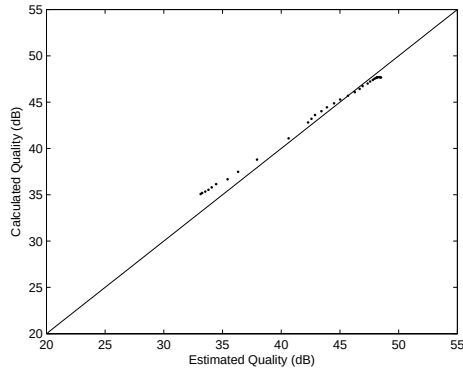
$$hs = 15. \overline{Q_{wimg}|_{hs}} = 24.7533 \text{ dB in PSNR.}$$

$$hs' = 8. \overline{Q_{wimg}|_{hs'}} = 30.0929 \text{ dB in PSNR.}$$

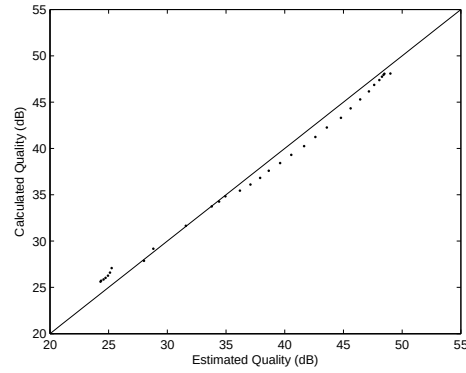
The accuracy of the quality evaluation under both $[ls, hs]$ and $[ls, hs']$ are evaluated and will be shown in the following sections. Note that $\overline{Q_{wimg}}$ is an averaged quality

value. It will be normal if some quality points smaller than $\overline{Q_{wimg}}$ are shown in the accuracy evaluation figures in the following sections.

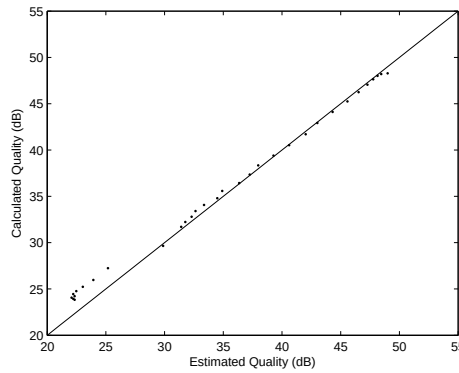
An example of the quality evaluation and accuracy evaluation in terms of PSNR under Gaussian low-pass filtering is shown in Fig. 6.3. The watermarked images in Fig. 6.2 (b), (d) and (f) are distorted under Gaussian low-pass filtering with the standard deviation, $\sigma_{filter} = [0.1 : 0.1 : 0.3, 0.31 : 0.01 : 0.5, 0.55 : 0.1 : 1.5]$. Then, the distorted



(a) Quality evaluation for the image White flower under different distortion strengths. $MAE|_{hs} = 0.7736$ dB. $MAE|_{hs'} = 0.5850$ dB



(b) Quality evaluation for the image Barbara under different distortion strengths. $MAE|_{hs} = 0.8294$ dB. $MAE|_{hs'} = 0.8274$ dB.



(c) Quality evaluation for the image Baboon under different distortion strengths. $MAE|_{hs} = 0.7875$ dB. $MAE|_{hs'} = 0.5960$ dB.

Figure 6.3: Example of the image quality evaluation under Gaussian low-pass filtering.

watermark is extracted and is compared with the original watermark. The TDR values

calculated using Equ. 5.19 is used to evaluate the quality of the distorted images. The results are shown in Fig. 6.3. In the figure, the x -axis is the evaluated quality obtained using the proposed scheme. The y -axis is the calculated quality computed using the distorted watermarked image and the original image. Thus, the quality loss caused by both the watermark embedding process and the distortions is involved in the accuracy evaluation. The calculated quality values are used to evaluate the accuracy of the quality evaluation. The black points are achieved by distort the watermarked images using different distortion strengths. We name these points the quality points. The solid line is the matching line with a slope of 1. The distances between the quality points and the solid line indicate the accuracy of the quality evaluation. The more converged the quality points to the solid line, the more accurate the quality evaluation. If a quality point with coordinates (x, y) falls on the solid line, the coordinates $x = y$. In other words, for the quality points located on the solid line, the evaluated quality equals to the calculated quality and the quality evaluation achieves the highest accuracy.

The MAE calculated with both $ds \in [ls, hs]$ and $ds \in [ls, hs']$ are listed with the figures. The standard deviation of the low-pass filter, σ_{filter} , is selected from $[0.31, 0.5]$ more intensely because $\overline{Q_{wimg}}|_{\sigma=0.5} = 35$ dB in PSNR. We are more interested in evaluating the accuracy at the better qualities.

6.2.2 The empirical grouping thresholds

The grouping thresholds used in Equ. (5.5) can be empirically set up using any empirical thresholds. Inappropriate assignment of the grouping thresholds may reduce the accuracy of the proposed scheme. To achieve the possibly highest accuracy, these grouping thresholds are chosen by testing the first 50 images in the image library for every set of experiment using the following steps:

1. Based on the researches done in [20], set up the empirical A_{wb} as listed in Equ. (5.6).
2. With each set of A_{wb} , test the quality evaluation error using the 50 images.
3. For every test image, find the lowest quality evaluation error and the corresponding A_{wb} .
4. Calculate the complexity values for the 50 images using the local intensity difference threshold 0.17. The maximum intensity values for the images is 1.
5. According to the tested A_{wb} for the 50 images, categorize the calculated complexity values into 6 groups.
6. Set up the empirical thresholds according to the categorized complexity values.

For different sets of the experiments, the empirical grouping thresholds may be different. The thresholds used in this paper are listed in the Table 6.1. As found in the experiments, under Gaussian noise distortion and a fixed noise standard deviation, σ , the degraded quality of different distorted images vary very limitedly no matter how complex the images are. Thus, under Gaussian noise distortion, we use the same grouping thresholds as those used under JPEG compression for convenience.

All these thresholds are tested in the program development stage and are fixed in both the proposed watermark embedder and extractor. They do not cause any burden on transmission or computational efficiency.

Table 6.1: The empirical grouping thresholds used for the 16 sets of experiments.

	PSNR/wPSNR	JND	SSIM/VIF
JPEG	[0.65, 0.53, 0.42, 0.34, 0.3]	[0.8, 0.6, 0.46, 0.36, 0.3]	[0.8, 0.6, 0.45, 0.25, 0.2]
JPEG2000	[0.65, 0.53, 0.42, 0.34, 0.3]	[0.8, 0.6, 0.46, 0.36, 0.3]	[0.8, 0.6, 0.45, 0.25, 0.2]
Low-pass filtering	[0.78, 0.68, 0.5, 0.36, 0.3]	[1, 1, 0.5, 0.2, 0]	[1, 0.7, 0.46, 0.36, 0.2]
Gaussian noise	[0.65, 0.53, 0.42, 0.34, 0.3]	[0.8, 0.6, 0.46, 0.36, 0.3]	[0.8, 0.6, 0.45, 0.25, 0.2]

6.3 Experimental results and accuracy evaluation

In this section, the experimental results of image quality evaluation in terms of PSNR, wPSNR, Watson JND, SSIM and VIF achieved under JPEG compression, JPEG2000 compression, Gaussian low-pass filtering and Gaussian noise pollution are presented. Along with the evaluated quality values, the quality of the distorted watermarked images are also calculated comparing to the original images. These calculated quality values are used to evaluate the accuracy of the proposed scheme. The quality loss caused by both the watermark embedding process and the distortions is involved in the accuracy evaluation. The accuracy of the quality evaluation with both $ds \in [ls, hs]$ and $ds \in [ls, hs']$ will be evaluated using *MAE*, Pearson correlation coefficient and *RMSE*, respectively.

Under JPEG compression, the watermarked images are compressed using the quality factor from 100 to 5 with a step of -5 , where the quality factor is a JPEG compression parameter used in Matlab and indicates the compression strength. The lower the quality factor, the higher the compression strength. We write the quality factor as $QF = [100 : -5 : 5]$. The degraded quality of the JPEG compressed watermarked images are evaluated comparing to the original images.

Under JPEG2000 compression, the watermarked images are compressed using the compression rate $CR = [1 : -0.05 : 0.8, 0.7 : -0.1 : 0.1, 0.09 : -0.02 : 0.01]$. The

smaller the compression rate, the stronger the compression strength. The quality of the JPEG2000 compressed images are evaluated comparing to the original images.

0.0113	0.0838	0.0113
0.0838	0.6193	0.0838
0.0113	0.0838	0.0113

Figure 6.4: A 3×3 Gaussian filter mask with standard deviation, $\sigma = 0.5$.

Under Gaussian low-pass filtering, the proposed scheme is tested by low-pass filtering the 100 watermarked images using a set of 3×3 rotationally symmetric Gaussian low-pass filters. The strength of the low-pass filtering is controlled by the standard deviation of the filter mask: $\sigma_{filter} \in [0.1, 1.5]$. The bigger the standard deviation, the stronger the low-pass filtering effect. The quality of the low-pass filtered watermarked image are evaluated comparing to the original images. An example of one Gaussian filter mask is illustrated in Fig. 6.4.

Under Gaussian noise distortion, the distortion strength is controlled by the standard deviation of the Gaussian noise, σ . The mean of the Gaussian noise is set as 0. The 100 watermarked images are tested under the noise distortion with $\sigma = [0 : 0.5 : 15]$. The quality of the noisy watermarked images are evaluated comparing to the original images.

Under each distortion, the quality of the distorted watermarked images are evaluated in terms of PSNR, wPSNR, Watson JND, SSIM and VIF, respectively. Thus, totally 20 sets of experimental results are presented in this section. In each set of experiment, one “Ideal Mapping Curve” is generated.

6.3.1 Quality evaluation in terms of PSNR under different distortions

As mentioned in the previous chapter, the “Ideal Mapping Curve” experimentally defines the relationship between the degraded quality of the distorted images and the degradation of the watermark which is denoted using TDR defined in Equ. 5.19. Thus, it is a TDR-Quality curve. The “Ideal Mapping Curves” shown in the following subsections are generated using the distortion strengths $ds \in [ls, hs]$.

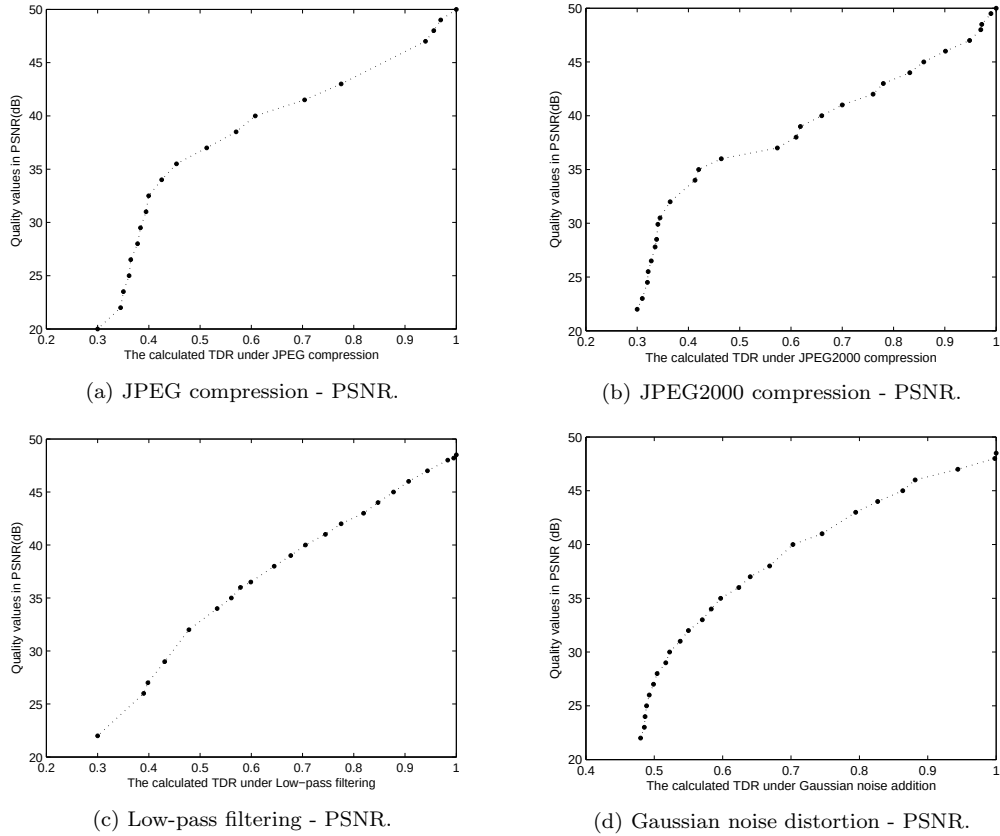


Figure 6.5: The generated “Ideal Mapping Curves” in terms of PSNR under JPEG compression, JPEG2000 compression, Gaussian low-pass filtering and Gaussian noise pollution.

The “Ideal Mapping Curves” generated in terms of PSNR under JPEG compression,

JPEG2000 compression, Gaussian low-pass filtering and Gaussian noise pollution are shown in Fig. 6.5(a)-(d), respectively. In the figure, the y axis is the calculated TDR values and the x axis the quality values in terms of PSNR. As shown in the figure, with the decrease of the quality factor, the calculated PSNR value and the calculated TDR value decrease monotonically.

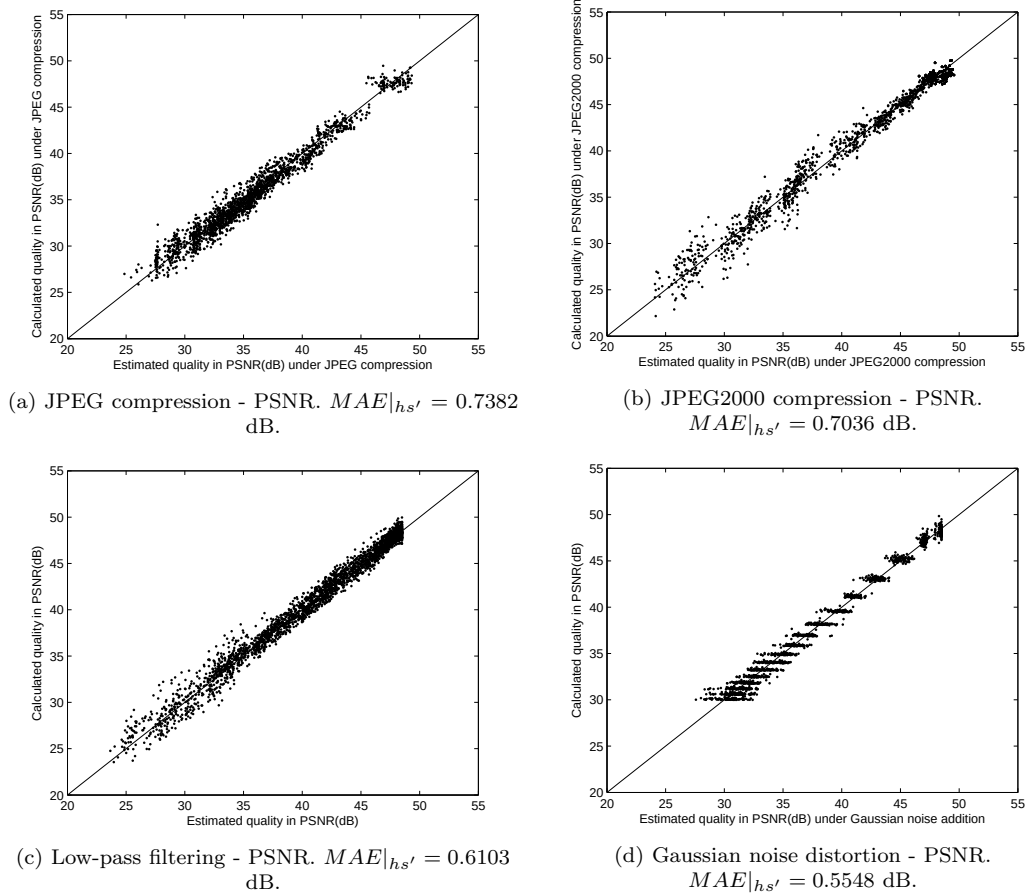


Figure 6.6: Experimental results of image quality evaluation in terms of PSNR under different distortions with distortions strengths $ds \in [ls, hs']$.

In Fig. 6.5(a), there is a slope change point at $TDR = 0.4$ and $Q|_{TDR=0.4} = 32.5$ dB. When TDR decreases from 0.4 to 0.345, Q drops from 32.5 dB to 22 dB. This means we need to use a TDR value varying in a 0.055 value range to evaluate the PSNR quality

varying in a 10.5 dB range. In this case, a small variation of the TDR value may cause relatively bigger quality evaluation error. Thus, it can be predicted that, when the calculated quality $Q < 32.5$ dB, the quality points will be more scattering. Similar analysis can be made to Fig. 6.5(b) and (d).

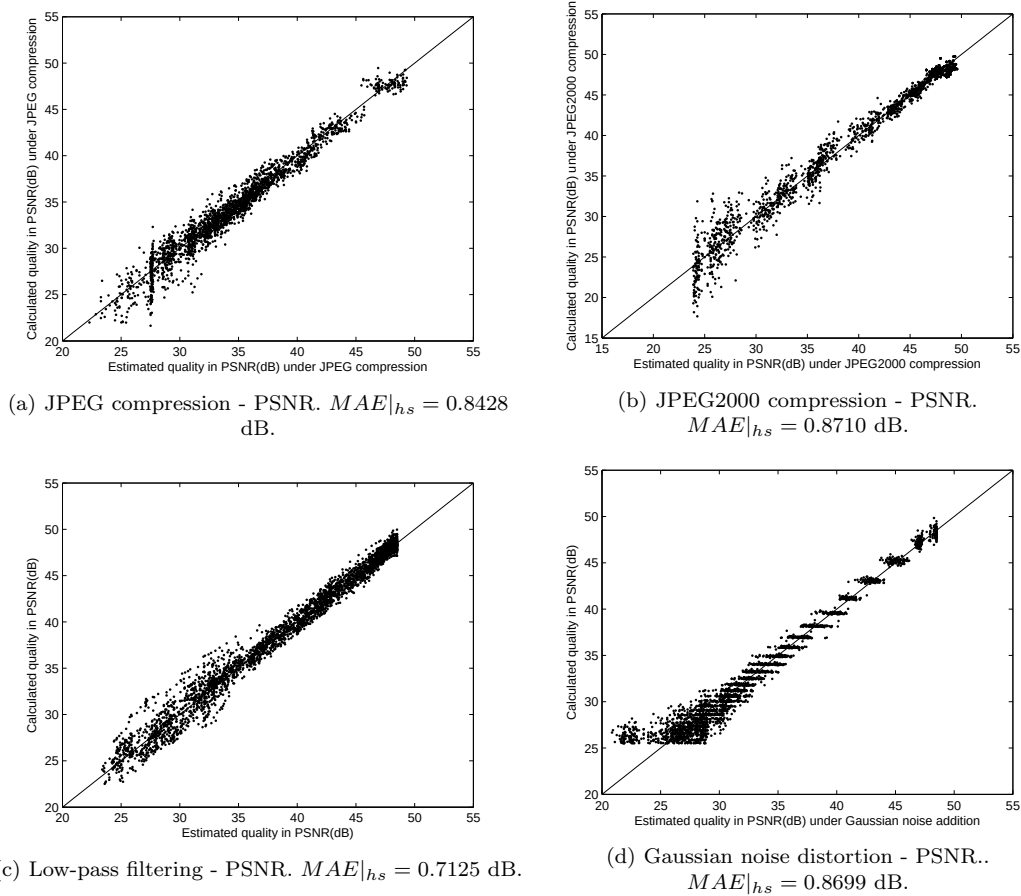


Figure 6.7: Experimental results of image quality evaluation in terms of PSNR under different distortions with distortions strengths $ds \in [ls, hs]$.

