THESE

Pour l'obtention du Grade de

DOCTEUR DE L'UNIVERSITE DE POITIERS

(Faculté des Sciences Fondamentales et Appliquées)
(Diplôme National - Arrêté du 25 mai 2016)

Ecole Doctorale : SISMI

Secteur de Recherche : Traitement du Signal et des images

Présentée par : Xinwen XIE

.....

****

Quality Strategy and Method for Transmission: Application to Image

****

****

Soutenue le 10th Januray 2019

devant la Commission d'Examen

****

JURY

Abdelhakim SAADANE	Rapporteur, Polytech Nantes
Patrick CORLAY	Rapporteur, Université Polytechnique Hauts-de-France
Frédéric DUFAUX	Président, Laboratoire L2S Supelec
Philippe CARRE	Directeur de thèse, Université de Poitiers
Jianhua WU	CoDirecteur de thèse, Nanchang University
Clency PERRINE	Coencadrant Université de Poitiers
Yannis POUSSET	Examinateur , Université de Poitiers

Acknowledgements

Firstly, I would like to appreciate all of my supervisors, professor Philippe CARRE and professor Nanrun ZHOU, and accociate professor Clency PERRINE as well as professor Yannis POUSSET and professor Jianhua WU. At all stages of my PhD study, they always provide clear direction and valuable advice. Secondly, I would like to thank all the colleagues in ICONES group of the Xlim Institute of the University of Poitiers as well as colleagues in the Department of Electronic Information Engineering of Nanchang University. The fruitful discussions with them are helpful for extension of thinking and inspiration. A special thank devotes to all the volunteers participating in the experiment for database construction. They are PhD and engineering students of the University of Poitiers, postgraduate students of Nanchang University, and undergraduate students and teachers of Jiujiang University.

I would also like to thank my family. I would like to dedicate this thesis to my father in heaven. His encouragement is the driving force of my persistence. During my Ph.D. study, my mother, Zhilian CHEN, and my wife, Wan CAO, have to spend more time to take care of two kids. Without their devotion, I do not have enough time to complete my dissertation. I want to thank my two kids, Yucheng XIE and Yuxi XIE, since they give me much pleasure time. I would like to thank all my friends who provided helps for me during the PhD study. Special thanks to Yu FAN, Pascal LEFEVRE and Chennan JIANG for their important help in various official application procedures.

I want to thank the University of Poitiers in France and the Nanchang University in China for providing the opportunity of joint training of PhD thesis, and I appreciate Jiujiang University for the financial support.

This work is supported by the National Natural Science Foundation of China (Grant Nos. 61462061 and 61662047) and the Major Academic Discipline and Technical Leader of Jiangxi Province (Grant No. 20162BCB22011).

Abstract

This thesis focuses on the study of image quality strategies in wireless communication systems and the design of new quality evaluation metrics:

Firstly, based on statistical model in complex wavelet transform domain, a new reduced-reference image quality metric has been proposed. The magnitude and the relative phase information of the Dual-tree Complex Wavelet Transform coefficients is modelled by using probability density function and the parameters of models served as the reduced-reference features will be transmitted to the receiver. Then, the mapping relation between the reduced-reference features and a objective score is constructed with a Generalized Regression Neural Network approach.

Secondly, with the new metric, a new decoding strategy is proposed for a realistic wireless transmission system, which can improve the quality of experience (QoE) while ensuring the quality of service (QoS). For this, a new database including large physiological vision tests has been constructed to collect the visual preference of people when they are selecting the images with different decoding configurations, and a classifier based on support vector machine or K-nearest neighboring is utilized to automatically select the decoding configuration.

Finally, according to the specific property of distortions and people's preference, an improved metric has been proposed by combining global features with local features. It is demonstrated that the improved metric performs well in optimization of the decoding strategy.

The experimental results validate the effectiveness of the proposed image quality metrics and the quality strategies.