The experimental results of the image quality evaluation in terms of PSNR achieved under different distortions are shown in Fig. 6.6 (a)-(d). These results are tested with the distortion strength $ds \in [ls, hs']$. The correspondingly calculated MAE values are listed with the results. The accuracy of the experimental results assessed using the

MAE, Pearson correlation coefficient and RMSE is listed in Table 6.2.

Table 6.2: Accuracy of the experimental results in terms of PSNR achieved under different distortions.

	MAE (dB)		$Corr_p$		$RMSE$	
	hs'	hs	hs'	hs	hs'	hs
JPEG	0.7382	0.8428	0.9830	0.9801	0.9223	1.1126
JPEG2000	0.7036	0.8710	0.9914	0.9890	0.9359	1.2709
Low-pass filtering	0.6103	0.7125	0.9922	0.9916	0.7976	0.9616
Gaussian noise	0.5548	0.8699	0.9923	0.9827	0.8034	1.7133

The experimental results tested with the distortion strengths $ds \in [ls, hs]$ are presented in Fig. 6.7 as well. It is clearly shown in the figure that when $Q < 30$ dB, the quality points becomes more scattering. Thus, it is normal if the evaluated accuracy of the experimental results in Fig. 6.7 is lower than that of the results in Fig. 6.6. The accuracy of the results in Fig. 6.7 is evaluated using the MAE, the Pearson correlation, $Corr_p$, and RMSE, and is listed in Table 6.2 as well.

Table 6.3: The average quality (dB) of the distorted images in terms of PSNR under interested distortion strengths.

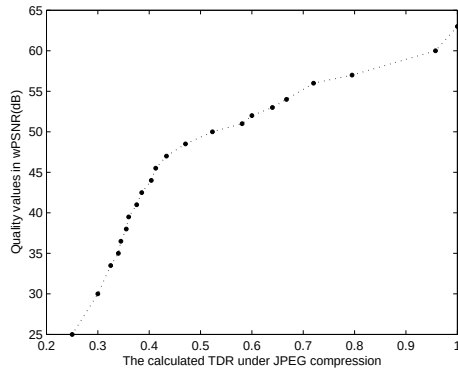
	$Q _{JPEG}$	$Q _{JPEG2000}$	$Q _{filtering}$	$Q _{noise}$
ls	47.7092	48.0957	48.3144	48.1919
hs'	30.2940	30.1790	29.7629	30.0929
hs	25.0870	24.4588	27.7597	24.7533

From the evaluated accuracy, we can see that the MAE values of the results shown in Fig. 6.6 are lower than 0.74 dB and the MAE values of the results in Fig. 6.7 are lower than 0.88 dB. All the Pearson correlations are higher than 0.98. It can be concluded that the proposed adaptive watermarking and tree structure based quality evaluation scheme can be used to evaluation image quality in terms of PSNR under

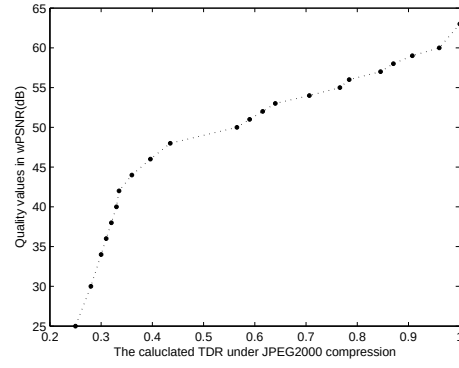
JPEG compression, JPEG2000 compression, Gaussian low-pass filtering and Gaussian noise pollution with good accuracy.

As additional observations, the averaged quality of the distorted watermarked images tested in terms of PSNR under the four distortions at $ds = ls$, $ds = hs'$ and $ds = hs$ are listed in Table 6.3 for reference.

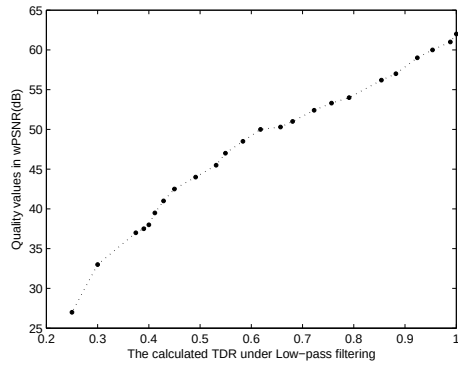
6.3.2 Quality evaluation in terms of wPSNR under different distortions



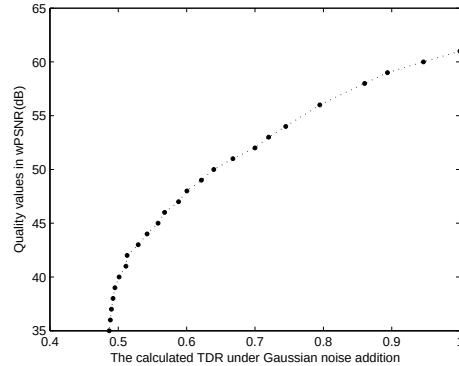
(a) JPEG compression - wPSNR.



(b) JPEG2000 compression - wPSNR.



(c) Low-pass filtering - wPSNR.



(d) Gaussian noise distortion - wPSNR.

Figure 6.8: The generated “Ideal Mapping Curves” in terms of wPSNR under JPEG compression, JPEG2000 compression, Gaussian low-pass filtering and Gaussian noise pollution.

The “Ideal Mapping Curves” generated to evaluate image quality in terms of wPSNR under JPEG compression, JPEG2000 compression, Gaussian low-pass filtering and Gaussian noise pollution are shown in Fig. 6.8(a) - (d), respectively. As shown in the figure, with the increase of the distortion strength, both the TDR values and the quality values drop monotonically. Similar to Fig. 6.5, the curves in Fig. 6.8(a), (b) and (d) has bigger slopes at the left ends. Thus, for JPEG compression, JPEG2000 compression and Gaussian noise pollution, the tested quality points at the left ends will be more scattering.

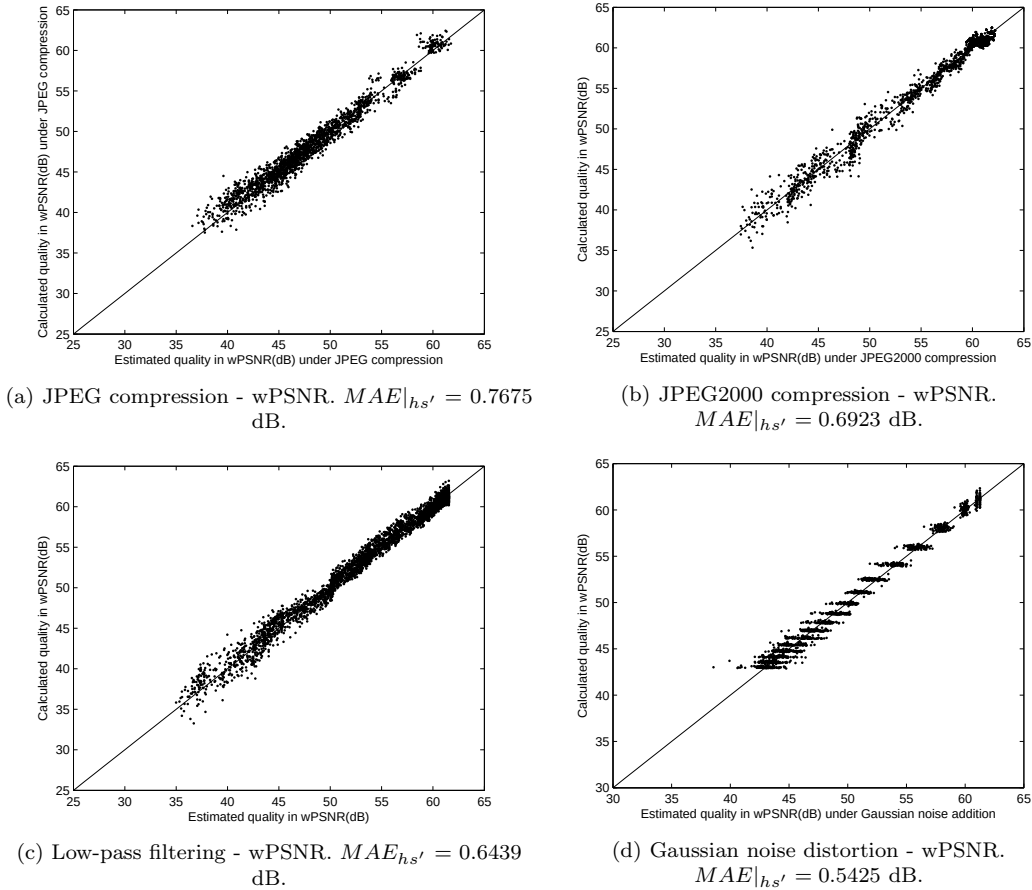


Figure 6.9: Experimental results of image quality evaluation in terms of wPSNR under different distortions with distortions strengths $ds \in [ls, hs']$.

The experimental results of the image quality evaluation in terms of wPSNR achieved under different distortions are shown in Fig. 6.9 (a) - (d). These results are tested with the distortion strength $ds \in [ls, hs']$. The correspondingly calculated MAE values are listed with the results.

The experimental results tested with the distortion strengths $ds \in [ls, hs]$ are presented in Fig. 6.10 as well. It is clearly shown in the figure that when $Q < 45$ dB, the

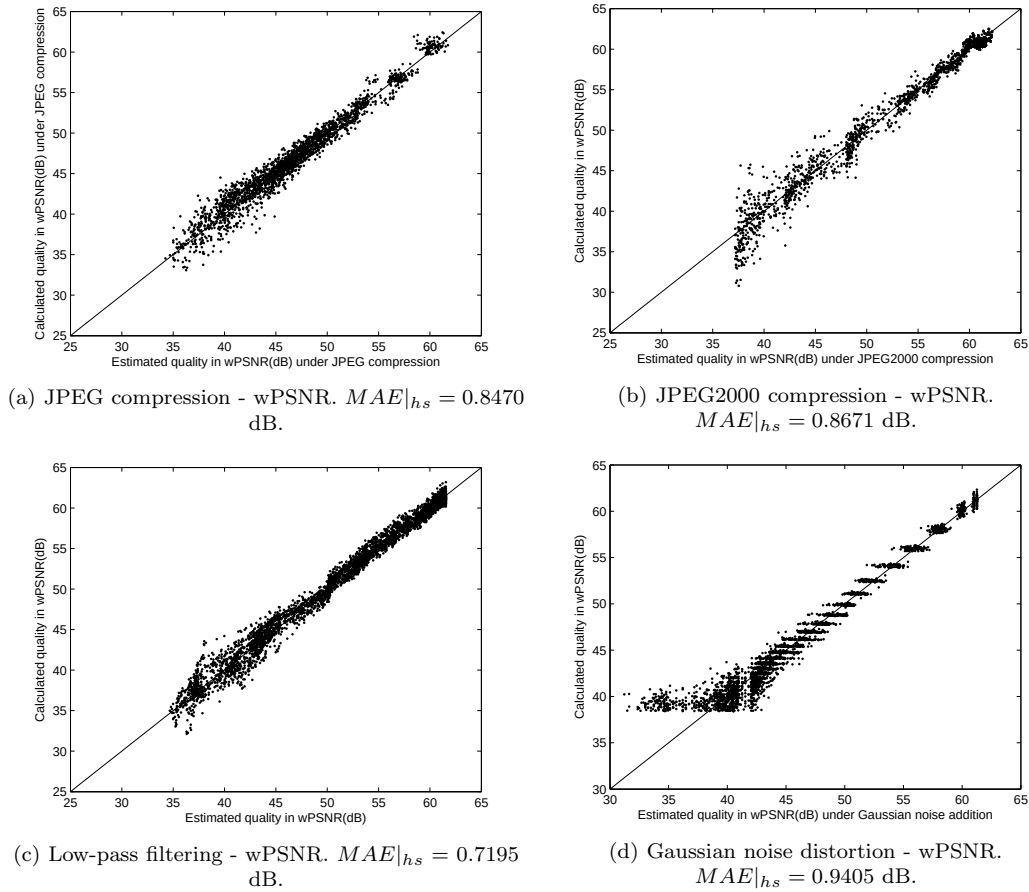


Figure 6.10: Experimental results of image quality evaluation tested in terms of wPSNR under different distortions with distortions strengths $ds \in [ls, hs]$.

quality points becomes more scattering. Thus, it is normal if the evaluated accuracy of the experimental results in Fig. 6.10 is lower than that of the results in Fig. 6.9.

The accuracy of the experimental results in Fig. 6.9 and Fig. 6.10 is evaluated using the MAE, the Pearson correlation, $Corr_p$, and RMSE, and is listed in Table 6.4.

Table 6.4: Accuracy of the experimental results in terms of wPSNR achieved under different distortions.

	MAE (dB)		$Corr_p$		$RMSE$	
	hs'	hs	hs'	hs	hs'	hs
JPEG	0.7675	0.8470	0.9834	0.9825	0.9807	1.1099
JPEG2000	0.6923	0.8671	0.9914	0.9888	0.9095	1.2691
Low-pass filtering	0.7125	0.7195	0.9916	0.9929	0.6620	0.9422
Gaussian noise	0.5245	0.9405	0.9925	0.9772	0.8467	1.5765

From the evaluated accuracy, we can see that the MAE values of the results shown in Fig. 6.9 are lower than 0.77 dB and the MAE values of the results in Fig. 6.10 are lower than 0.95 dB. All the Pearson correlations are higher than 0.98. All these results indicate that the proposed scheme has similar accuracy in estimating image quality in PSNR and wPSNR. Finally, we can conclude that the proposed adaptive watermarking and tree structure based quality evaluation scheme can be used to evaluate image quality in terms of wPSNR under JPEG compression, JPEG2000 compression, Gaussian low-pass filtering and Gaussian noise pollution with good accuracy.

Table 6.5: The average quality (dB) of the distorted images in terms of wPSNR under interested distortion strengths.

	$Q _{JPEG}$	$Q _{JPEG2000}$	$Q _{filtering}$	$Q _{noise}$
ls	60.6581	60.9263	61.2034	61.0705
hs'	42.1364	43.0372	41.0301	43.0292
hs	37.1933	37.4423	39.9028	38.5926

As additional observations, the averaged quality of the distorted watermarked images tested in terms of wPSNR under the four distortions at $ds = ls$, $ds = hs'$ and

$ds = hs$ are listed in Table 6.5 for reference.

6.3.3 Quality evaluation in terms of Watson JND under different distortions

The “Ideal Mapping Curves” generated to evaluate image quality in terms of JND under the JPEG compression, JPEG2000 compression, Gaussian low-pass filtering and Gaussian noise pollution are shown in Fig. 6.11 (a) - (d), respectively. With the

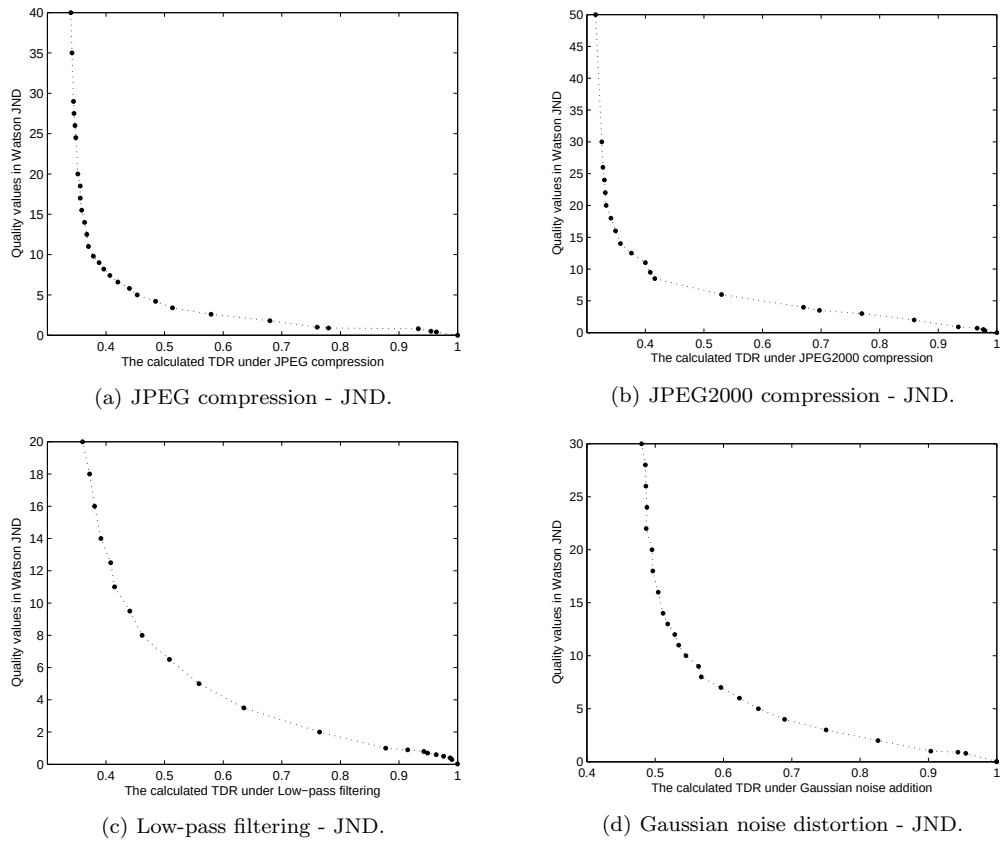
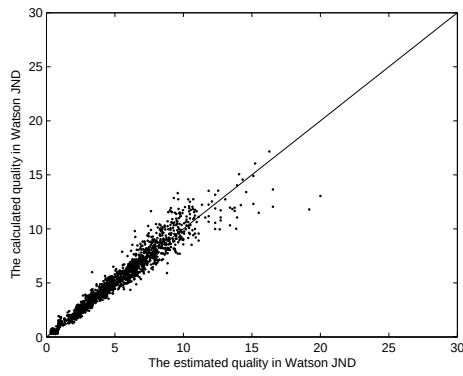


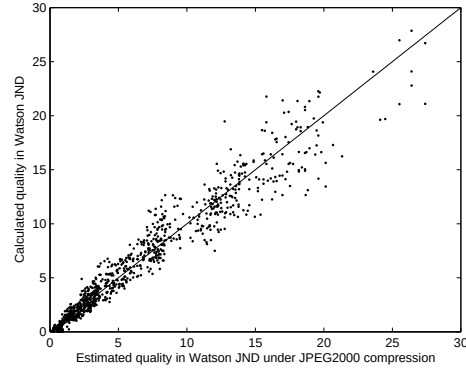
Figure 6.11: The generated “Ideal Mapping Curves” in terms of Watson JND under JPEG compression, JPEG2000 compression, Gaussian low-pass filtering and Gaussian noise pollution.

increasing of the distortion strengths, the calculated TDR decreases and the calculated

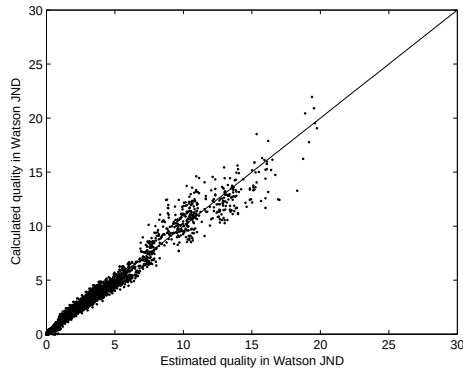
JND value increases. When TDR is 1, the JND is 0. In Fig. 6.11 (a), $TDR = 0.4$ is a slope change point and $Q|_{TDR=0.4} \doteq 8$ JND, where Q is the quality. When $TDR < 0.4$, the JND value increases rapidly with the decreasing of the TDR value. When TDR decreases from 0.4 to 0.31, the JND of the distorted images increases from 8 to 40. Thus, we need to use a TDR value varying in a 0.09 value range to evaluate the JND quality varying in a 32 value range. In this case, a small variation of the TDR value may result in a big quality evaluation error. Therefore, we can predict that this rapid increase of



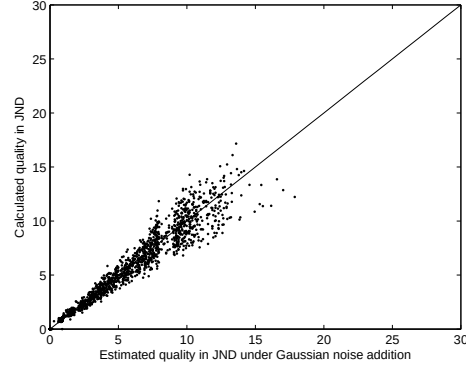
(a) JPEG compression - JND. $MAE|_{hs'} = 0.4991$ JND. (When $JND \leq 8$, $MAE = 0.3892$ JND.)