Keywords: Image quality metric, Statistical feature, Dual-tree complex wavelet transform, Kullback-Leibler divergence, Quality of experience, Decoding strategy, Machine learning, Image transmission, Realistic wireless channel.

Résumé

Cette thèse porte sur l'étude des stratégies d'amélioration de la qualité d'image dans les systèmes de communication sans fil et sur la conception de nouvelles métriques d'évaluation de la qualité.

Tout d'abord, une nouvelle métrique de qualité d'image à référence réduite, basée sur un modèle statistique dans le domaine des ondelettes complexes, a été proposée. Les informations d'amplitude et de phase relative des coefficients issus de la transformée en ondelettes complexes sont modélisés à l'aide de fonctions de densité de probabilité. Les paramètres associés à ces fonctions constituent la référence réduite qui sera transmise au récepteur. Ensuite, une approche basée sur les réseaux de neurones à régression généralisée est exploitée pour construire la relation entre les caractéristiques de la référence réduite et le score objectif.

Deuxièmement, avec la nouvelle métrique, une nouvelle stratégie de décodage est proposée pour la transmission d'image sur un canal de transmission sans fil réaliste. Ainsi, la qualité d'expérience (QoE) est améliorée tout en garantissant une bonne qualité de service (QoS). Pour cela, une nouvelle base d'images a été construite et des tests d'évaluation subjective de la qualité de ces images ont été effectués pour collecter les préférences visuelles des personnes lorsqu'elles sélectionnent les images avec différentes configurations de décodage. Un classificateur basé sur les algorithmes SVM et des k plus proches voisins sont utilisés pour la sélection automatique de la meilleure configuration de décodage.

Enfin, une amélioration de la métrique a été proposée permettant de mieux prendre en compte les spécificités de la distorsion et la préférence des utilisateurs,. Pour cela, nous avons combiné les caractéristiques globales et locales de l'image conduisant ainsi à une amélioration de la stratégie de décodage.

Les résultats expérimentaux valident l'efficacité des métriques de qualité d'image et des stratégies de transmission d'images proposées.

Mots-clés: Métrique de qualité d'image, Caractéristique statistique, Transformation en ondelettes complexes à deux arbres, Divergence de Kullback-Leibler, JPEG2000 Wireless, Qualité de l'expérience, Stratégie de décodage, Apprentissage automatique, transmission d'image, canal sans fil réaliste.

Table of contents

List of figures vii

List of tables xi

Nomenclature xiii

1 Introduction 1
- 1.1 Context and motivation . . . 1
- 1.2 Purpose of the thesis . . . 2
- 1.3 Thesis organization . . . 3

2 Background and state-of-the-art 5
- 2.1 Image quality assessment . . . 5
  - 2.1.1 Subjective IQA . . . 6
  - 2.1.2 Objective IQA methods and the state-of-the-art . . . 8
  - 2.1.3 Evaluation of IQA method . . . 11
- 2.2 Wavelet transform . . . 16
  - 2.2.1 Discrete Wavelet Transform . . . 16
  - 2.2.2 Dual-Tree Complex Wavelet Transform . . . 17
- 2.3 Wireless channel . . . 24
  - 2.3.1 Channel properties . . . 24
  - 2.3.2 Stochastic models . . . 25
  - 2.3.3 Deterministic models . . . 27
  - 2.3.4 Adaptive strategies to guarantee QoS . . . 30
- 2.4 Chapter summary . . . 31

3 Proposed metric for image quality assessment 33
- 3.1 Motivation . . . 33
- 3.2 Framework of the RR-IQA metric FMRP . . . 34

3.3 The crucial features . . . 36
3.4 Modelization . . . 38
3.4.1 Optimizing the histogram by IC . . . 39
3.4.2 Model of magnitude . . . 40
3.4.3 Model of relative phase . . . 42
3.5 Influence of image degradation on RR features . . . 44
3.6 Machine learning method for regression . . . 47
3.7 Test results . . . 50
3.7.1 Test databases and evaluation criteria . . . 50
3.7.2 Decomposition scales and the features . . . 50
3.7.3 Comparison with other metrics . . . 53
3.7.4 Robustness and cross-validation . . . 55
3.7.5 Complexity and two inaccurate examples . . . 56
3.8 Chapter summary . . . 59

4 Optimization of JPWL decoding in realistic wireless transmission 61
4.1 Motivation . . . 62
4.2 Global scheme and classifier deployment . . . 66
4.2.1 Global scheme . . . 66
4.2.2 Classifier deployment . . . 67
4.3 Database construction . . . 68
4.3.1 Reference and distorted images . . . 68
4.3.2 Testing methodology and environment . . . 70
4.3.3 Post-processing and properties . . . 72
4.4 Assessment of representative metrics with database . . . 76
4.4.1 Classifier design . . . 76
4.4.2 Effectiveness validation . . . 77
4.4.3 Inaccurate samples analysis . . . 80
4.5 A new metric FMRP-LSD . . . 81
4.5.1 FMRP feature and LSD feature . . . 81
4.5.2 Transmission binary cost . . . 88
4.5.3 Improvement for decoding strategy . . . 92
4.5.4 Validation on comprehensive database . . . 94
4.6 Chapter summary . . . 95

5 Conclusion and perspective **97**
5.1 Summary of thesis 97
5.2 Future research 98

References **101**

List of figures

2.1 Quality and impairment scales and corresponding scores: (a) Quality scale and (b) Impairment scale. 6
2.2 Two pairwise images evaluated with MOS and DMOS, where (a) and (b) are from TID2013 database, and (c) and (d) are from LIVE database 2, with MOS measuring the similarity between the distorted image and the reference one (range: scaled to 0-8, larger MOS means less distortion), and DMOS measuring the difference between the distorted image and the reference one (range: scaled to 0-100, larger DMOS means larger distortion). 7
2.3 Simplified framework for the deployment of RR-IQA systems. 9
2.4 Comparison of scatter plot with higher and lower PLCC: (a) A scatter plot with higher PLCC and (b) A scatter plot with lower PLCC. 12
2.5 Comparison of scatter plot with higher and lower SROCC value: (a) A scatter plot with higer SROCC and (b) A scatter plot with lower SROCC. 13
2.6 Comparison of scatter plot with higher and lower OR values: (a) A scatter plot with lower OR and (b) A scatter plot with higher OR. 14
2.7 Analysis filter bank of the DT-CWT. 19
2.8 Six orientation sub-bands of DT-CWT with real and imaginary parts. 19
2.9 The magnitude of DT-CWT coefficients and its histogram: (a) The magnitude of image "caps", (b) The histogram, and (c) The approximation of histogram. 21
2.10 The phase of DT-CWT coefficients and its histogram: (a) Section of image "Boats", (b) Phase of the DT-CWT coefficients, (c) The corresponding histogram, and (d) The histogram of relative phase (π shifted histogram). . 23
2.11 Multipath propagation in wireless channels. 25
2.12 Illustration of ray-launching model. 28
2.13 Rx far away from Tx may be missed by rays. 29
2.14 Illustration of ray-tracing model. 29