(b) JPEG2000 compression - JND. $MAE|_{hs'} = 0.7456$ JND. (When $JND \leq 8$, $MAE = 0.5348$ JND.)



(c) Low-pass filtering - JND. $MAE|_{hs'} = 0.3833$ JND. (When $JND \leq 8$, $MAE = 0.2715$ JND.)



(d) Gaussian noise distortion - JND. $MAE|_{hs'} = 0.6713$ JND. (When $JND \leq 8$, $MAE = 0.4412$ JND.)

Figure 6.12: Experimental results of image quality evaluation in terms of Watson JND under different distortions with distortions strengths $ds \in [ls, \mathbf{hs}]$.

JND will greatly reduce the accuracy of the quality evaluation and the quality points whose $TDR < 0.4$ will be very scattering. In this case, the quality evaluation accuracy when the calculated quality is smaller than 8 JND is evaluated in MAE for additional observations. Similar analysis can be made to Fig. 6.11 (b), (c), (d).

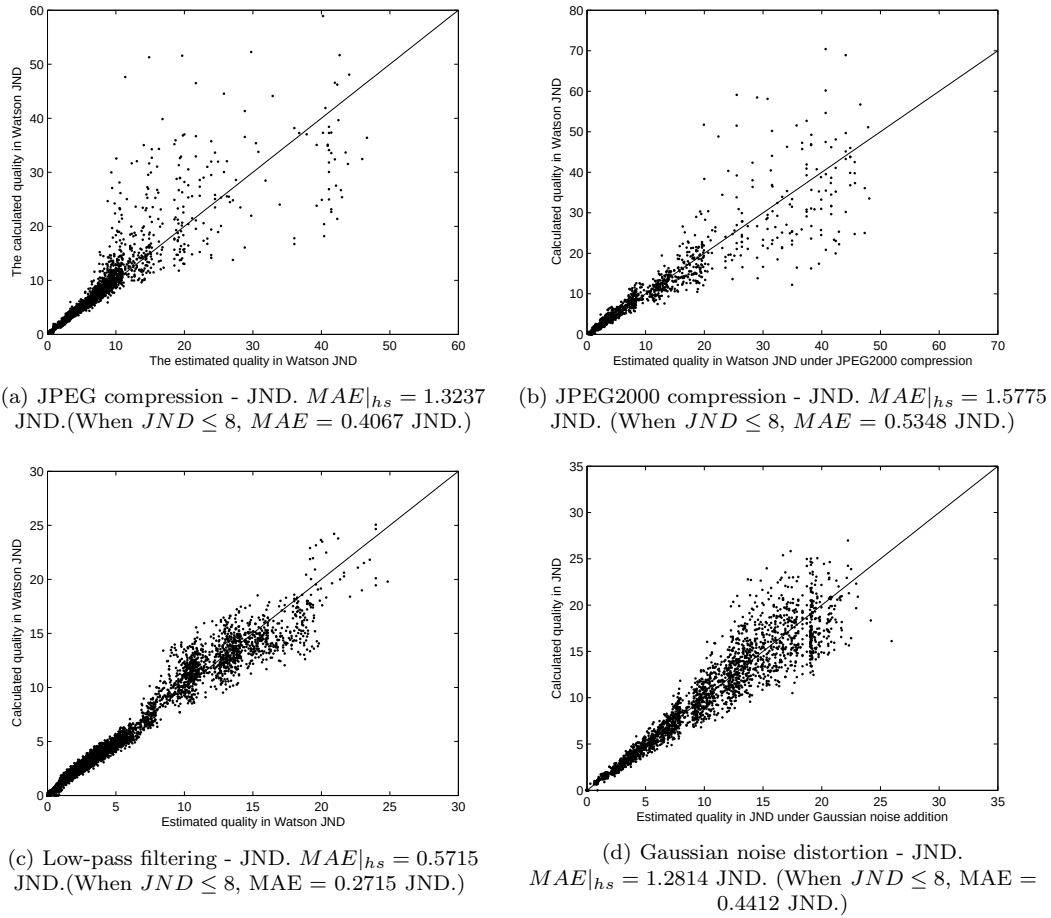


Figure 6.13: Experimental results of image quality evaluation in terms of Watson JND under different distortions with distortions strengths $ds \in [ls, hs]$.

The experimental results of the image quality evaluation in terms of Watson JND achieved under different distortions are shown in Fig. 6.12 (a) - (d). These results are tested with the distortion strength $ds \in [ls, hs']$. In each sub-figure, the MAE of all the

quality points and the MAE of the quality points whose calculated quality is smaller than 8 JND are calculated and listed with the sub-figures.

The experimental results tested with the distortion strengths $ds \in [ls, hs]$ are presented in Fig. 6.13 as well. Comparing Fig. 6.13 to Fig. 6.12, we can see that when the distortion strengths $ds \in [hs', hs]$, the achieved quality points become very scattering. Based on our experiments, we find that the more subjective the quality metric, the harder to track when the quality of the distorted images is relatively lower. The accuracy of the results in both Fig. 6.12 and Fig. 6.13 is evaluated using the MAE, the Pearson correlation, $Corr_p$, and RMSE, and is listed in Table 6.6.

Table 6.6: Accuracy of the experimental results in terms of JND achieved under different distortions.

	MAE (JND)		$Corr_p$		RMSE	
	hs'	hs	hs'	hs	hs'	hs
JPEG	0.4991	1.3237	0.9683	0.8852	0.7672	2.3352
JPEG2000	0.7456	1.5775	0.9780	0.9347	1.1156	2.6961
Low-pass filtering	0.3833	0.5715	0.9859	0.9827	0.6295	0.9877
Gaussian noise	0.6713	1.2814	0.9632	0.9485	1.0198	2.0308

From Table 6.6, we can see that the proposed scheme has the highest accuracy in estimating image quality in terms on Watson JND under Gaussian low-pass filtering and has very good accuracy under the other three distortions when $ds \in [ls, hs']$. When $ds \in [ls, hs]$, the accuracy of the results under JPEG compression, JPEG2000 compression and Gaussian noise pollution is relatively lower because the quality points achieved with $ds \in [hs', hs]$ are very scattering. However, in our experiments, we deem the proposed scheme has high accuracy if the calculated Pearson correlation is higher than 0.9 and has good accuracy if the calculated Pearson correlation locates between 0.8 and 0.9.

As additional observations, the averaged quality of the distorted watermarked images tested in terms of JND under the four distortions at $ds = ls$, $ds = hs'$ and $ds = hs$ are listed in Table 6.7 for reference.

Table 6.7: The average quality (JND) of the distorted images in terms of JND under interested distortion strengths.

	$Q _{JPEG}$	$Q _{JPEG2000}$	$Q _{filtering}$	$Q _{noise}$
ls	0.4623	0.2086	0	0
hs'	10.8677	14.5725	12.8172	11.5625
hs	31.1412	36.9529	14.6243	19.3357

6.3.4 Quality evaluation in terms of SSIM under different distortions

Fig. 6.14 (a), (b), (c) and (d) are the “Ideal Mapping Curves” used to evaluate image quality in terms of SSIM under JPEG compression, JPEG2000 compression, Gaussian low-pass filtering and Gaussian noise pollution. In Fig. 6.14 (a), the left half of the curve has a very big slope. When the TDR is changing from 0.42 to 0.31, the calculated SSIM changes from 0.9871 to 0.7. The SSIM values drop rapidly corresponding a small change of TDR value. It is straightforward that a small variation of the TDR value may cause relatively bigger quality evaluation error and the divergence of the quality points referring to the solid matching line may be magnified. The right half of the curve has a very shallow slope. When the TDR is changing from 1 to 0.42, the calculated SSIM changes from 1 to 0.9871. SSIM usually evaluates image quality in the range of $[0, 1]$. When $SSIM = 0$, the quality of the distorted image is the worst. When $SSIM = 1$, the image quality is the best. On the other hand, when the TDR is around 0.42, the quality of the distorted image is lower than 30 dB in terms of PSNR as shown in Fig.

6.5 (a) and is about 8 JND as shown in Fig. 6.11 (a). Thus, SSIM appears relatively robust in the quality evaluation.

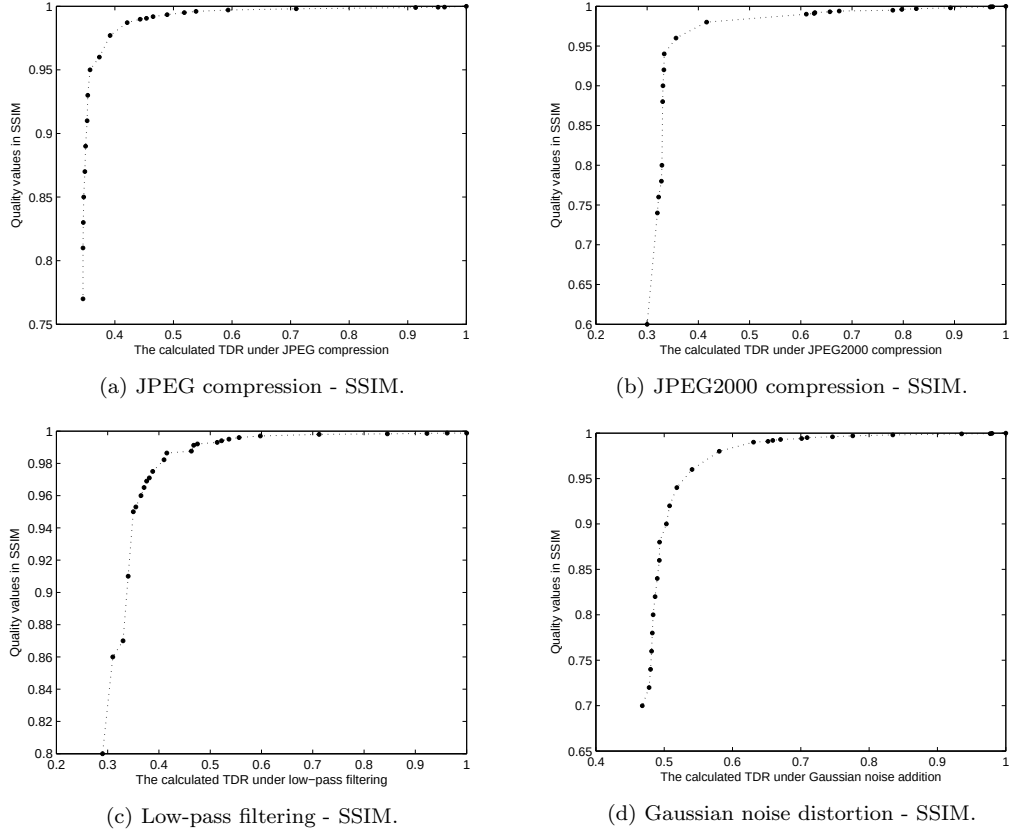


Figure 6.14: The generated “Ideal Mapping Curves” in terms of SSIM under JPEG compression, JPEG2000 compression, Gaussian low-pass filtering and Gaussian noise pollution.

The experimental results of the quality evaluation in terms of SSIM under JPEG compression, JPEG2000 compression, Gaussian low-pass filtering and Gaussian noise pollution are presented in Fig. 6.15 (a) - (d), respectively. As shown in the figures, all the quality points in terms of SSIM locate in the range of $[0.95, 1]$ under the JPEG, JPEG2000 compression and Gaussian low-pass filtering. Under Gaussian noise pollution, all the quality points locate in $[0.9, 1]$. The SSIM evaluates image quality by

comparing the structure features of both the distorted image and the original image. Under the 4 types of distortions, when the quality of the distorted watermarked images drops from 48 dB to 25 dB in PSNR or 0 JND to 20 JND or 1 to 0.3 VIF, the SSIM provides quality scores varying from 1 to 0.95. This may be caused by the structure information used by the SSIM. Generally speaking, the SSIM is a good and popular quality metric that can provide quality evaluation highly correlated to subjective quality.

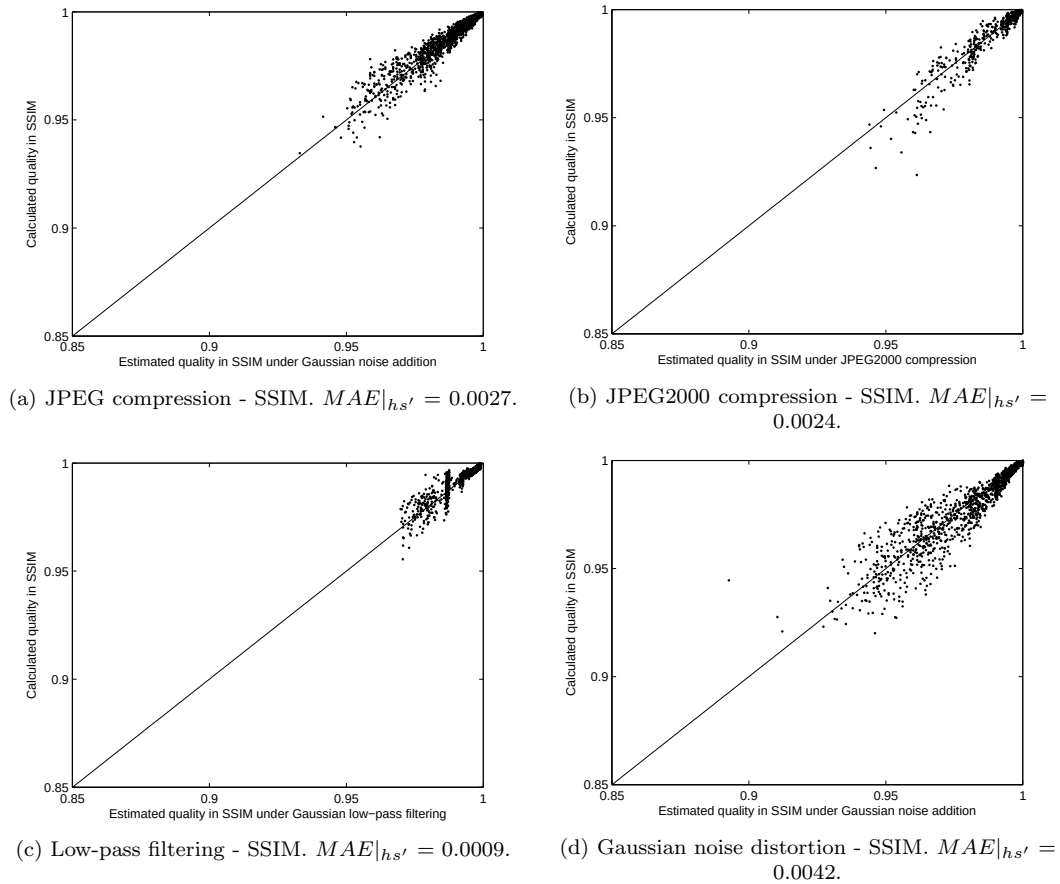


Figure 6.15: Experimental results of image quality evaluation in terms of SSIM under different distortions with distortions strengths $ds \in [ls, hs']$.

The accuracy of the results achieved with both $ds \in [ls, hs']$ and $ds \in [ls, hs]$ is

evaluated in both MAE, Pearson correlation and RMSE, and is listed in Table 6.8. From the calculated Pearson correlation, we can confirm that the proposed scheme has high accuracy to evaluate image quality in terms of SSIM under JPEG compression, JPEG2000 compression, Gaussian low-pass filtering and Gaussian noise pollution with distortions strengths $ds \in [ls, hs']$ and good accuracy under the distortions with $ds \in [ls, hs]$.

Table 6.8: Accuracy of the experimental results in terms of SSIM achieved under different distortions.

	MAE		$Corr_p$		RMSE	
	hs'	hs	hs'	hs	hs'	hs
JPEG	0.0027	0.0082	0.9513	0.8723	0.0038	0.0289
JPEG2000	0.0024	0.0125	0.9609	0.8363	0.0035	0.0308
Low-pass filtering	0.0009	0.0017	0.9420	0.9291	0.0008	0.0056
Gaussian noise	0.0042	0.0298	0.9331	0.8209	0.0066	0.0500

As additional observations, the averaged quality of the distorted watermarked images tested in terms of SSIM under the four distortions at $ds = ls$, $ds = hs'$ and $ds = hs$ are listed in Table 6.9 for reference.

Table 6.9: The average quality of the distorted images in terms of SSIM under interested distortion strengths.

	$\overline{Q _{JPEG}}$	$\overline{Q _{JPEG2000}}$	$\overline{Q _{filtering}}$	$\overline{Q _{noise}}$
ls	0.9991	0.9990	0.9988	0.9990
hs'	0.9593	0.9404	0.9814	0.9322
hs	0.8119	0.7713	0.9762	0.8251

6.3.5 Quality evaluation in terms of VIF under different distortions

The “Ideal Mapping Curves” used to conduct image quality evaluations in terms of

VIF under JPEG compression, JPEG2000 compression, Gaussian low-pass filtering and Gaussian noise addition are shown in Fig. 6.16. With the increasing of the distortion strength, the TDR value monotonically decreases, so does the calculated quality values. As presented in Section 1.1.1.6, because we did not do any image enhancement in this thesis, the quality values in terms of VIF calculated or estimated for the distorted watermarked images vary in $[0, 1]$.

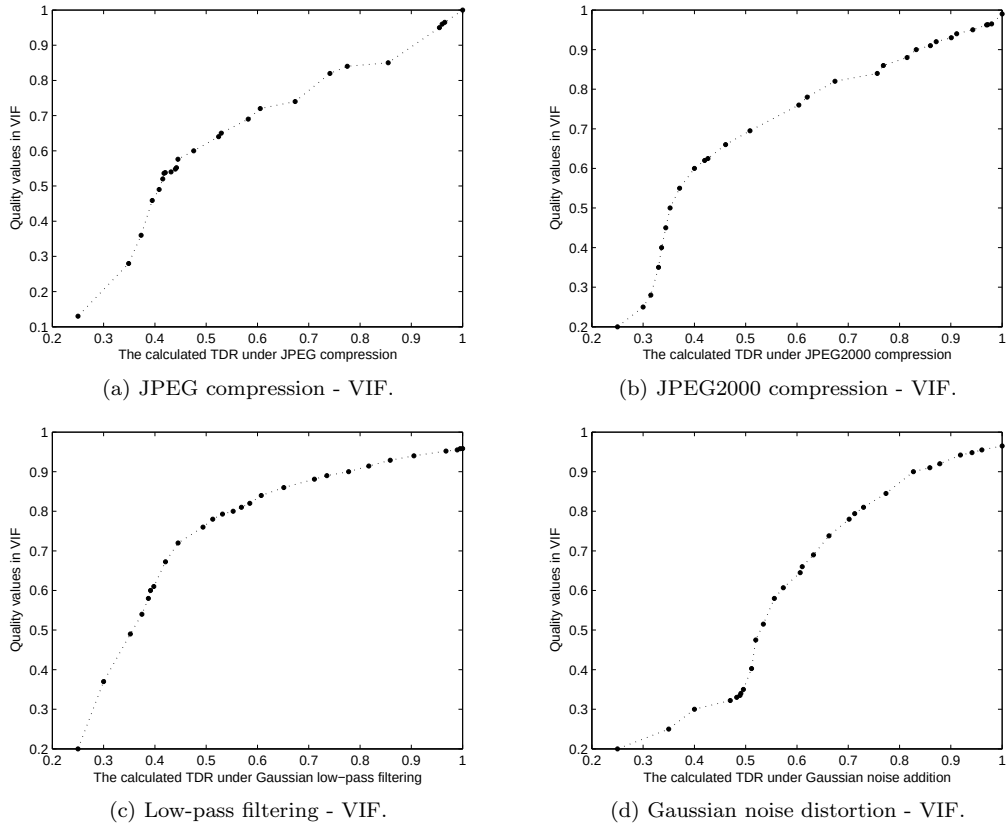
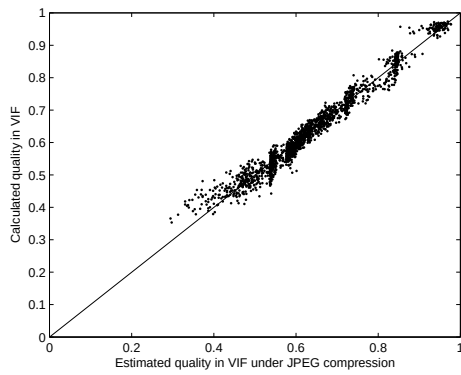


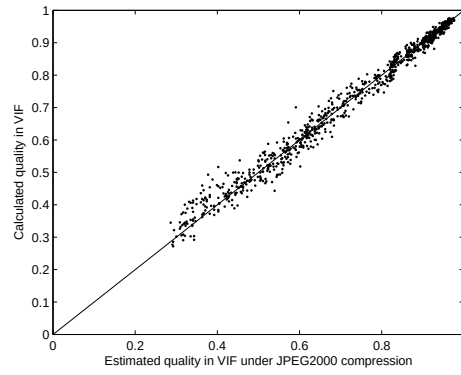
Figure 6.16: The generated “Ideal Mapping Curves” in terms of VIF under JPEG compression, JPEG2000 compression, Gaussian low-pass filtering and Gaussian noise pollution.

The experimental results of the quality evaluation under the JPEG compression, JPEG2000 compression, low-pass filtering and Gaussian noise addition with distortion

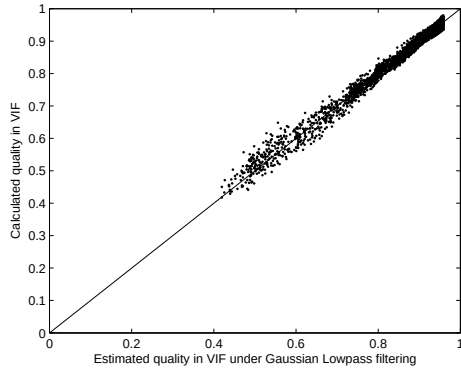
strength $s \in [ls, hs']$ are shown in Fig. 6.17. The horizontal axis is the estimated VIF quality values and the vertical axis is the calculated VIF values. From the figure we can see that the quality points are convergent to the solid quality matching line. With each set of the experimental results, the correspondingly calculated MAE is listed.



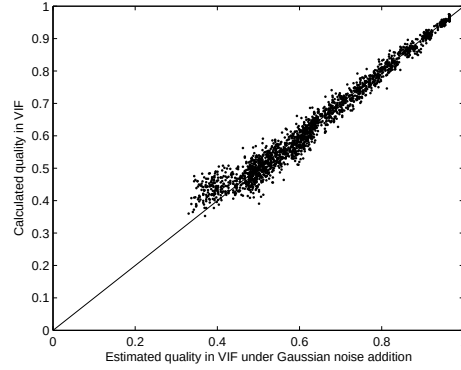
(a) JPEG compression - VIF. $MAE|_{hs'} = 0.0186$.



(b) JPEG2000 compression - VIF. $MAE|_{hs'} = 0.0125$.



(c) Low-pass filtering - VIF. $MAE|_{hs'} = 0.0112$.



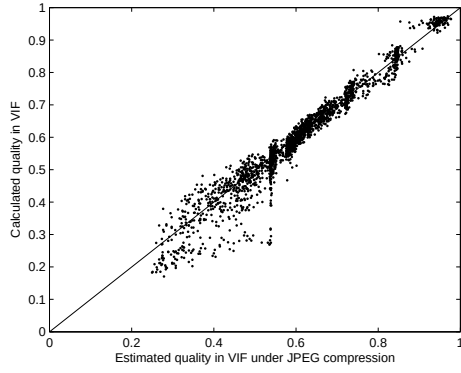
(d) Gaussian noise distortion - VIF. $MAE|_{hs'} = 0.0198$.

Figure 6.17: Experimental results of image quality evaluation in terms of VIF under different distortions with distortions strengths $ds \in [ls, hs']$.

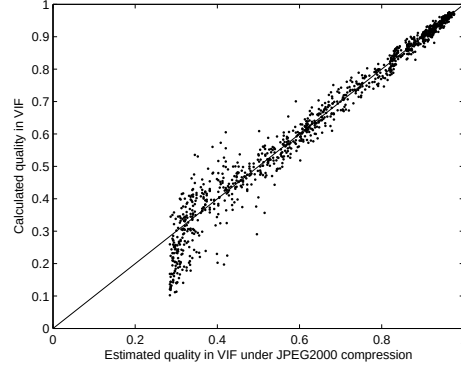
For additional observation, the experimental results obtained under the four mentioned distortions with distortion strength $ds \in [ls, hs]$ are illustrated in Fig. 6.18. The corresponding MAE is listed with the results. Comparing Fig. 6.18 to Fig. 6.17, the proposed tree structure based scheme works the most effectively under Gaussian

low-pass filtering in terms of VIF.

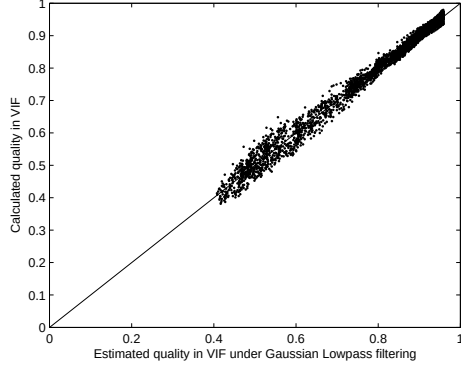
The accuracy of the results in Fig. 6.17 and Fig. 6.18 is respectively evaluated in MAE, Pearson correlation and RMSE and is listed in Table 6.10. The evaluated



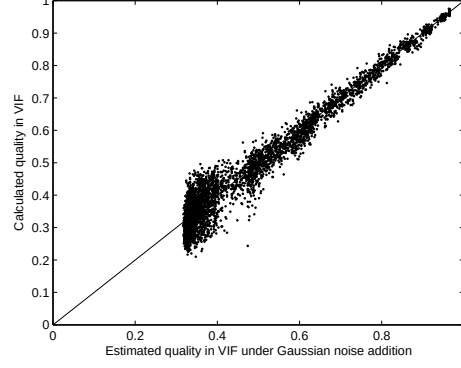
(a) JPEG compression - VIF. $MAE|_{hs} = 0.0260$.



(b) JPEG2000 compression - VIF. $MAE|_{hs} = 0.0208$.



(c) Low-pass filtering - VIF. $MAE|_{hs} = 0.0129$.



(d) Gaussian noise distortion - VIF. $MAE|_{hs} = 0.0289$.

Figure 6.18: Experimental results of image quality evaluation in terms of VIF under different distortions with distortions strengths $ds \in [ls, hs]$.

accuracy shows that the proposed scheme can be used to estimate image quality in terms of VIF with good accuracy and the VIF can be successfully used with the tree structure based scheme.

As additional observations, the averaged quality of the distorted watermarked images tested in terms of VIF under the four distortions at $ds = ls$, $ds = hs'$ and $ds = hs$

Table 6.10: Accuracy of the experimental results in terms of VIF achieved under different distortions.

	MAE		$Corr_p$		RMSE	
	hs'	hs	hs'	hs	hs'	hs
JPEG	0.0186	0.0260	0.9835	0.9720	0.0242	0.0377
JPEG2000	0.0125	0.0208	0.9950	0.9888	0.0196	0.0384
Low-pass filtering	0.0112	0.0129	0.9934	0.9946	0.0151	0.0174
Gaussian noise	0.0198	0.0289	0.9871	0.9810	0.0277	0.0387

are listed in Table 6.11 for reference.

Table 6.11: The average quality of the distorted images in terms of VIF under interested distortion strengths.

	$Q _{JPEG}$	$Q _{JPEG2000}$	$Q _{filtering}$	$Q _{noise}$
ls	0.9544	0.9617	0.9629	0.9630
hs'	0.4618	0.4560	0.5562	0.4364
hs	0.2469	0.2169	0.5207	0.2796

6.3.6 Summary of the evaluation accuracy

To give the readers an overall impression about the effectiveness of the proposed adaptive watermarking and tree structure based image quality evaluation scheme, the accuracy of the 20 sets of experimental results is summarized in this section.

The accuracy of the 20 sets of experiments evaluated using MAE is summarized in Table 6.12. The smaller the calculated MAE , the more accurate the quality evaluation. In the table, hs or hs' are used to indicate the range of the distortion strengths $[ls, hs]$ or $[ls, hs']$ and hs is the stronger distortion strength comparing to hs' . The degraded quality of the 100 distorted images under hs' is about 30 dB in average for all the four distortions. The degraded quality of the test images under hs is about 25 dB in average

for all the distortions. In Table 6.12, “LPF” indicates the Gaussian low-pass filtering and “Gnoise” means the Gaussian noise pollution.

Table 6.12: Summary of the quality evaluation accuracy in MAE.

	PSNR (dB)		wPSNR (dB)		JND		SSIM		VIF	
	hs'	hs	hs'	hs	hs'	hs	hs'	hs	hs'	hs
JPEG	0.7382	0.8428	0.7675	0.8470	0.4991	1.3237	0.0027	0.0082	0.0186	0.0260
JPEG2000	0.7036	0.8710	0.6923	0.8671	0.7456	1.5775	0.0024	0.0125	0.0125	0.0208
LPF	0.6103	0.7125	0.6439	0.7195	0.3833	0.5715	0.0009	0.0017	0.0112	0.0129
Gnoise	0.5548	0.8699	0.5245	0.9405	0.6713	1.2814	0.0042	0.0298	0.0198	0.0289

The accuracy of the 20 sets of experiments evaluated using the Pearson correlation are summarized in Table 6.13. The higher the calculated correlation, the more accurate the quality evaluation. If the calculated correlation is higher than 0.8, we think the accuracy is good.

Table 6.13: Summary of the quality evaluation accuracy in Pearson correlation coefficient.

	PSNR (dB)		wPSNR (dB)		JND		SSIM		VIF	
	hs'	hs	hs'	hs	hs'	hs	hs'	hs	hs'	hs
JPEG	0.9830	0.9801	0.9834	0.9825	0.9683	0.8852	0.9513	0.8723	0.9835	0.9720
JPEG2000	0.9914	0.9890	0.9914	0.9888	0.9780	0.9347	0.9609	0.8363	0.9950	0.9888
LPF	0.9922	0.9916	0.9929	0.9929	0.9859	0.9827	0.9420	0.9291	0.9934	0.9946
Gnoise	0.9923	0.9827	0.9925	0.9772	0.9632	0.9485	0.9331	0.8209	0.9871	0.9810

The accuracy of the 20 sets of experiments evaluated using MAE is summarized in Table 6.12. The smaller the calculated $RMSE$, the more accurate the quality evaluation.

From Table 6.12 to Table 6.14, we can see that:

1. The proposed tree structure based scheme has good accuracy to evaluate image quality in terms of PSNR and wPSNR. Under the four mentioned distortions, with

Table 6.14: Summary of the quality evaluation accuracy in RMSE

	PSNR (dB)		wPSNR (dB)		JND		SSIM		VIF	
	hs'	hs	hs'	hs	hs'	hs	hs'	hs	hs'	hs
JPEG	0.9223	1.1126	0.9807	1.1099	0.7672	2.3352	0.0038	0.0289	0.0242	0.0377
JPEG2000	0.9359	1.2709	0.9095	1.2691	1.1156	2.6961	0.0035	0.0308	0.0196	0.0384
LPF	0.7976	0.9616	0.6620	0.9422	0.6295	0.9877	0.0008	0.0056	0.0151	0.0174
Gnoise	0.8034	1.7133	0.8467	1.5765	1.0198	2.0308	0.0066	0.0500	0.0277	0.0387

both the distortion strengths $[ls, hs]$ and $[ls, hs']$, the calculated *MAE* for PSNR and wPSNR are respectively lower than 0.74 dB and 0.8710 dB. The calculated MAE for the quality estimation in wPSNR are respectively lower than 0.77 dB and 0.9405 dB. The calculated correlations are all higher than 0.977 which indicate the high correlations between the evaluated quality values and the calculated quality values.

2. To evaluate image quality in terms of JND, the proposed scheme has the highest accuracy under low-pass filtering with both the distortion strengths $[ls, hs]$ and $[ls, hs']$; good accuracy under other distortions with distortion strengths $[ls, hs']$; acceptable accuracy with distortion strengths $[ls, hs]$.
3. To evaluate image quality in terms of SSIM, the proposed scheme has good accuracy under all the four distortions with distortion strengths $[ls, hs']$ and acceptable accuracy with distortion strengths $[ls, hs]$.
4. The assessed accuracy show that the proposed scheme can be used to evaluate image quality in terms of VIF with good accuracy.

Under JPEG compression, the accuracy of the proposed tree structure based scheme is comparable to that of the quantization based scheme proposed in Chapter 4. The

quantization based quality evaluation scheme is tested by estimating image quality in terms of PSNR, wPSNR, JND and SSIM under JPEG compression and achieved high accuracy. The quality factors $[100 : -5 : 20]$ are used in the testing. The accuracy of the image quality evaluation of both the quantization based scheme and the tree structure based scheme under quality factors $[100 : -5 : 20]$ are listed in Table 6.15 for comparison. From both the calculated MAE and the calculated Pearson correlations, $Corr_p$, we can see that the two schemes has comparable accuracy.

Table 6.15: Accuracy comparison between the quantization based scheme and the tree structure based scheme.

	Proposed quantization based scheme		Proposed tree structure based scheme	
	MAE	$Corr_p$	MAE	$Corr_p$
PSNR	0.7877	0.9872	0.7382	0.9830
wPSNR	0.7967	0.9876	0.7675	0.9834
JND	0.5440	0.9714	0.4991	0.9683
SSIM	0.0027	0.9583	0.0027	0.9513

6.4 Advantages of the proposed scheme

Based on the experimental results and analysis in Section 6.3, the advantages of the proposed adaptive watermarking and tree structure based quality metric are listed in the following:

1. The computational efficiency. As mentioned in Chapter 4, the quantization based quality evaluation scheme needs an average of 12-13 recursive loops to find the optimal watermark embedding strengths for one test image. With the tree structure based scheme, the extensive iterative process is not needed and the computation efficiency is improved about 17 times over that of the quantization based scheme.

2. The quality loss caused by the watermark embedding is very small. With the tree structure based scheme, the quality of the 150 watermarked images is 48.1476 dB on average. In the quantization based scheme, the quality of the watermarked images is about 40 dB. Thus, with the tree structure based scheme, the quality degradation caused by the watermark embedding is reduced about 7 dB.
3. The watermarking-based quality metric can be used to evaluate image quality in terms of PSNR, wPSNR, Watson JND, SSIM and VIF with good accuracy. Moreover, it can be used to evaluate image quality degradation under different distortions, such as JPEG compression, JPEG2000 compression, Gaussian low-pass filtering and Gaussian noise pollution. The tree structure based scheme can be developed to evaluate image quality in terms of other possible quality measurements and or under other possible distortions.
4. It works as a Reduced- or No-Reference quality evaluation scheme which is more desirable in communication systems. With the watermarking-based quality metric, there is no need to transmit the cover signal to the receiver side for the purpose of quality evaluation. We only need to transmit the original watermark which is about 1 KB to 2 KB and a few bits of additional information. If the original watermark is known at the receiver side, the transmission bandwidth will be further saved.

Chapter 7

Video quality evaluation using SPIHT tree structure and HVS based watermarking

With good computational efficiency, the proposed framework for the watermarking based quality evaluation and the proposed SPIHT tree structure and HVS based quality evaluation scheme are further adapted for video signals as shown in Fig. 7.1 and Fig. 7.2. The watermark is embedded in the luminance components of the raw video sequence. After watermark embedding, the watermarked video signal is compressed for more efficient transmission. The H.264 is one widely used video coding standards [90][91][92]. During the transmission, the watermarked video signal may be degraded by the channel distortions like packet loss [93][94]. In this chapter, the effectiveness of the proposed tree structure based scheme will be evaluated on the I-frames and P-frames of different resolutions videos, such as CIF and 720p, under different distortions, such as H.264 compression and packet loss related distortion.

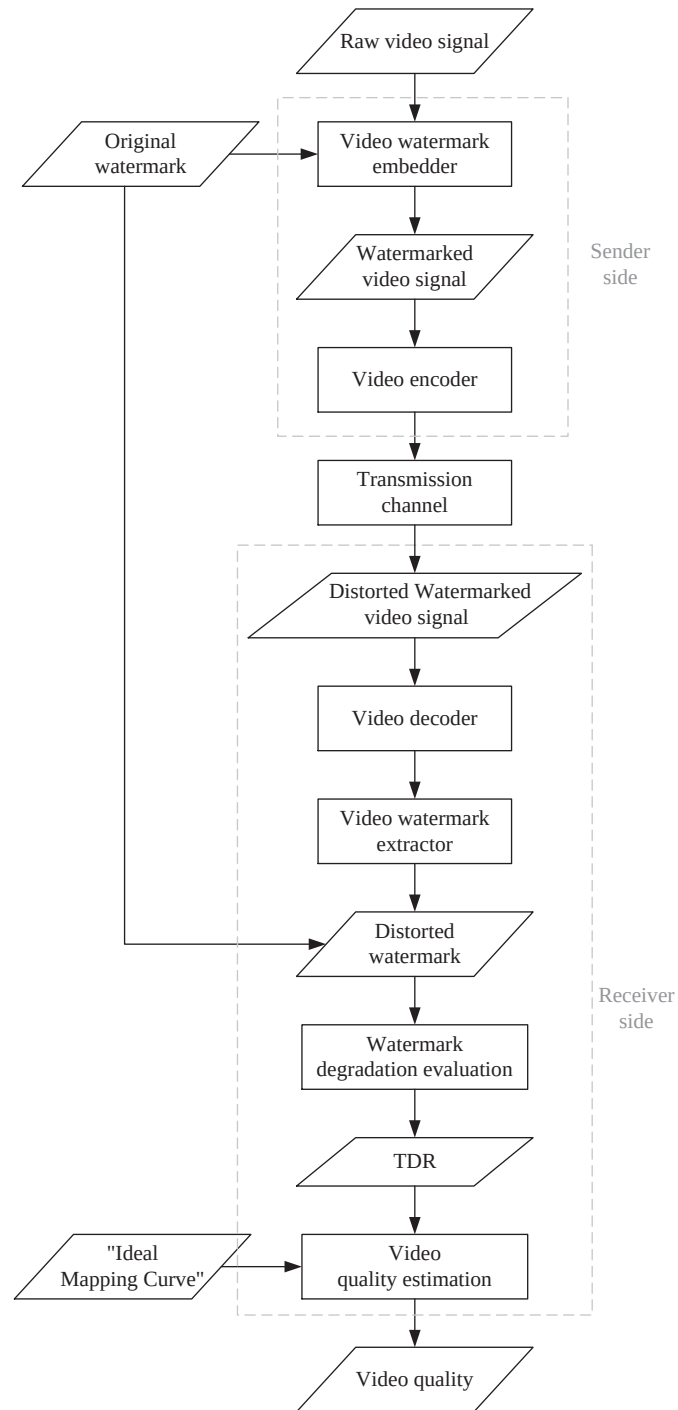


Figure 7.1: The framework for watermarking based video quality evaluation.

In the proposed scheme, to evaluate the overall quality of the video signal, the watermark is embedded in multiple selected frames of a raw video sequence. In one video frame, the watermark is embedded throughout the frame content using the strategies presented in Chapter 5. Based on the preliminary experiments, the video frames in one shot have quite similar degradation characteristics because they have very close content complexities. For these frames, the watermark is embedded using the same watermark bit assignment, A_{wb} . However, for different video frames in the same shot, the generated HVS masks may have slight differences. Thus, the bitplanes selected for the watermark embedding may be different for different video frames. Recall that the adaptive watermark embedding strength is controlled by A_{wb} and the selected bitplanes, the adaptive watermark embedding strengths may be different for different video frames in the same shot and are different for the video frames in different video shots.