3.1 Framework of the proposed RR-IQA method for a wireless system. 35

3.2 Test images from LIVE database 2. 36
3.3 Natural scene image "Plane" and its histograms of magnitude and relative phase in the DT-CWT domain. (a) Original image, (b) Histogram of magnitude, (c) Histogram of relative phase, (e) Histogram of magnitude after thresholding, and (f) Histogram of relative phase after thresholding. 38
3.4 Comparison of the histograms using and without using IC (Histograms taken from the image "Bikes", the decomposition scale is 1, and the orientation is 15 degree): (a) Empirical histogram of the magnitude without using IC, (b) Empirical histogram of the relative phase without using IC, (c) Optimal histogram of the magnitude using IC and (d) Optimal histogram of the relative phase using IC. 40
3.5 Comparison of two PDF candidates (Histogram taken from the image "Bikes", the decomposition scale is 1, and the orientation is 45 degree). (a) Curve fitting of MRD and (b) Curve fitting of IGD. 42
3.6 Comparison of two PDF candidates (Histogram taken from the image "Bikes", the decomposition scale is 1, and the orientation is 15 degree). (a) Curve fitting of WCD, (b) Curve fitting of WCD, histogram after a phase shift of π, and (c) Curve fitting of GGD, histogram after a phase shift of π. 44
3.7 Comparison of the PDFs of magnitude and relative phase from different distortions. The first column shows the original image and the distorted images, the second column shows the PDF of the magnitude, and the third column shows the PDF of the relative phase. (The red dashed curve and the blue solid curve are the PDF of the model and that of the distribution, respectively.) 46
3.8 The objective scores of FMRP of image "Buildings" in LIVE database 2. . . 49
3.9 Two inaccurate examples of the FMRP metric in TID2013 database. 58

4.1 Content-based hierarchical transmission system with UPA and UEP technologies. 62
4.2 Image quality comparsion, where four images are obtained over a realistic wireless channel, (b) and (d) have better QoS, and (a) and (c) have better QoE. 63
4.3 The quality comparison between the $l-1$ layers and the l layers decoding images, where (a) decoded with $l-1$ layers and (b) decoded with l layers ($l=2$, because of the bad channel condition), and (b) has a better quality than (a). Figure (c) decoded with $l-1$ layers and (d) decoded with l layers ($l=4$, because of the good channel condition), but (d) has a worse quality than (c). 65
4.4 The global scheme of realistic MIMO wireless channel. 66

4.5 Optimization of decoding configuration with the classifier. 68
4.6 The reference images used in the database. 69
4.7 The topology of the experimental environment and the gain variation of the system: (left) Topology of the transmission scene and (right) gain evolution of the MIMO channel. 70
4.8 The interface used in the experiment, $-2, 0, 2$ represent 3 classifications. . . 71
4.9 The optimization of decoding configuration with a classifier. 73
4.10 Image with blur and ringing distortion. 75
4.11 Image with local and random distortion. 75
4.12 Classifier design for classic IQA metric. 76
4.13 Comparison of classification results of HVS, PSNR and FMRP, where all the points correspond to the classification label -2 and the number of isolated points indicates the degree of difference. 78
4.14 Comparison of two misclassified samples, where for the HVS, (a) has a better quality than (b), and (c) and (d) have the same quality; for the FMRP metric, (a) and (b) have the same quality, and (c) has a better quality than (d). . . . 81
4.15 Comparison of two images with different decoding configurations: (a) and (b) are respectively decoded with 3 and 4 layers configuration, (c) and (d) are the comparison of magnitude's PDFs between the model and the image, (e) and (f) are the comparison of relative phase's PDFs between the model and the image. 83
4.16 Block diagram of new metric. 85
4.17 Image blocking method and LSD map, where "Std 1" denotes the standard deviation of luminance value for block 1. 86
4.18 Comparison of three features for classification, where KLD is the mean of all sub-bands, DCC is the new feature for detecting local distortion, and S is the preference score of HVS. 87
4.19 The influences of parameter variation on the model error: (a) Variation of parameter λ and (b) Variation of parameter μ_v. 91
4.20 Two wrongly classified samples for FMRP metric, which are correctly classified by FMRP-LSD metric. 93