At the receiver side, the multiple degraded watermarks are extracted. For each distorted watermark, its degradation is evaluated by calculating TDR referring to the original watermark. For one video frame, its quality is evaluated by mapping the corresponding TDR referring to the “Ideal Mapping Curve”. The overall quality of one video sequence is assessed by averaging the framewise quality evaluations. As reviewed in Chapter 2, most of the No-Reference video quality metrics evaluate video quality in terms of MSE or PSNR. In our proposed scheme, the quality of the distorted video signals are evaluated in terms of PSNR.

The SPIHT tree structure and HVS based video watermark embedding process is shown in Fig. 7.2. The watermark is embedded in the luminance components of the selected frames. For one video shot, the watermark bit assignment, A_{wb} is assigned by analyzing the content complexity of the luminance component of the first video frame. The adaptive watermarking and tree structure based watermark extractor shown in

Fig. 5.11 is used to extract the distorted watermarks from the luminance frames of the distorted video sequence. The YUV raw video sequences are used for the testing, where Y is the luminance component of a video frame.

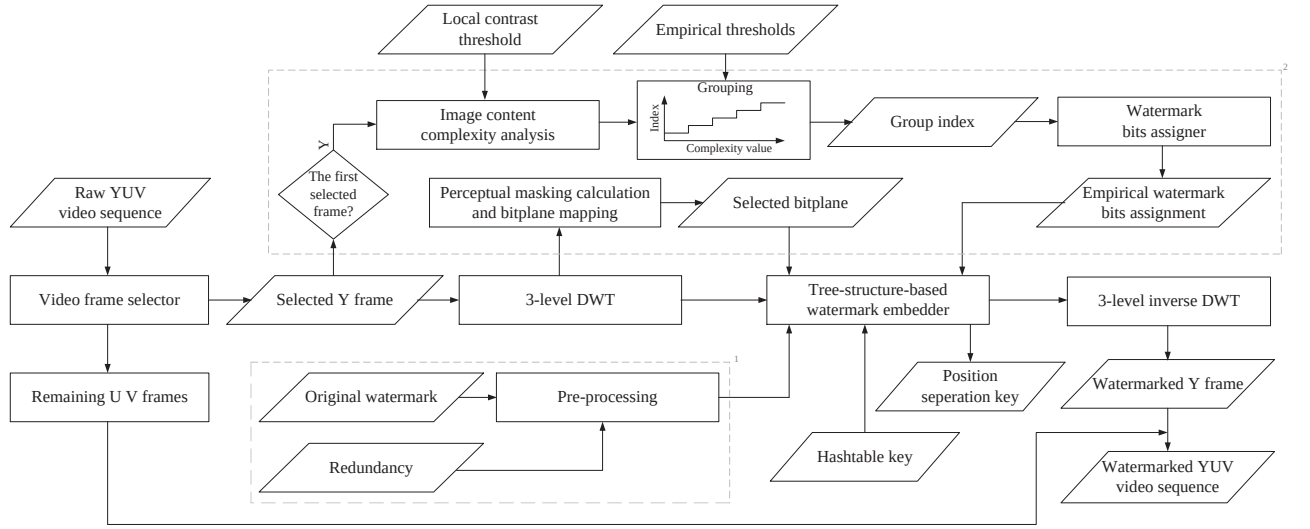


Figure 7.2: The SPIHT tree structure and HVS based video watermark embedding process.

7.1 The density of watermark embedding

To serve the purpose of quality evaluation, it is desirable to embed the redundant watermark sequence throughout an image or a video frame and any two neighboring watermarked trees are not expected to separate too far away from each other to avoid the failure of quality monitoring under local tamper distortion. In Chapter 6, we embedded the 48×48 original watermark with *Redundancy* of 3 throughout an 512×512 image, where *Redundancy* indicates the times we repeat the original watermark to obtain the redundant watermark sequence. Considering the video sequences have different spatial resolutions, such as CIF, 720p and 1080p, to make the proposed tree structure

based scheme effective for videos with different resolutions, we define the density of watermark embedding which quantifies the percentage of the trees selected for watermark embedding over the total number of tree positions available for selection across a video frame. With the definition of density of watermark embedding, watermark sequence with variable length is embedded into videos with different resolutions. Because we use the 48×48 original watermark shown in Fig. 6.1 for video quality evaluation, the variable length of the watermark sequence is adjusted by changing the value of *Redundancy* according to the video resolution. The calculation of *Redundancy* for different resolution videos will be presented in this section in details.

The density of watermark embedding is mathematically depicted using Equ. (7.1).

$$\rho_{we} = \frac{w_{segs}}{T_{NP}} \quad (7.1)$$

where, ρ_{we} is the density of the watermark embedding for an image or a video frame and $\rho_{we} \leq 1$; w_{segs} is the number of trees selected for watermark embedding and $w_{segs} = \left\lfloor \frac{Redundancy \times len}{\sum A_{wb}} \right\rfloor$ as defined in Equ. (5.1); $\sum A_{wb}$ indicates the number of watermark bits embedded in one selected tree and is defined in Equ. (5.7); $len = 48 \times 48$ is the length of the original watermark. T_{NP} is the total number of tree positions available for selection and $T_{NP} = \frac{M}{2^3} \cdot \frac{N}{2^3} \cdot \frac{3}{4} \cdot \frac{1}{3} = \frac{M \times N}{256}$ as defined in Section 5.3.2; $M \times N$ is the size of an image or video frame. Thus, the smaller the ρ_{we} , the less trees are selected for watermark embedding, and vice versa. On the other hand, $\lfloor 1/\rho_{we} \rfloor - 1$ indicates the number of trees separated between any two neighboring watermarked trees and is the position separation key, which is consistent with the definition in Equ. (5.2).

With the definitions of w_{segs} and T_{NP} , Equ. (7.1) can be further developed as:

$$\rho_{we} = \left\lfloor \frac{Redundancy \times len}{\sum A_{wb}} \right\rfloor \cdot \frac{256}{M \times N} \quad (7.2)$$

Recall Equ. (5.6) and Equ. (5.7), the watermark bit assignment we used for image quality evaluation in Chapter 6 is listed in the following:

$$A_{wb} = \begin{cases} [27 \ 0 \ 0], & Gindex = 1 \\ [19 \ 7 \ 1], & Gindex = 2 \\ [13 \ 12 \ 2], & Gindex = 3 \\ [8 \ 15 \ 4], & Gindex = 4 \\ [1 \ 16 \ 4], & Gindex = 5 \\ [0 \ 8 \ 4], & Gindex = 6 \end{cases}$$

and

$$\sum A_{wb} = \begin{cases} 27, & \text{when } Gindex \in \{1, 2, 3, 4\} \\ 21, & \text{when } Gindex = 5 \\ 12, & \text{when } Gindex = 6 \end{cases}$$

In Chapter 6, for a 512×512 image, $Redundancy = 3$. When $\sum A_{wb} = 27$, $\rho_{we} = 1/4$. When $\sum A_{wb} = 12$, $\rho_{we} = 1/2$. Thus, for a video sequence with spatial resolution $M \times N$, we set $\rho_{we} = 1/4$ when $\sum A_{wb} = 27$. This can be done by adjusting the value of $Redundancy$ according to the video resolution $M \times N$, which makes that the length of the embedded redundant watermark sequence is different for videos with different spatial resolutions. The $Redundancy$ values for different resolution video sequences calculated using Equ. (7.2) are listed in the following:

1. For the CIF video sequences, the spatial resolution is 352×288, $Redundancy = 1$.
2. For video sequences with spatial resolution of 512×512, $Redundancy = 3$.

3. For 720p video sequences, the spatial resolution is 1280×720 , *Redundancy* = 10.
4. For 1080p video sequences, the spatial resolution is 1920×1080 , *Redundancy* = 23.

By defining the density of watermark embedding, the watermark extraction is modified as the following:

$$w_e(i, j) = \begin{cases} 1, & N_1 \geq N_0 + \text{const} \\ 0, & N_1 < N_0 \end{cases} \quad (7.3)$$

and

$$\text{const} = \max \left(0, \left\lfloor \frac{\text{Redundancy}}{2} \right\rfloor - 1 \right) \quad (7.4)$$

where, $w_e(i, j)$ is the extracted watermark bit with coordinates (i, j) ; N_1 is the number of extracted 1s and N_0 is the number of extracted 0s.

Note, Equ. (7.3) is compatible with Equ. (5.18) and is consistent with the implementations in Chapter 6.

In this chapter, we will test the effectiveness of the density of watermark embedding with the tree structure based scheme on a number of different resolution video sequences in the following few sections.

7.2 Summary of the experiments on video quality evaluation

Under H. 264 compression, quality evaluation is conducted on the I-frames and IP-frames of videos with different resolution, such as 352×288 (CIF), 512×512 , 1280×640

(640p) and 1280×720 (720p). Quality evaluation is also done on both I-frames and P-frames of CIF videos under the packet loss related distortions. Thus, totally 7 sets of experimental results will be presented in this chapter.

The original watermark shown in Fig. 6.1 is used in the testing. The grouping thresholds used under JPEG/JPEG2000 compression listed in Table 6.1 are used under H.264 compression.

In experiments, twelve CIF video sequences shown in Fig. A-4 and nine 720p and 640p video sequences shown in Fig. A-5 are tested [95]. Each of these videos consists of 30 frames. In implementation, for each video, all the 30 frames are distorted and the quality evaluation is conducted. The framewise accuracy of quality evaluation is assessed using MAE_i , Pearson correlation coefficients, $Corr_{pi}$, and $RMSE_i$, where $i \in [1, 30]$. Finally, the overall quality evaluation accuracy of one distorted watermarked video sequence is assessed by averaging the framewise MAE_i , $Corr_{pi}$, or $RMSE_i$ and is denoted as $\overline{MAE}$, $\overline{Corr_p}$ and $\overline{RMSE}$, respectively.

7.3 Quality evaluation in terms of PSNR under H.264 compression

The JM H.264 codec [96] is used to compress the watermarked video sequence and reconstruct the compressed watermarked video sequence. Quantization parameter, Qp , is used to control the strength of H.264 compression. The higher the quantization parameter, the stronger the compression strength. In the JM H.264 codec, the maximum quantization parameter allowed to use is 51 for both the I- and P-frames.

Similar to the strategies presented in Section 6.2.1, the boundaries of the distortion strength is chosen as follows:

H.264 video compression: the distortion strength, ds , is the quantization parameter.

$$ls = 1. \quad \overline{Q_{wvdo}|_{ls}} = 47.8380 \text{ dB in PSNR.}$$

$$hs = 50. \quad \overline{Q_{wvdo}|_{hs}} = 25.1567 \text{ dB in PSNR.}$$

$$hs' = 41. \quad \overline{Q_{wvdo}|_{hs'}} = 29.8361 \text{ dB in PSNR.}$$

The $\overline{Q_{wvdo}}$ is calculated using all the test video sequences. The accuracy of the experimental results are evaluated using MAE, Pearson correlation and RMSE under both $ds \in [ls, hs]$ and $ds \in [ls, hs']$.

In the testing, the quantization parameters $Qp = [1, 10, 15 : 3 : 50]$ are used to compress the videos with different resolution. For each of the test video, the watermark bit A_{wb} is assigned to the whole video sequence by analyzing the content complexity of the first video frame.

7.3.1 A preliminary experiment

A preliminary experiment on video quality evaluation is conducted on the I-frames of 512×512 videos to provide additional observations on the performance of the proposed tree structure based scheme comparing to the experimental results listed in Chapter 6. In this experiment, the twelve 512×512 videos shown in Fig. A-3 are tested. In this experiment, all the watermarked video frames are intra coded under H.264 compression using the quantization parameters, $Qp = [1, 10, 15 : 3 : 50]$. Because the spatial resolution of the videos is 512×512, for convenience, the “Ideal Mapping Curve” is generated under H.264 compression by testing the first 50 gray images in the image library shown in Appendix A using the procedure presented in Section 5.6. The generated “Ideal Mapping Curve” is shown in Fig. 7.3. The x-axis is the TDR value calculated under different quantization parameters. The y-axis indicates the correspondingly calculate quality values in terms of PSNR. As illustrated in the figure, with the increasing distor-

tion strength, the watermark degrades more and more severely and the quality values monotonically decrease.

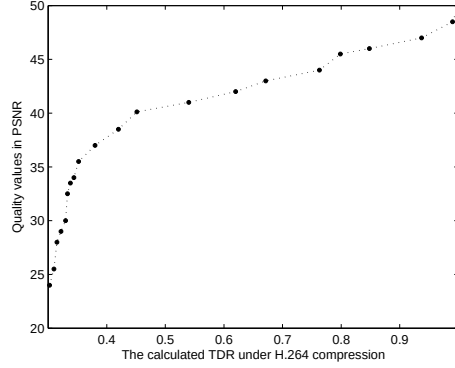


Figure 7.3: The “Ideal Mapping Curve” generated to evaluate the quality of 512×512 videos under H.264 compression.

The framewise quality evaluation tested on two sample 512×512 videos, Kimono and Spin calendar, are shown in Fig. 7.4. The watermark bit assignment for video Kimono is [8, 15, 4] and the quality evaluation results tested on the 30 video frames are shown in Fig. 7.4 (b).

The video Spin calendar and its results are shown in Fig. 7.4 (c) and (d), respectively. Associated with Fig. 7.4 (b) and (d), the $\overline{MAE}$ is calculated by averaging the MAE values of all the 30 frames and is used as a quality indicator for the whole video sequence Kimono or Spin calendar.

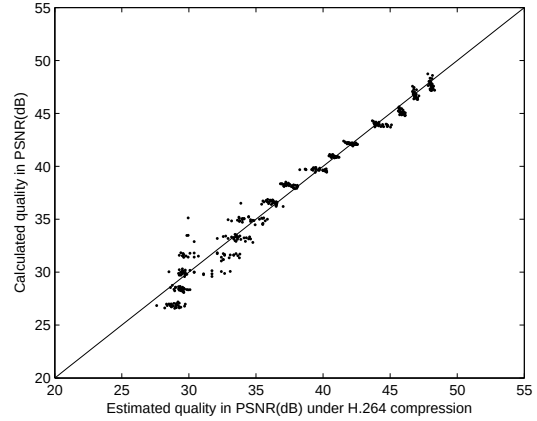
The accuracy evaluated using $\overline{MAE}$, Pearson correlation coefficient, $\overline{Corr_p}$ and $\overline{RMSE}$ for all the 512×512 videos shown in Fig. A-3 are listed in Table 7.1. In the table, hs' and hs respectively indicate the compression done with $Qp \in [1, 41]$ and $Qp \in [1, 50]$.

The proposed tree structure based scheme is further tested on the I-frames of the CIF videos and 720p videos to verify the effectiveness of the scheme and the definition:

the density of watermark embedding. To test the maximum accuracy in terms of PSNR that can be obtained, one “Ideal Mapping Curve” is generated for the CIF videos and one “Ideal Mapping Curve” is generated for the 720p videos using the 10 video frames shown in Fig. A-2. These 10 video frames are all 1920×1080 in size. So, to generate mapping curves for different resolution videos, downsampling and cropping may be needed.



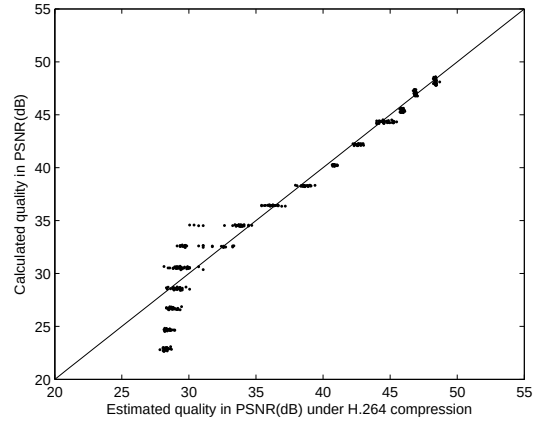
(a) Original video Kimono.



(b) Framewise quality evaluation on video Kimono.
 $\overline{MAE}|_{hs'} = 0.6525$ dB and $\overline{MAE}|_{hs} = 0.7788$ dB.



(c) Framewise quality evaluation on video Spin Calendar.



(d) Framewise quality evaluation on video Spin Calendar.
 $\overline{MAE}|_{hs'} = 0.6911$ dB and
 $\overline{MAE}|_{hs} = 1.3068$ dB.

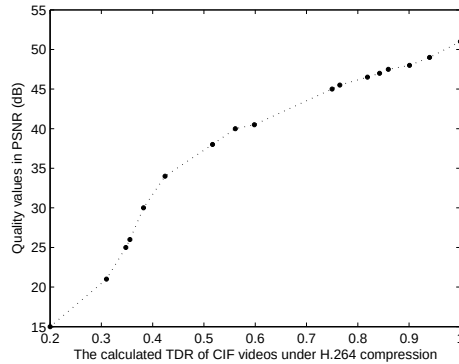
Figure 7.4: Illustration of the framewise quality evaluation on the I-frames of two sample 512×512 videos.

Table 7.1: Quality evaluation in terms of PSNR on I-frames of 512×512 videos

	MAE (dB)		$Corr_p$		$RMSE$	
	hs'	hs	hs'	hs	hs'	hs
Kimono	0.6525	0.7788	0.9868	0.9876	0.8961	1.0466
Crew	0.7567	0.9353	0.9881	0.9811	0.9393	1.2153
Cyclist	0.7575	0.8586	0.9871	0.9848	1.1185	1.1888
Park Scene	1.3145	1.5339	0.9826	0.9860	1.6152	1.8387
Tractor	0.7865	0.9783	0.9906	0.9825	1.1457	1.3721
Station	0.8667	0.8741	0.9901	0.9883	1.1572	1.1389
Spin Calendar	0.6911	1.3068	0.9906	0.9740	1.0064	1.8090
Blue sky	1.0778	1.6599	0.9774	0.9802	1.3400	2.0231
Mobcal	0.9136	1.4044	0.9918	0.9846	1.0636	1.8527
Riverbed	1.2453	1.4566	0.9885	0.9808	1.4275	1.7125
Shields	1.4591	2.0164	0.9929	0.9909	1.6363	2.4355
China Speed	1.7130	2.3594	0.9629	0.9706	1.9892	2.8950

7.3.2 Experimental results tested on I-frames of CIF videos

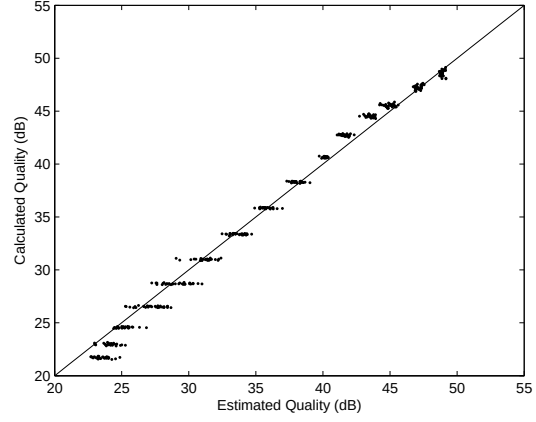
The “Ideal Mapping Curve” generated for quality evaluation under H.264 compression for the CIF videos is shown in Fig. 7.5. With the increasing of the compression strength, both the calculated TDR values and the PSNR values decrease monotonically.

**Figure 7.5:** The “Ideal Mapping Curve” generated to evaluate the quality of CIF videos under H.264 compression.

One sample CIF YUV video, Tempete, is shown in Fig. 7.6 (a). All its 30 frames



(a) Original video Tempete.



(b) Framewise quality evaluation on video Tempete.

 $\overline{MAE}|_{hs'} = 0.6286$ dB and $\overline{MAE}|_{hs} = 0.7389$ dB.

Figure 7.6: Illustration of the framewise quality evaluation on the I-frames of one sample CIF video.

Table 7.2: Quality evaluation in terms of PSNR on I-frames of CIF videos

	$\overline{MAE}$ (dB)		$\overline{Corr}_p$		$\overline{RMSE}$	
	hs'	hs	hs'	hs	hs'	hs
Big buck bunny	0.6145	0.6914	0.9962	0.9953	0.7471	0.8606
Bus	0.6590	0.8201	0.9950	0.9958	0.8176	1.0449
Foreman	0.7536	1.5628	0.9834	0.9687	1.0869	2.0223
Silent	0.8810	1.2535	0.9962	0.9953	1.0571	1.6045
Tempete	0.6286	0.7389	0.9949	0.9963	0.7856	0.9380
Waterfall	1.1450	1.0999	0.9935	0.9913	1.3176	1.2634
Big buck bunny 2	1.0470	1.1723	0.9876	0.9872	1.5683	1.6848
Riverbed	1.0353	0.9869	0.9889	0.9932	1.1259	1.0732
Station	0.7540	0.9949	0.9901	0.9890	0.9645	1.2755
Sunflower	0.6274	0.6269	0.9960	0.9959	0.8768	0.8544
Kimono	0.6077	0.6365	0.9938	0.9950	0.7714	0.8033
Pedestrian area	0.6646	0.8614	0.9947	0.9949	0.8745	1.1522

are intra-coded. The framewise quality evaluation is shown in Fig. 7.6 (b). The average MAE for the 30 frames quality evaluation is 0.6286 dB with quantization parameter $Qp \in [1, 41]$ and the average MAE is 0.7389 dB with $Qp \in [1, 50]$.

The quality evaluation is also conducted on the rest of the CIF videos shown in Fig. A-4. The quality evaluation accuracy assessed using the average MAE, average Pearson correlation coefficient and the average RMSE is listed in Table 7.2. hs' and hs respectively indicates the compression done with the quantization parameters $Qp \in [1, 41]$ and $Qp \in [1, 50]$.

7.3.3 Experimental results tested on I-frames of 720p videos

The “Ideal Mapping Curve” generated for the quality evaluation under H. 264 compression on the I-frames of 720p videos is shown in Fig. 7.7. This curve is also obtained

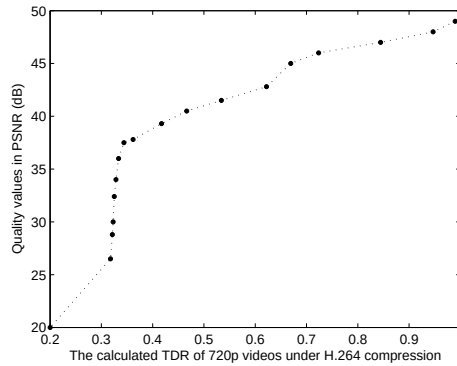


Figure 7.7: The “Ideal Mapping Curve” generated to evaluate the quality of 720p or 640p videos under H.264 compression.

by testing the 10 video frames in Fig. A-2 under H.264 compression with quantization parameters $Qp = [1, 10, 15 : 3 : 50]$. The “Ideal Mapping Curve” shows the monotonic relationship between the calculated TDR and the quality values.

One sample 720p video, Crew, is shown in Fig. 7.8 (a). All the 30 frames of video

Crew are intra-coded under H.264 with quantization parameters $Qp = [1, 10, 15 : 3 : 50]$. The tested framewise quality evaluation results are shown in Fig. 7.8 (b). The average MAE for the results obtained with $Qp \in [1, 41]$ is 0.6953 dB and the average MAE for the results tested with $Qp \in [1, 50]$ is 0.8643 dB.

The quality evaluation is also conducted on the rest of the 720p videos shown in Fig. A-5. The accuracy of the quality evaluation is assessed using the average MAE, average Pearson correlation and average RMSE and is listed in Table 7.3.

The hs' and hs respectively indicates the compression done with quantization parameters $Qp \in [1, 41]$ and $Qp \in [1, 50]$.

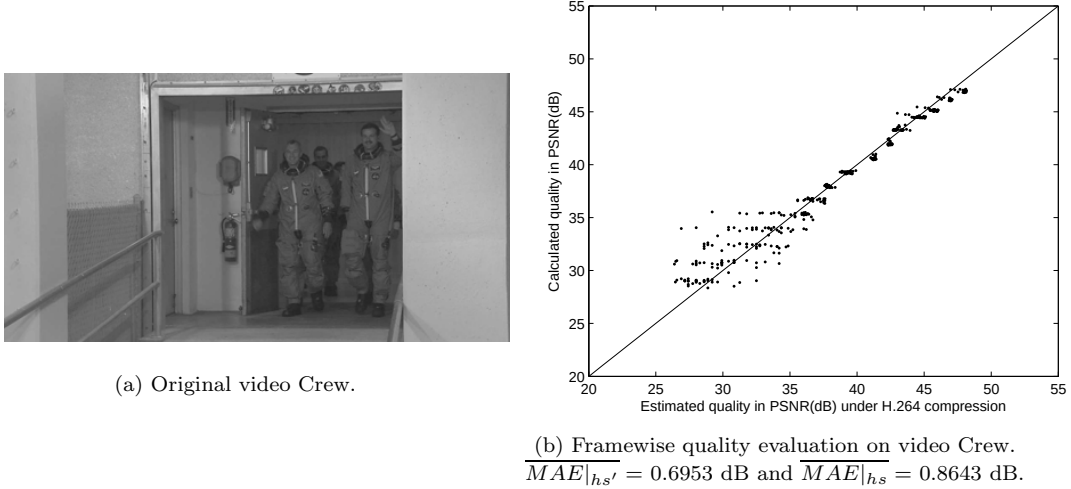


Figure 7.8: Illustration of the framewise quality evaluation on the I-frames of one sample 720p video.

All the results listed in Table 7.1, Table 7.2 and Table 7.3 confirm that the proposed scheme works effectively for the quality evaluation on I-frames of different resolution videos. Experiments are extended to test the effects of quality evaluation on both I- and P-frames of CIF and 720p videos. For this extended experiment, each of the test videos is coded into IPPPI $\dots$ sequences. To avoid involving too many “Ideal Mapping

Curves”, the mapping curves shown in Fig. 7.5 and Fig. 7.7 are respectively used for the quality evaluation of the I- and P-frames for CIF and 720p videos.

Table 7.3: Quality evaluation in terms of PSNR on I-frames of 720p and 640p videos

	MAE (dB)		$Corr_p$		$RMSE$	
	hs'	hs	hs'	hs	hs'	hs
Jets	0.6264	0.8600	0.9860	0.9838	0.8036	1.1997
Crew	0.6953	0.8643	0.9790	0.9817	1.0554	1.2766
Blue sky	1.1461	1.5770	0.9602	0.9633	1.3851	2.0736
Tractor	1.0190	1.3314	0.9620	0.9779	1.5595	1.8617
Slide editing	0.8191	0.9462	0.9826	0.9841	1.0831	1.2870
Kimono	0.9917	1.1389	0.9765	0.9696	1.3620	1.5224
Pedestrian area	0.6899	0.9280	0.9843	0.9740	0.9436	1.3282
Cyclists	0.7639	0.9156	0.9764	0.9794	1.0674	1.2659
Park scene	0.8522	1.1110	0.9727	0.9827	1.2595	1.5431

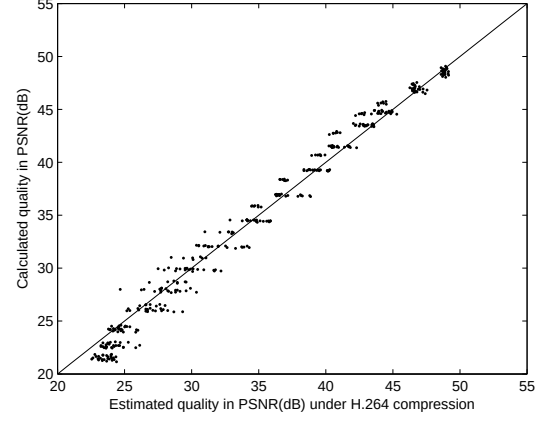
7.3.4 Experimental results tested on I- and P-frames of CIF videos

With the “Ideal Mapping Curve” shown in Fig. 7.5, the framewise quality evaluation results tested on the IPPPI $\dots$ frames of the CIF video Tempete are illustrated in Fig. 7.9(b). The calculated average MAE is associated with the results. Comparing Fig. 7.9(b) to Fig. 7.6 (b), we can see that, although we use the “Ideal Mapping Curve” generated on I-frames to evaluate the quality of P-frames, the quality points only become slightly divergent and the average MAE of Fig. 7.9(b) has a slight drop comparing to that of Fig. 7.6 (b).

The quality evaluation are also tested on the IPPPI $\dots$ frames of the other CIF videos and the assessed accuracy is listed in Table 7.4. These results confirm that both the I-frames and the P-frames can use the same “Ideal Mapping Curve” for quality



(a) Original video Tempete.

(b) Framewise quality evaluation on video Tempete.
 $\overline{MAE}|_{hs'} = 0.8363$ dB and $MAE|_{hs} = 0.9385$ dB.**Figure 7.9:** Illustration of the framewise quality evaluation on the I- and P-frames of one sample CIF video.**Table 7.4:** Quality evaluation in terms of PSNR on I- and P-frames of CIF videos

	MAE (dB)		$Corr_p$		$RMSE$	
	hs'	hs	hs'	hs	hs'	hs
Big buck bunny	0.6301	0.7256	0.9964	0.9960	0.7598	0.8592
Bus	0.7895	0.9550	0.9950	0.9920	1.0141	1.2203
Foreman	1.0562	1.8579	0.9655	0.9476	1.4640	2.6591
Silent	0.9597	1.3367	0.9938	0.9939	1.1337	1.6867
Tempete	0.8363	0.9385	0.9898	0.9924	1.0634	1.1930
Waterfall	0.8156	0.8581	0.9946	0.9945	0.9542	0.9799
Big buck bunny 2	1.2062	1.0607	0.9856	0.9903	1.4633	1.3380
Riverbed	1.0630	0.8666	0.9879	0.9919	1.2182	1.0829
Station	0.7173	0.7515	0.9946	0.9925	0.8906	0.9536
Sunflower	1.0230	0.9578	0.9960	0.9923	1.4167	1.3269
Kimono	0.8763	1.0445	0.9968	0.9979	1.1107	1.2750
Pedestrian area	0.8236	0.8951	0.9909	0.9936	1.0834	1.1477

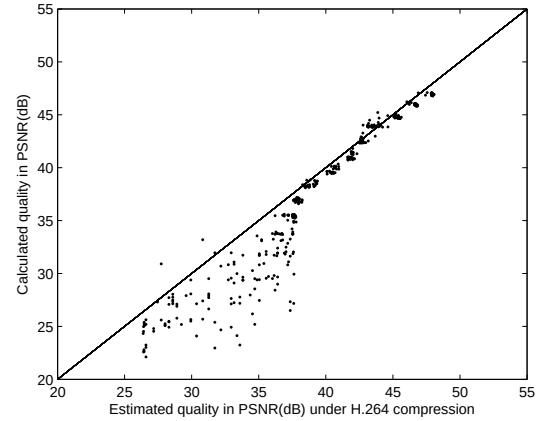
evaluation and the quality evaluation accuracy in terms of PSNR is still good for CIF videos.

7.3.5 Experimental results tested on I- and P-frames of 720p videos

With the “Ideal Mapping Curve” shown in Fig. 7.7, the quality evaluation is conducted on the IPPPI $\dots$ frames of the 720p video, Crew, and the framewise quality evaluation results are illustrated in Fig. 7.10 (b). Comparing to Fig. 7.8 (b), the quality points in Fig. 7.10 (b) become divergent after the calculated PSNR < 37 dB and the average MAE has a 0.5 dB drop when $Qp \in [1, 41]$ and 1 dB drop when $Qp \in [1, 51]$.



(a) Original video Crew.



(b) Framewise quality evaluation on video Crew.
 $\overline{MAE}|_{h.s'} = 1.2587$ dB and $MAE|_{h.s} = 1.8642$ dB.

Figure 7.10: Illustration of the framewise quality evaluation on the I- and P-frames of one sample 720p video.

The experiment is also conducted on other 720p videos. The assessed quality evaluation accuracy is listed in Table 7.5, which indicate that the proposed tree structure based scheme works with acceptable accuracy on the quality evaluation done on the

IPPP I . . . frames of 720p videos using the “Ideal Mapping Curves” generated on I-frames.

Table 7.5: Quality evaluation in terms of PSNR on I- and P-frames of 720p and 640p videos

	$\overline{MAE}$ (dB)		$\overline{Corr}_p$		$\overline{RMSE}$	
	hs'	hs	hs'	hs	hs'	hs
Jets	0.9799	1.6946	0.9747	0.9738	1.4333	2.4923
Crew	1.2587	1.8642	0.9670	0.9633	1.7848	2.7606
Blue sky	1.4684	1.9947	0.9505	0.9557	1.7899	2.6476
Tractor	1.6650	1.8957	0.9692	0.9720	2.7523	2.8474
Slide editing	1.2846	1.4956	0.9776	0.9685	1.3481	2.4773
Kimono	0.9068	1.9521	0.9662	0.9582	1.2791	2.8197
Pedestrian area	1.0903	1.7441	0.9748	0.9745	1.6852	2.6141
Cyclists	0.7081	1.3786	0.9830	0.9821	0.9991	2.0476
Park scene	0.9823	1.1093	0.9895	0.9887	1.1891	1.3176

7.4 Quality evaluation in terms of PSNR under packet loss

Video compression is widely used in many applications and the encoded bit stream is transmitted to receiver side through some transmission channel. During the transmission, the packet loss could happen on the transportation layer which causes the received video bit stream syntax error [97]. To mitigate the quality loss, error concealment techniques are introduced to minimize the video quality degradation caused by the packet loss. Most recent video coding standards, like H.264 or MPEG 4-AVC, have a set of coding tools to minimize the packet loss impact on the video quality [98][99], including data partitioning, redundant slices, etc. Data partitioning can rearrange data partition for different kinds of video data, such as macro-block type, motion vector

data or residual data, so that more important information can be better protected from possible corruptions [100][101][102]. Also, error control coding is used to provide error correction mechanism to a certain degree. All these tools can facilitate the decoder to detect and locate the syntax error, then to perform the error concealment better. Once the corrupted macro-blocks are located, error concealment techniques are used to suppress the quality loss. Generally, there are two kinds of error concealment techniques, the spatial concealment [103] and the temporal concealment [104][105] because of the temporal and spatial correlation in video contents.

When the motion information for a macro-block is lost, either a zero motion vector (temporal replacement) is applied or the motion vector is interpolated or extrapolated based on the neighboring macro-blocks and the co-located macro-block from the previous frame [106]. Then the deducted motion vector will be used for the current macro-block motion compensation calculation and reconstruction. For example, the boundary matching technique can be used to find the best matching macro-block in the reference frame using the neighboring macro-block motion vectors to locate the candidate macro-block for matching calculation together with the boundary pixels of spatial neighboring blocks [107][108][109]. Regarding the corrupted intra macro-block, the spatial concealment is used in which pixels from neighboring blocks are used to interpolate the current blocks.

Video error concealment technique has a direct impact on the objective/subjective quality of the decoded and error concealed video. Some error concealment techniques may focus on minimizing subjective quality loss regarding to MOS, while others may be more focusing on minimizing the objective quality loss with respect to PSNR, SSIM, etc. It can be seen that a lot of techniques are involved in the video error concealment area.

To evaluate the impact of the packet loss on the video quality, a complete end to end simulation system is desirable which includes video encoder, transmission channel with noise simulation, video decoder and error concealment module. The packet loss detection can be done on the transportation layer and the result can be sent to the decoder for syntax error detection. Then the error concealment module can be used to minimize the quality loss of the decoded video. However, building such a system would be too time consuming and there are too many variables in this system which directly affect the final video quality. Also, the video syntax error detection and the error concealment are both broad research areas which do not have standardized techniques and benchmark. It is necessary to design a practical and simplified system to emulate the video quality degradation caused by the packet loss.

In our experiments, the quality loss caused by the packet loss is restricted to the macro-block level. The frame dropping caused by packet loss would not be taken into consideration. Two most common types of quality loss caused by packet loss are simulated: the corrupted intra macro-blocks and the corrupted inter macro-blocks. In I-frames, a certain number of macro-blocks in random positions are discarded, then those macro-blocks are recovered using the weighted averaging method [110]. In P-frames, a certain number of macro-blocks are discarded. In these macro-block, a ratio of intra macro-blocks and inter macro-blocks are predefined. For the intra macro-blocks, the weighted averaging method is used to recover the discarded intra macro-blocks. For the discarded inter macro-block, its content will be directly copied from macro-block in the corresponding position of the reference frame [111]. The number of the discarded macro-blocks is the factor directly controlling how much the video quality would be affected by the pack loss. In our experiment, we only model the isolated packet loss pattern.

The weighted averaging method is illustrated in Fig. 7.4. In Fig. 7.4, the dark grids stand for the discarded macro-block. The gray blocks are the neighboring pixels that will be used to recover the discarded macro-block. The gray block P is the current pixel in the discarded macro-block to be recovered. The pixels P_T , P_B , P_L and P_R with the corresponding distances, d_T , d_B , d_L and d_R , from pixel P are used to bi-linearly interpolate the pixel P .

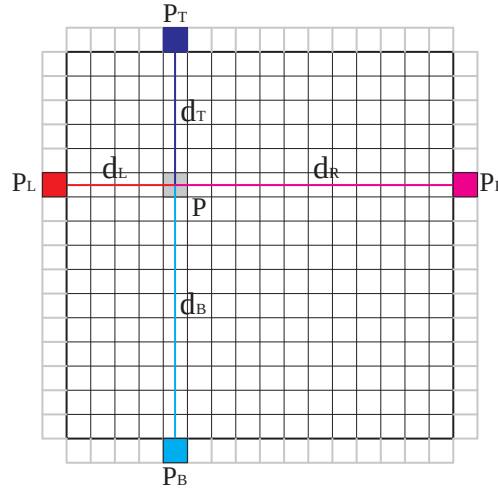


Figure 7.11: The weighted averaging error concealment on the intra-blocks corrupted by packet loss.

Equ. (7.5) mathematically depicts the weighted averaging method.

$$P = \frac{d_T P_T + d_B P_B + d_L P_L + d_R P_R}{d_T + d_B + d_L + d_R} \quad (7.5)$$

Although the error concealment in our test system is not optimal in terms of the state-of-art error concealment techniques, it is closely correlated with how a real error concealment system could behave under packet loss. With this simplified system, we test the feasibility of our proposed scheme in evaluating the video quality degradation under packet loss.

With the framework we proposed in Chapter 3, we made the assumption that the distortion first affects the higher frequency components and the insignificant bits of the cover image or video signals. Then, with the increasing of the distortion strengths, the distortion affects the lower frequency components and the more significant bits of the cover signals. With this assumption, we estimate the watermark embedding strengths for different textured image or video by analyzing their content complexity. The packet loss and error concealment techniques affects the signal quality in a totally different way from our assumption. In this case, it is difficult to properly estimate the watermark embedding strength by only estimate the signal content complexity. Thus, in the implementation under packet loss, we use the same watermark embedding strengths as those used under the H.264 compression for the test videos to observe the effectiveness of the proposed scheme.

7.4.1 Experimental results tested under packet loss

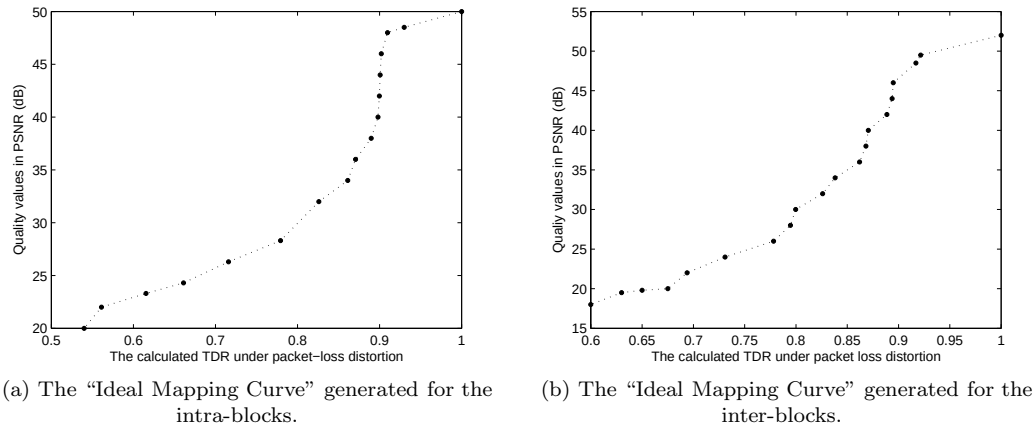


Figure 7.12: The “Ideal Mapping Curve” under packet loss.