List of tables

2.1 Information of five extensively used image databases 16

3.1 Comparison of KLD of the models MRD and IGD. 41
3.2 Comparison of KLD of the GGD and WCD models. 43
3.3 Information of the image databases. 50
3.4 Prediction accuracy of 2, 3 and 4 decomposition scales. 51
3.5 KLD of the sub-bands of the magnitude (from image "Building2" of LIVE database 2). 51
3.6 KLD of the sub-bands of the relative phase (from image "Buildings2" of LIVE database 2). 52
3.7 Comparison of the features of DWT and DT-CWT. 53
3.8 Performance comparison of the objective quality metrics. 54
3.9 Performance comparison of RR metrics for different distortions. 55
3.10 Performance of cross-validation for FMRP. 56
3.11 Performance of cross-database validation for FMRP and two other RR metrics. 56
3.12 The size of features of different metrics. 57
3.13 Time consumption of different metrics. 57

4.1 Subjective evaluation sets 71
4.2 The demographic data of subjective assessors 72
4.3 Classified statistics of the database 73
4.4 Performance of the objective quality metrics based on the proposed database 78
4.5 Performances of decoding strategies (ratio of correct classification samples based on the HVS) 80
4.6 KLDs of six sub-bands of images (a) and (b) 82
4.7 Variation of accuracy with feature dimensions 89
4.8 Classification accuracy varies with feature dimensions 89
4.9 Determination of quantization bits 91

4.10 The quantization bits of each parameter 92
4.11 Comparison of decoding strategies with FMRP, DCC and FMRP-LSD . . . 94
4.12 Comparison between FMRP and FMRP-LSD metrics on TID2013* database 94

Nomenclature

Acronyms / Abbreviations

ADI	Average Direction Information
BER	Bit Error Rate
CL	Closed Loop
CSI	Channel State Information
CWT	Complex Wavelet Transform
DCT	Discrete Cosine Transform
DMOS	Difference Mean Opinion Score
DNT	Divisive Normalization Transform
DT-CWT	Dual-Tree Complex Wavelet Transform
DWT	Discrete Wavelet Transform
ECC	Error-Correcting Codes
ECOC	Error-Correcting Output Codes
EPB	Error Protect Block
FB	Filter Bank
FMRP	Feature of Magnitude and Relative Phase
FMRP-LSD	Metric with FMRP and LSD features
FR	Full Reference

GGD	General Gaussian Distribution
GRNN	Generalized Regression Neural Network
HVS	Human Visual System
IGD	Inverse Gaussian Distribution
IQA	Image Quality Assessment
ISI	Inter-Symbol Interference
JP2K-CD	JPEG 2000 Compression Distortion
JP2K-TED	JPEG 2000 Transmission Error Distortion
JPWL	JPEG 2000 WireLess
KLD	Kullback-Leibler Divergence
K-NN	K-Nearest Neighbors
LOS	Line-Of-Sight
LSD	Local Standard Deviation
LTE	Long-Term Evolution
MIMO	Multiple-Input-Multiple-Output
MOS	Mean Opinion Score
MRD	Mixture Rayleigh Distribution
NLOS	Non-Line-Of-Sight
NR	No Reference
OFDM	Orthogonal Frequency Division Multiplexing
OR	Outlier Ratio
PDF	Probability Density Function
PLCC	Pearson Linear Correlation Coefficient
QoE	Quality of Experience

QoS	Quality of Service
RBF	Radial Basis Function
RR	Reduced Reference
Rx	Receiver
SNR	Signal Noise Ratio
SROCC	Spearman Rank-Order Correlation Coefficient
SSIM	Structural SIMilarity
SVM	Support Vector Machine
SVR	Support Vector Regression
Tx	Transmitter
UEP	Unequal Error Protection
UPA	Unequal Power Allocation
VQEG	Video Quality Expert Group
WBRRM	Wavelet-Based RR Method
WCD	Wrapped Cauchy Distribution

Chapter 1

Introduction

1.1 Context and motivation

Over the last two decades, the rapid development of multimedia wireless communication technology has greatly promoted the multimedia services and devices, for example, digital cameras, tablet and mobile smart-phones as well as the growth of wireless network facilities. On the other hand, the popularity of high-quality multimedia services and devices, in turn, requires the multimedia wireless communication system with better Quality of Service (QoS) and better Quality of Experience (QoE). However, the better QoS and the better QoE conflict with the well-known constraints in wireless transmission such as multipath frequency selective fading due to multipath transmission, Doppler effect because of the mobility of the users, and limited bandwidth. These challenges bring the research boom of communication technologies including kinds of methods to exploit the time, space and frequency domain diversity to overcome the constraints. However, all these technologies only consider the transmission parameters to improve the QoS of systems and do not involve the QoE of terminal users. It is highly desirable to introduce image quality assessment in the wireless communication system to improve the QoE of users.