With our simplified packet loss and error concealment simulation system, the distortion strength is controlled by the percentage of corrupted intra or inter macro-blocks

over the total number of macro-blocks in the current frame. The proposed scheme is tested on the CIF videos shown in Fig. A-4 when the percentage of corrupted macro-blocks respectively equals to $[0.1\%, 1\%, 2\%, 3\%, 4\% : 5\% : 50\%]$. The “Ideal Mapping Curve” generated for the error concealment on the intra-blocks and inter-blocks are shown in Fig. 7.12(a) and (b). The experimental results tested on the I-frames and P-frames are shown in Fig. 7.13 (a) and (b). The results in Fig. 7.13 (a) are obtained

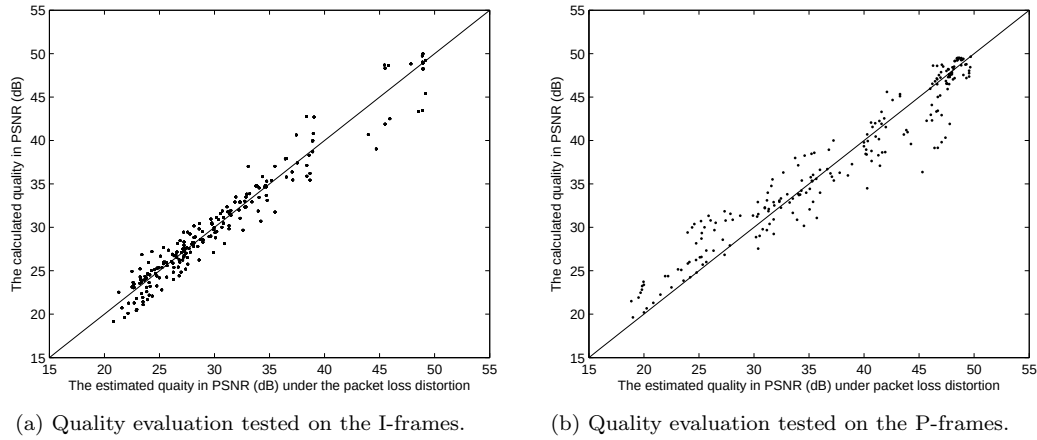


Figure 7.13: Experimental results tested on CIF videos under packet loss.

by testing the first I-frame of all the CIF videos in Fig. A-4. The results in Fig. 7.13 (b) are achieved by testing a randomly selected P-frame from each of the CIF videos in Fig. A-4.

Table 7.6: Quality evaluation in terms of PSNR under packet loss

	$MAE(dB)$	Pearson correlation	$RMSE$
Results tested on I-frames	1.2011	0.9728	1.6200
Results tested on P-frames	1.7766	0.9678	2.4258

The assessed accuracy for the video quality evaluation under packet loss listed in Table 7.6 indicates that the proposed scheme is a feasible approach for the quality

evaluation under packet loss.

7.5 Summary

In this chapter, the proposed tree structure based scheme is further implemented for video quality evaluation. The adaptive watermark embedding strength is assigned to a video sequence by estimating the quality degradation characteristics of the first video frame. The watermark is embedded in the luminance component of the video signals. The concept of the density of watermark embedding is introduced to make the tree structure based scheme effective for quality evaluation of videos with different resolutions, such as 352×288 (CIF), 512×512 , 1280×640 (640p) and 1280×720 (720p). The experimental results show that the tree structure based scheme can evaluate video quality under H.264 and packet loss related distortion.

Chapter 8

Conclusions and future work

In this thesis, a digital watermarking based quality evaluation framework is proposed to conduct quality evaluation in terms of the existing objective quality metrics, such as PSNR, wPSNR, Watson JND, SSIM and VIF, under JPEG compression, JPEG2000 compression, Gaussian low-pass filtering, Gaussian noise distortion, H.264 compression and the packet loss distortion. With the proposed framework, the image or video quality can be evaluated in terms of the user-preferred objective quality metric without the reference image or video signal. The “Ideal Mapping Curve” is introduced as a model which is experimentally generated to quantify the relationship between the quality change (with respect to certain type of quality metric such as PSNR and JND) of the host signal and the watermark signal under certain distortion. The SPIHT tree structure and HVS masking are used to adaptively adjust the watermark embedding by utilizing the host signal characteristics. Based on the proposed framework, three schemes are implemented for image quality evaluation using quantization based watermarking, image quality evaluation using SPIHT tree structure and HVS based watermarking and video quality evaluation using SPIHT tree structure and HVS based

watermarking. The three schemes are designed to assess image and video quality in terms of any Full-Reference quality metrics, such as PSNR, wPSNR, JND and SSIM. All these schemes are implemented in the DWT domain and work with high accuracy.

The quantization based image quality evaluation scheme has relatively low computational efficiency. The lack of consideration of the human perception characteristics introduces relatively more significant quality loss caused by the watermark embedding process to the cover images. The quantization based scheme works effectively under JPEG compression. The tree structure based image quality estimation scheme utilizes the SPIHT structure and HVS mapping to guide the watermark embedding process. The tree structure based scheme has several performance improvements over the quantization based scheme. First, due to the adaptive watermark embedding process, the quality of the watermarked images is increased by 7 dB in PSNR on average over that from the quantization based scheme. The tree structure based image quality evaluation scheme is tested effective under JPEG compression, JPEG2000 compression, Gaussian low-pass filtering and Gaussian noise distortion. In the meanwhile, the computational efficiency of the tree structure based scheme is improved greatly over the quantization based scheme. With the good computational efficiency, the tree structure based quality evaluation scheme is further tested on I-frames of the video sequences in terms of PSNR under H.264 compression. The experimental results show the effectiveness.

In the tree structure based image quality estimation scheme, the empirical grouping thresholds are set up by observing the complexity values categorized in the 6 groups as presented in Section 6.2.2. In the experiments, we found out that the complexity values of 2 neighboring groups have an overlap as shown in Fig. 8.1. The empirical thresholds are chosen as the median value of the overlap. In the figure, t_1 and t_2 are two thresholds. As we mentioned previously, the empirical grouping affects the accuracy of the proposed

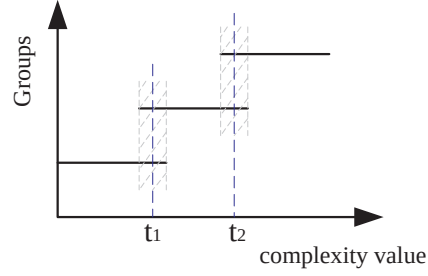


Figure 8.1: Strategy of choosing empirical grouping thresholds for the tree structure based scheme.

scheme. Thus, the smaller the overlap, the more accurate the empirical grouping and the more accurate the quality estimation. The optimization of the grouping is part of our future work. It is desirable to minimize the overlaps of any 2 neighboring groups and maximize the accuracy of the quality estimation.

As the future work, the proposed scheme can be further tested with the quality evaluation in terms of MOS. Moreover, the quality of image or video signals affected by multiple distortions can also be tested using the proposed scheme.

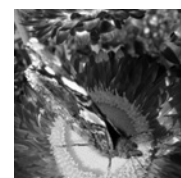
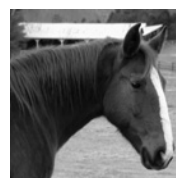
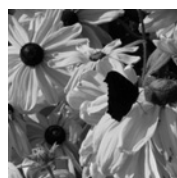
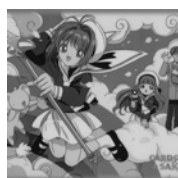
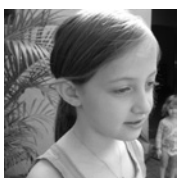
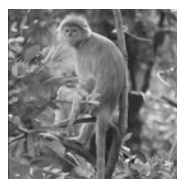
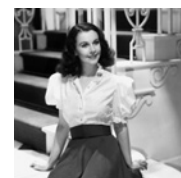
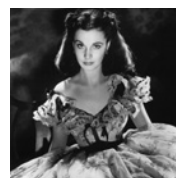
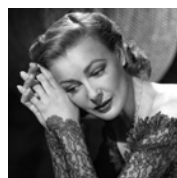
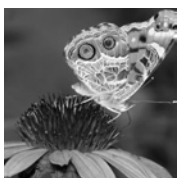
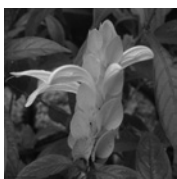
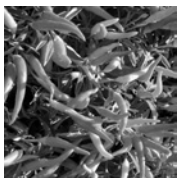
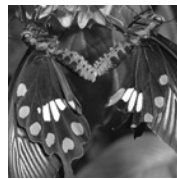
Appendix A

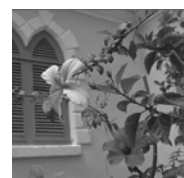
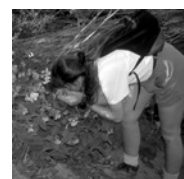
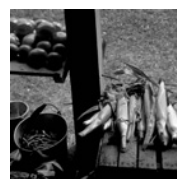
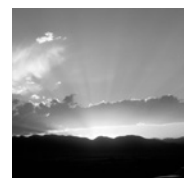
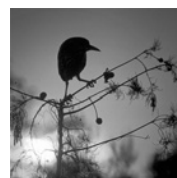
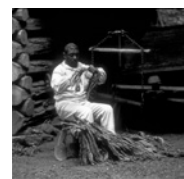
There are 150 gray images in our image library. All of these images are 512×512 in size. The images include different types of textures, such as portraits, plants, animals, animations, sceneries, buildings and crowd.

The first 20 images are used to generate the “Ideal Mapping Curves” for the quantization based image quality estimation scheme. The first 50 images are used to generate the “Ideal Mapping Curves” for the tree structure based quality estimation schemes. The 51st to the 150th images are used to evaluate the accuracy of the tree structure based scheme. The 12 video sequences shown in Chapter 7 are used to evaluate the accuracy of the proposed video quality estimation scheme.









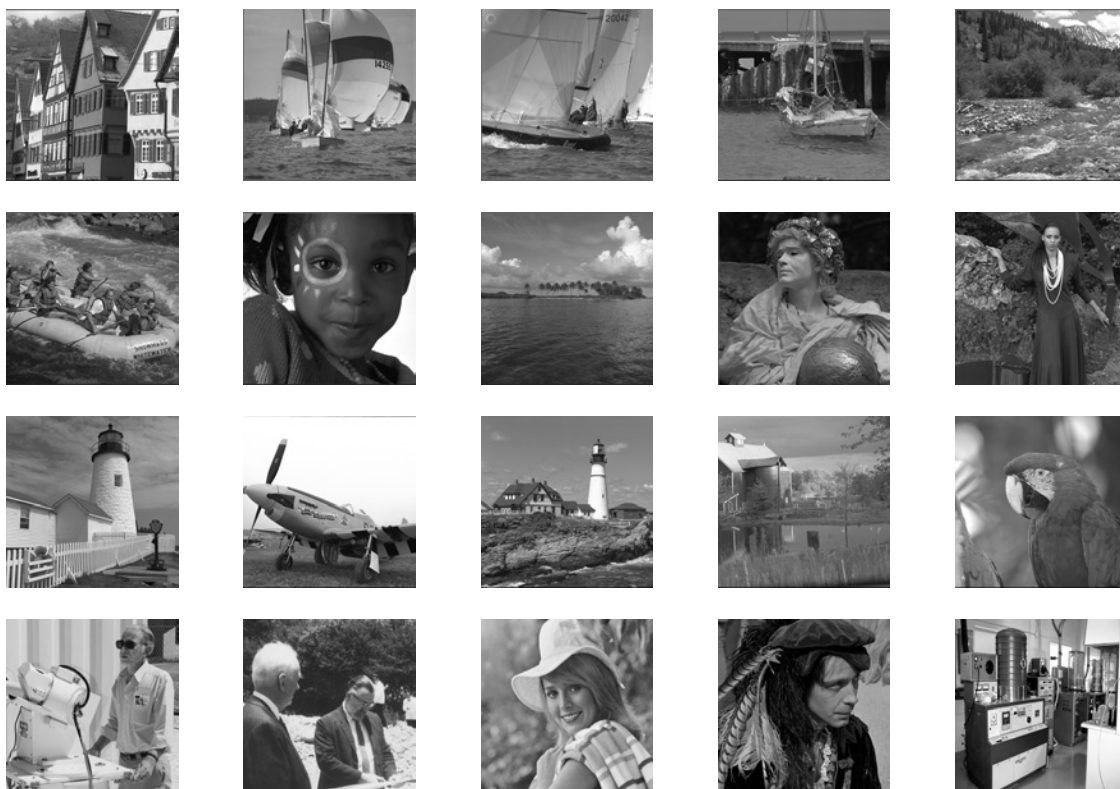


Figure A-1: The 150 original images used in the experiments.



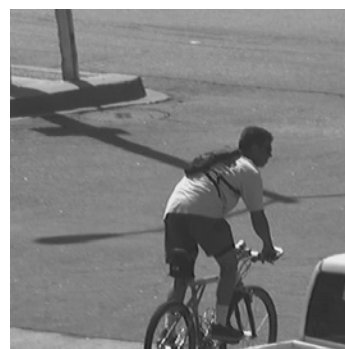
Figure A-2: The 10 video frames used to generate “Ideal Mapping Curves”.



Kimono.



Crew.



Cyclist.



Park Scene.



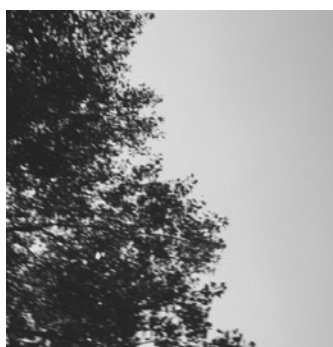
Tractor.



Station.



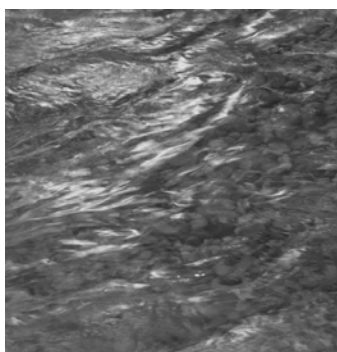
Spin calendar.



Blue sky.



Calendar.



Riverbed.

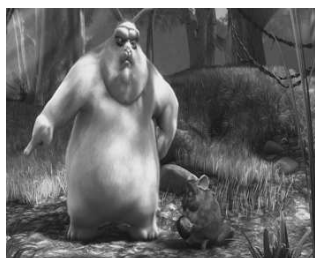


Shields.



China speed.

Figure A-3: The 12 original 512×512 video sequences.



Big buck bunny.



Bus.



Foreman.



Silent.



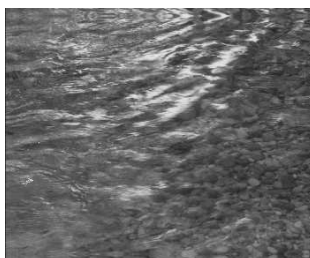
Tempete.



Waterfall.



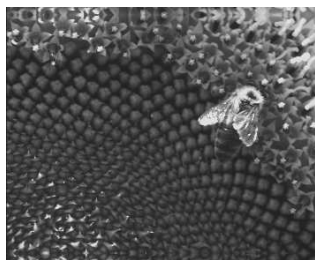
Big buck bunny 2.



Riverbed.



Station.



Sunflower.



Kimono.



Pedestrian area.

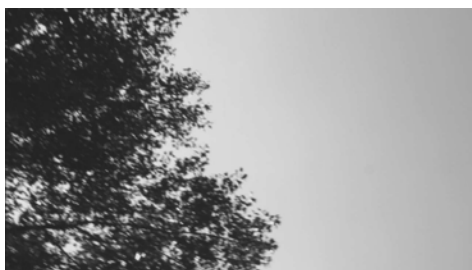
Figure A-4: The 12 original CIF video sequences.



Jets.



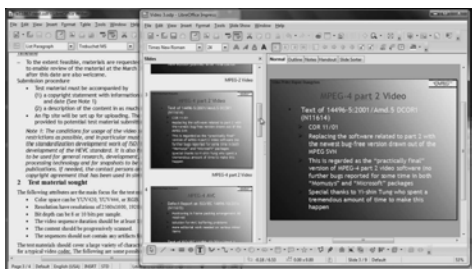
(a) Crew.



Blue sky.



Tractor2.



Slide editing.



Kimono.



Pedestrian area.



Cyclists.



Park scene.

Figure A-5: The 9 original 720p or 640p video sequences.

Bibliography

- [1] M. Miyahara, K. Kotani, and V. Algazi. Objective picture quality scale (PQS) for image coding. IEEE Transactions on Communications, 46(9):1215–1226, September 1998.
- [2] I. Cox, M. Miller, J. Bloom, J. Fridrich, and T. Kalker. Digital watermarking and steganography. ISBN: 978-0-12-372585-1. Morgan Kaufmann Publishers, 2008.
- [3] A. Ahumada and H. Peterson. Luminance-model-based DCT quantization for color image compression. In Processings of SPIE, Human Vision, Visual Processing, and Digital Display III, volume 1666, pages 365–374, 10-14 February 1992.
- [4] Z. Wang, A. Bovik, H. Sheikh, and E. Simoncelli. Image quality assessment: From error visibility to structural similarity. IEEE Transactions on Image Processing, 13(4):600–612, April 2004.
- [5] H. Sheikh and A. Bovik. Image information and visual quality. IEEE Transactions on Image Processing, 15(2):430–444, February 2006.
- [6] H. Sheikh and A. Bovik. A visual information fidelity approach to video quality

- assessment. In International Workshop on Video Processing and Quality Metrics for Consumer Electronics, pages 23–26, 23-25 January 2005.
- [7] P. Lecallet, C. Viard-Gaudin, S. Pechard, and E. Caillault. No reference and reduced reference video quality metrics for end to end QoS monitoring. IEICE Transactions on Communications, E85-B(2):289–296, February 2006.
- [8] M. Farias, M. Carli, and S. Mitra. Objective video quality metric based on data hiding. IEEE Transactions on Consumer Electronics, 51(3):983–992, August 2005.
- [9] A. Mishra, A. Jain, M. Narwaria, and C. Agarwal. An experimental study into objective quality assessment of watermarked images. International Journal of Image Processing, 5(2):199–219, May 2011.
- [10] O. Ekici, B. Coskun, U. Naci, and B. Sankur. Comparative assessment of semi-fragile watermarking techniques. In Proceedings of SPIE, Multimedia Systems and Applications IV, volume 4518, pages 177–188, 2024 August 2001.
- [11] Q. Cheng and T. Huang. Robust optimum detection of transform domain multiplicative watermarks. IEEE Transactions on Signal Processing, 51(4):906–924, April 2004.
- [12] M. Hernandez, M. Miyatake, and H. Meana. Analysis of a DFT-based watermarking algorithm. In International Conference on Electrical and Electronics Engineering, pages 44–47, 7-9 September 2005.
- [13] F. Deguillaume, G. Csurka, J. O’Ruanaidh, and T. Pun. Robust 3D DFT video watermarking. In Proceedings of SPIE on Security and Watermarking of Multimedia Contents, volume 3657, pages 1–12, 23-29 January 1999.

- [14] D. Zheng, J. Zhao, W. Tam, and F. Speranza. Image quality measurement by using digital watermarking. In IEEE International Workshop on Haptic, Audio and Visual Environments and Their Applications, volume 1, pages 65–70, 20–21 September 2003.
- [15] S. Lin and C. Chen. A robust DCT-based watermarking for copyright protection. IEEE Transactions on Consumer Electronics, 46(3):415–421, August 2000.
- [16] D. Kundur and D. Hatzinakos. Digital watermarking for telltale tamper proofing and authentication. Proceedings of the IEEE, 87(7):1167–1179, July 1999.
- [17] D. Kundur and D. Hatzinakos. A robust digital image watermarking method using wavelet based fusion. In IEEE International Conference on Image Processing, volume 1, pages 544–547, 26–29 October 1997.
- [18] L. Cai and J. Zhao. Evaluation of speech quality using digital watermarking. IEICE Electronics Express, 1(13):380–385, October 2004.
- [19] L. Cai and J. Zhao. Speech quality evaluation: A new application of digital watermarking. In IEEE Instrumentation and Measurement Technology Conference, pages 726–731, 17–19 May 2005.
- [20] S. Wang, D. Zheng, J. Zhao, W. Tam, and F. Speranza. An image quality evaluation method based on digital watermarking. IEEE Transactions on Circuits and Systems for Video Technology, 17(1):98–105, January 2007.
- [21] H. Qi, D. Zheng, and J. Zhao. Human visual system based adaptive digital image watermarking. Elsevier Journal: Signal Processing, 88(1):174–188, January 2008.

- [22] L. Cai, R. Tu, J. Zhao, and Y. Mao. Speech quality evaluation: a new application of digital watermarking. IEEE Transactions on Instrumentation and Measurement, 56(1):45–55, February 2007.
- [23] G. Doerr and J. Dugelay. A guide tour of video watermarking. Signal Processing: Image Communication, 18(4):263–282, April 2003.
- [24] S. Bhattacharya, T. Chattopadhyay, and P. Arpan. A survey on different video watermarking techniques and comparative analysis with reference to H.264/AVC. In IEEE International Symposium on Consumer Electronics, pages 1–6, 28 June - 1 July 2006.
- [25] J. Lu, Z. Yang, F. Yang, and L. Li. A MPEG2 video watermarking algorithm based on DCT domain. In IEEE Digital Media and Digital Content Management, pages 194–197, 15-16 May 2011.
- [26] N. Sakr, N. Georganas, J. Zhao, and E. Petriu. Multimodal vision-haptic perception of digital watermarks embedded in 3D meshes. IEEE Transactions on Instrumentation and Measurement, 59(5):1047–1055, May 2010.
- [27] N. Sakr, N. Georganas, J. Zhao, and E. Petriu. Perceptibility of digital watermarks embedded in 3D meshes using unimodal and bimodal haptic-visual interaction. In IEEE International Instrumentation and Measurement Technology Conference, pages 987–991, 5-7 May 2009.
- [28] N. Sakr, N. Georganas, and J. Zhao. Perceptibility of watermarked 3D meshes using a multimodal visuo-haptic interface. In IEEE International Conference on Virtual Environments, Human-Computer Interfaces, and Measurement Systems Istanbul, pages 200–204, 14-16 July 2008.

- [29] Q. Li and Z. Dong. Novel text watermarking algorithm based on Chinese characters. In International Symposium on Computer Science and Computational Technology, volume 2, pages 348–351, 20-22 December 2008.
- [30] C. Thomoborson, J. Nagra, R. Somaraju, and C. He. Tamper-proofing software watermarks. In Proceedings of the Second Workshop on Australasian Information Security, Data Mining and Web Intelligence, and Software Internationalisation, volume 32, pages 27–36, 18-20 March 2004.
- [31] R. Agrawal and J. Kiernan. Watermarking relational databases. In International Conference on Very Large Databases, pages 155–166, 20-23 August 2002.
- [32] Y. Chen and L. Chang. A secure and robust digital watermarking technique by the block cipher RC6 and secure hash algorithm. In IEEE International Conference on Image Processing, volume 2, pages 518–521, 7-10 October 2001.
- [33] I. Cox, J. Kilian, F. Leighton, and T. Shamoon. Secure spread spectrum watermarking for multimedia. IEEE Transactions on Image Processing, 6(12):1673–1687, December 1997.
- [34] Y. Wang, J. Doherty, and R. Van-Dyck. A wavelet-based watermarking algorithm for ownership verification of digital images. IEEE Transactions on Image Processing, 11(2):77–88, February 2002.
- [35] A. Adelsbach and A. Sadeghi. Zero-knowledge watermark detection and proof of ownership. In International Workshop on Information Hiding, pages 273–288, 25-27 April 2001.
- [36] L. Qiao and K. Nahrstedt. Watermarking methods for MPEG encoded video: to-

- wards resolving rightful ownership. In IEEE Multimedia Computing and Systems, pages 276–285, 28 June - 1 July 1998.
- [37] M. Wu and B. Liu. Watermarking for image authentication. In IEEE International Conference on Image Processing, volume 2, pages 437–441, 4-7 October 1998.
- [38] B. Mobasseri, M. Sieffert, and R. Simard. Content authentication and tamper detection in digital video. In Proceedings of IEEE Conference on Image Processing, volume 1, pages 458–461, 10-13 September 2000.
- [39] L. Liu, X. Li, and P. Li. Diversity technique based secure video watermarking for broadcasting monitoring. In World Congress on Computer Science and Information Engineering, volume 7, pages 285–289, 1-2 April 2009.
- [40] R. Petrovic, B. Tehranchi, and J. Winograd. Security of copy-control watermarks. In International Conference on Telecommunications in Modern Satellite, Cable and Broadcasting Services, pages 117–126, 26-28 September 2007.
- [41] D. Zheng, S. Wang, and J. Zhao. An RST invariant image watermarking algorithm with mathematical modeling and analysis of the watermarking processes. IEEE Transactions on Image Processing, 18(5):1055–1068, May 2009.
- [42] D. Zheng, J. Zhao, and A. Saddik. RST invariant digital image watermarking based on log-polar mapping and phase correlation. IEEE Transactions on Circuits and Systems for Video Technology: Special Issue on Authentication, Copyright Protection and Information Hiding, 13(8):753–765, August 2003.
- [43] D. Zheng, Y. Liu, J. Zhao, and A. Saddik. A survey of RST invariant image watermarking algorithms. ACM Computing Surveys, 39(2):1–91, June 2007.

- [44] D. Zheng, Y. Liu, and J. Zhao. RST invariant digital image watermarking based on a new phase-only filtering method. Elsevier Journal: Signal Processing, 85(12):2354–2370, December 2005.
- [45] Y. Liu, D. Zheng, and J. Zhao. A rectification scheme for RST invariant image watermarking. IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences, Special Section on Cryptography and Information Security, E88-A(1):314–318, January 2005.
- [46] C. Cruz, R. Reyes, M. Nakano, and H. Perez. Image content authentication system based on semi-fragile watermarking. In Midwest Symposium on Circuits and Systems, pages 303–309, 10-13 August 2008.
- [47] C. Hernandez and C. Huitzi. A fragile watermarking scheme for image authentication in mobile devices. In International Conference on Electrical Engineering Computing Science and Automatic Control, pages 1–6, 26-28 October 2011.
- [48] D. Zhang, Z. Pan, and H. Li. A contour-based semi-fragile image watermarking algorithm in DWT domain. In International Workshop on Education Technology and Computer Science, volume 3, pages 228–231, 6-7 March 2010.
- [49] ITU-T. Objective picture quality measurement method by use of in-service test signals. In ITU-T J.147, 2002.
- [50] S. Wang, D. Zheng, J. Zhao, W. Tam, and F. Speranza. A digital watermarking and perceptual model based video quality measurement. In IEEE Instrumentation and Measurement Technology Conference, pages 1729–1734, 17-19 May 2005.
- [51] S. Wang, J. Zhao, W. Tam, and F. Speranza. Image quality measurement by

- using watermarking based on discrete wavelet transform. In The 32nd Biennial Symposium on Communications, pages 210–212, 1-3 June 2004.
- [52] S. Wang, D. Zheng, J. Zhao, W. Tam, and F. Speranza. An accurate method for image quality evaluation using digital watermarking. IEICE Electronics Express (ELEX), 2(20):523–529, October 2005.
- [53] S. Bossi, F. Mapelli, and R. Lancini. Semi-fragile watermarking for video quality evaluation in broadcast scenario. In IEEE International Conference on Image Processing, volume 1, pages 161–164, 11-14 September 2005.
- [54] S. Bossi, F. Mapelli, and R. Lancini. Objective video quality evaluation by using semi-fragile watermarking. In Proceedings of Picture Coding Symposium, pages 1154–1158, 4-7 December 2004.
- [55] F. Yang, X. Wang, Y. Chang, and S. Wan. A no-reference video quality assessment method based on digital watermark. In IEEE Proceeding on Personal, Indoor and Mobile Radio Communications, volume 3, pages 2707–2710, 7-10 September 2003.
- [56] M. Farias, S. Mitra, and M. Carli. Video quality objective metric using data hiding. In IEEE Workshop on Multimedia Signal Processing, pages 424–427, 9-11 December 2002.
- [57] F. Hartung and B. Girod. Watermarking of uncompressed and compressed video. Elsevier Journal: Singal Processing, 66(3):283–301, May 1998.
- [58] P. Campisi, M. Carli, G. Giunta, and A. Neri. Blind quality assessment system for multimedia communications using tracing watermarking. IEEE Transactions on Signal Processing, 51(4):996–1002, April 2003.

- [59] P. Campisi, M. Carli, G. Giunta, and A. Neri. Tracing watermarking for multimedia communication quality assessment. In IEEE International Conference on Communications, volume 2, pages 1154–1158, 22-25 September 2002.
- [60] S. Maity, M. K. Kundu, and P. Nandi. Watermarking scheme for blind quality assessment in multimedia mobile communication services. In Indian Conference on Computer Vision, Graphics and Image Processing, pages 376–381, 16-18 December 2004.
- [61] S. Maity, M. Kundu, and S. Maity. Dual purpose FWT domain spread spectrum image watermarking in real time. Computers and Electrical Engineering, 35(2):415–433, March 2009.
- [62] F. Benedetto, G. Giunta, and A. Neri. QoS assessment of 3G video-phone calls by tracing watermarking exploiting the new colour space ‘YST’. IET Communications, 1(4):696–704, August 2007.
- [63] M. Barni, F. Bartolini, and A. Piva. Improved wavelet-based watermarking through pixel-wise masking. IEEE Transactions on Image Processing, 10(5):783–791, May 2001.
- [64] T. Brandao and M. Queluz. Towards objective metrics for blind assessment of images quality. In IEEE International Conference on Image Processing, pages 2933–2936, 8-11 October 2006.
- [65] T. Brandao and M. Queluz. Blind PSNR estimation of video sequences through non-uniform quantization watermarking. In International Conference on Image Analysis and Recognition, volume 4141, pages 587–599, 18-20 September 2006.

- [66] E. Nezhadarya, Z. Wang, and R. Ward. Image quality monitoring using spread spectrum watermarking. In IEEE International Conference on Image Processing, pages 2233–2236, 7-10 November 2009.
- [67] N. Baaziz, D. Zheng, and D. Wang. Image quality assessment based on multiple watermarking approach. In IEEE International Workshop on Multimedia Signal Processing, pages 1–5, 17-19 October 2011.
- [68] Z. Wang, G. Wu, H. Sheikh, E. Simoncelli, E. Yang, and A. Bovik. Quality-aware images. IEEE Transactions on Image Processing, 15(6):1680–1689, June 2006.
- [69] A. Bhattacharya, S. Palit, N. Chatterjee, and G. Roy. Blind assessment of image quality employing fragile watermarking. In IEEE International Symposium on Image and Signal Processing and Analysis, pages 431–436, 4-6 September 2011.
- [70] S. Altous, M. Samee, and J. Gotze. Reduced reference image quality assessment for JPEG distortion. In IEEE International Symposium on Electronics in Marine, pages 97–100, 14-16 September 2011.
- [71] M. Samee and J. Gotze. Reduced reference image quality assessment for transmitted images using digital watermarking. In IEEE International Symposium on Image and Signal Processing and Analysis, pages 425 – 430, 4-6 September 2011.
- [72] M. Samee and J. Gotze. Increased robustness and ssecurity of digital watermarking using DS-CDMA. In IEEE International Symposium on Signal Processing and Information Technology, pages 189–193, 15-18 December 2007.
- [73] M. El-Mahallawy, A. Hashad, H. Ali, and H. Zaky. Quality estimation of video transmitted over an AWGN channel based on digital watermarking and wavelet

- transform. Journal of World Academy of Science, Engineering and Technology, 57(1):419–423, September 2009.
- [74] M. Farias, M. Carli, A. Neri, and S. Mitra. Video quality assessment based on data hiding driven by optical flow information. In Proceedings of SPIE Image Quality and System Performance, volume 5294, pages 190–200, 7-10 December 2004.
- [75] P. Campisi, G. Giunta, and A. Neri. Object-based quality of service assessment using semi-fragile tracing watermarking in MPEG-4 video cellular services. In IEEE International Conference on Image Processing, volume 2, pages 881–884, 22-25 September 2002.
- [76] K. Zeng and Z. Wang. Quality-aware video based on robust embedding of intra- and inter-frame reduced-reference features. In IEEE International Conference on Image Processing, pages 3229–3232, 26-29 September 2010.
- [77] S. Winkler. Visual fidelity and perceived quality: Toward comprehensive metrics. In Proceedings of SPIE, Human Vision and Electronic Image, volume 4299, pages 114–125, 22-25 January 2001.
- [78] A. Bovik (Editor). Handbook of image and video processing. ISBN: 978-0121197902. Academic Press, 2000.
- [79] D. Percival and A. Walden. Wavelet methods for time series analysis. ISBN: 0521640687. Cambridge University Press, 2000.
- [80] D. Sheskin. Handbook of parametric and nonparametric statistical procedures. ISBN:1-58488-133-X. 2nd Edition. Chapman & Hall CRC, 2000.

- [81] R. Tu and J. Zhao. A novel semi-fragile audio watermarking scheme. In IEEE International Workshop on Haptic, Audio and Visual Environments and their Applications, pages 89–94, 20–21 September 2003.
- [82] Dctune2.0. <http://vision.arc.nasa.gov/dctune/>, last visited in September 2011.
- [83] D. Taubman and M. Marcellin. JPEG2000: image compression fundamentals, standards and practice. ISBN: 0-7923-7519-X1. Kluwer Academic Publishers, 2002.
- [84] A. Said and W. Pearlman. A new, fast and efficient image codec based on set partitioning in hierarchical trees. IEEE Transactions on Circuits and Systems for Video Technology, 6(3):243–250, June 1996.
- [85] M. Jamzad and F. Yaghmaee. Achieving higher stability in watermarking according to image complexity. Scientia Iranica Journal, 13(4):404–412, April 2006.
- [86] F.Yaghmaie and M.Jamzad. Introducing a new method for estimation image complexity according to calculate watermark capacity. In IEEE International Conference on Intelligent Information Hiding and Multimedia Signal Processing, pages 981–984, 15–17 August 2008.
- [87] F.Yaghmaie and M.Jamzad. Computing watermark complexity in images according to their Quad-tree. In IEEE International Symposium on Signal Processing and Information Technology, pages 823–826, 18–21 December 2005.
- [88] A. Lewis and G. Knowles. Image compression using the 2-D wavelet transform. IEEE Transactions on Image Processing, 1(2):244–250, April 1992.

- [89] R. Safranek and J. Johnson. A perceptually tuned sub-band image coder with image dependent quantization and post-quantization data compression. In IEEE International Conference on Acoustics, Speech, and Signal Processing, volume 3, pages 1945–1948, 23-26 May 1989.
- [90] JVT. Draft ITU-T recommendation and final draft international standard of joint video specification (ITU-T rec. H.264-ISO/IEC 14496-10 AVC), March 2003.
- [91] S. Kwon, A. Tamhankar, and K Rao. Overview of H.264/MPEG-4 part10. Elsevier Journal: Visual Communication and Image Representation, 17(2):186–216, April 2006.
- [92] I. Richardson. H.264 and MPEG-4 video compression: video coding for next-generation multimedia. ISBN: 0-470-84837-5. John Wiley & Sons Inc., 2003.
- [93] Q. Zhang, Q. Peng, J. Chen, and C. Zhu. Video transmission channel distortion estimation model for H.264/AVC based on channel simulation. In International Conference on Multimeida and Ubiquitous Engineering, pages 64–69, 4-6 June 2009.
- [94] Y. Zhou, C. Hou, W. Xiang, and F. Wu. Channel distortion modeling for multi-view video transmission over packet-switched networks. IEEE Transcations on Circuits and Systems for Video Technology, 21(11):1679–1692, November 2011.
- [95] YUV video download. <http://media.xiph.org/video/derf/>, last visited in June 2012.
- [96] H.264/AVC reference codec. <http://iphone.hhi.de/suehring/tml/>, last visited in January 2012.

- [97] K. Kumazoe, K. Kawahara, T. Takine, and Y. Oie. Analysis of packet loss in transport layer over ATM networks. In Global Telecommunications Conference: The Key to Global Prosperity, volume 1, pages 42–49, 18–22 November 1996.
- [98] W. Wang, Z. Xia, H. Cui, and K. Tang. Robust H.264/AVC transmission with optimal mode selection and data partitioning. In IEEE International Symposium on Communications and Information Technology, volume 2, pages 1444–1447, 12–14 October 2005.
- [99] M. Dissanayake, C. Hewage, S. Worrall, W. Fernando, and A. Kondoz. Redundant motion vectors for improved error resilience in H.264/AVC coded video. In IEEE International Conference on Multimedia and Expo, pages 25–28, 23–26 June 2008.
- [100] R. Razavi, M. Fleury, M. Altaf, H. Sammak, and M. Ghanbari. H.264 video streaming with data-partitioning and growth codes. In IEEE International Conference on Image Processing, pages 909–912, 7–10 November 2009.
- [101] T. Stockhammer and M. Bystrom. H.264/AVC data partitioning for mobile video communication. In IEEE International Conference on Image Processing, volume 1, pages 545–548, 24–27 October 2004.
- [102] B. Barmada, M. Ghandi, E. Jones, and M. Ghanbari. Prioritized transmission of data partitioned H.264 video with hierarchical QAM. IEEE Signal Processing Letters, 12(8):577–580, August 2005.
- [103] Y. Chen, K. Yu, J. Li, and S. Li. An error concealment algorithm for entire frame loss in video transmission. In Picture Coding Symposium, pages 1–4, 15–17 December 2004.

- [104] J. Zheng and L. Chau. A temporal error concealment algorithm for H.264 using Lagrange interpolation. In International Symposium on Circuits and Systems, volume 2, pages 133–136, 23-26 May 2004.
- [105] D. Kim, S. Yang, and J. Jeong. A new temporal error concealment method for H.264 using adaptive block sizes. In IEEE International Conference on Image Processing, volume 3, pages 928–931, 11-14 September 2005.
- [106] Q. Peng, T. Yang, and C. Zhu. Block-based temporal error concealment for video packet using motion vector extrapolation. In IEEE International Conference on Communications, Circuits and Systems and West Sino Expositions, volume 1, pages 10–14, 29 June - 1 July 2002.
- [107] Y. Chen, Y. Hu, O. Au, H. Li, and C. Chen. Video error concealment using spatio-temporal boundary matching and partial difference equation. IEEE Transactions on Multimedia, 10(1):2–15, January 2008.
- [108] L. Atzori, F. De-Natale, and C. Perra. A spatio-temporal concealment technique using boundary matching algorithm and mesh-based warping. IEEE Transactions on Multimedia, 3(3):326–338, September 2001.
- [109] E. Kim, S. Choi, and H. Kim. Weighted boundary matching algorithm for error concealment in the MPEG-2 video bit stream. Elsevier Journal: Signal Processing, 73(3):291–295, January 1999.
- [110] O. Nemethova, A. Al-Moghrabi, and M. Rupp. An adaptive error concealment mechanism for H.264/AVC encoded low-resolution video streaming. In European Signal Processing Conference, pages 1–5, 4-8 September 2006.

- [111] H. Ha, C. Yim, and Y. Kim. Packet loss resilience using unequal forward error correction assignment for video transmission over communication networks. Elsevier Journal: Computer Communications, 30(18):3676–3689, December 2007.