Image quality assessment (IQA) has become one of important research fields of image processing, thanks to the explosive growth of images and videos as well as relevant applications since the beginning of the 21st century. Objective IQA method is a method to assess image degradations by computer algorithms in a way analogous with the Human Visual System (HVS), automatically. The method generally quantifies image degradations by comparing the visual-relevant features extracted from the reference image and the distorted one. According to the availability of a reference image, objective IQA methods can be classified into Full Reference IQA (FR-IQA), Reduced Reference IQA (RR-IQA) and No Reference IQA (NR-IQA) methods. The FR-IQA method needs full access to the reference

image to evaluate the image quality. It is difficult to apply the FR-IQA metric for image quality evaluation in a wireless communication system, since it is impossible to access the reference image at the receiver in most cases. The NR-IQA metric is a good candidate but NR-IQA metrics are usually designed for specific applications and they are only available for few particular distortions [53, 77]. Another good candidate method is RR-IQA metric, which evaluates image quality with partial statistical information and usually has a higher accuracy and a better robustness than the NR-IQA method. Most of RR-IQA metrics [53, 77, 108] are based on the statistical model in the wavelet domain to summarize the image features. However, with the development of wavelet theory, the more efficient tool, complex wavelet transform, has been proposed and widely used in image processing. In this study, we assume that the RR-IQA metric in the complex wavelet domain will have better performance than the one in the classic wavelet domain. Thus, in this thesis, we focus on the RR-IQA metric and design new RR-IQA metrics in the complex wavelet domain to evaluate the image quality.

To improve the QoE of users, the RR-IQA metric will be embedded in the realistic wireless communication system to detect the distortions. A database is necessary for validating the effectiveness of the RR-IQA metric. However, most of the existing public databases do not contain the distort ion occurring in the wireless communication system. Although the representative databases LIVE 2 [91] and TID2013 [71] contain the distortions occurring in wireless communication systems, they were generated under a simulation environment rather than a realistic environment. Therefore, to verify the effectiveness of the RR-IQA metric, a new database including the distortions which occur in wireless communication will be constructed in this study.

In the wireless communication systems, classical transmission strategies guarantee the QoS by adjusting relevant system parameters. However, these strategies do not take into account the QoE of users. In this thesis, we propose a quality transmission strategy to improve the QoE by using RR-IQA metrics.

1.2 Purpose of the thesis

Based on the theories of image quality assessment, wavelet transform, computer vision and wireless communication, the RR-IQA metric and its application in a realistic wireless communication channel for optimizing the users' QoE are investigated in the thesis. The main purpose of the thesis are as follows:

1. Design of RR-IQA metric in the wavelet domain. By utilizing the advantages of multi-scale and multi-direction of the complex wavelet transform (CWT), the new RR-IQA metric will be designed to detect the distortions occurring in JPEG 2000

wireless transmission (JPWL-transmission). For this, the influences of the magnitude and the phase feature on image quality will be analyzed. Moreover, to summarize the image information, some statistical models for the distributions of the magnitude and the phase will be compared. Based on the information criterion, a strategy is designed to reduce the modelling error.

2. Database construction based on the distortions occurring in wireless communication channels. A new database including JPWL-transmission distortions and a large physiological visual test will be constructed. The effectiveness of different RR-IQA metrics will be verified based on the new database. The new database will provide a benchmark to judge whether the IQA metric can improve the QoE in wireless communication systems or not.

3. Investigation of the characteristics of JPWL-transmission distortions. Based on the new database, the characteristics of JPWL-transmission distortions will be studied and the relevant factors causing the characteristics will be revealed. According to the characteristics of the distortions, a new feature based on the local-standard-deviation map will be designed to detect the distortions.

4. Optimization decoding strategy with RR-IQA metric. The proposed RR-IQA metrics will be embedded in the wireless communication system to optimize the decoding strategy by selecting a better-quality image to improve the QoE. The performance and the transmission cost of the metric will be analyzed and discussed.

1.3 Thesis organization

The thesis is arranged as follows: Chapter 2 provides background and state-of-the-art including image quality assessment, classic wavelet transform and complex wavelet transform, and properties and models of wireless channel.

Chapter 3 proposes a new RR-IQA metric based on the statistical model in the complex wavelet domain. The metric utilizes the magnitude and the relative phase information of a complex wavelet transform to detect the image degradation. The parameters of the models of magnitude and relative phase serve as the crucial features, and the influence of image degradation on the features is detailed. The machine learning method for the objective score regression is also introduced in this chapter. The experimental validation and the summary are presented at the end of the chapter.

Chapter 4 is about the optimization of JPWL decoding in realistic wireless transmission. The motivation and the global scheme to apply the new RR-IQA metric are introduced. Then,

a new database is constructed to provide an evaluation benchmark for the decoding strategy and the details of the database construction are presented. Another new RR-IQA metric is proposed to handle the specific distortions that occur in JPWL transmission. At the end of the chapter, the comparative experiments are conducted to verify the performance of the decoding strategy with the new RR-IQA metric.

The conclusion and the perspective are presented in Chapter 5.

Chapter 2

Background and state-of-the-art

Image quality assessment is an important research direction in the fields of image processing and computer vision. How to assess the image quality automatically, accurately and efficiently by an objective algorithm has been the target of the researchers for years. In the subsequent chapters, we will investigate the image quality assessment metric based on the statistical features and apply it to the wireless communication system to improve quality of experience in a consistent way with users. This chapter first introduces the theoretical and technical basis of image quality assessment, and then introduces the basic knowledge and crucial technical indicators of wireless communication systems. Due to the limited space of the thesis, we only briefly introduce the knowledge that is closely related to subsequent chapters.

This chapter is organized as follows: Section 2.1 introduces the subjective and the objective IQA methods and the validation way. The wavelet transform and its main features, i.e., the foundation of statistical IQA methods, are described in Section 2.2. The wireless channel properties, indicators and corresponding technologies are introduced in Section 2.3. The chapter summary is presented in Section 2.4.

2.1 Image quality assessment

Thanks to the rapid development of computer, communication and network technologies, Image Quality Assessment (IQA) has become an important issue in many applications such as image acquisition system, storage medium, processing system or transmission equipment. Therefore, to maintain and control the quality of the image, it is important for IQA to be able to identify and quantify image quality degradation.

IQA methods can be divided into two categories: subjective evaluation methods by observers and objective evaluation methods by computer algorithms. The two kind of categories will be introduced in this section.

2.1.1 Subjective IQA

Subjective IQA method mainly evaluates the image degradation by observers and quantifies the degradation by a subjective score. According to the standard ITU-R BT. 500-11 [76], the process of the evaluation should be:

1. Construct an appropriate test environment, including requirements for light intensity of indoor environment, display equipment, subjects' vision, the distance between the subject and the display, test methodology and evaluation criteria.
2. Select a series of images obtained in different application scenarios with different levels of degradations (the degradation levels should cover all quality-degraded levels specified by the recommends of International Telecommunication Standardization Sector (ITU-T)),
3. Require the subjects to evaluate the image quality according to the quantified scales (the tests should be grouped and done in stages to avoid vision fatigue),
4. Obtain the average score of all the subjects as the final result (some abnormal subjective scores should be discarded with an outlier detection).

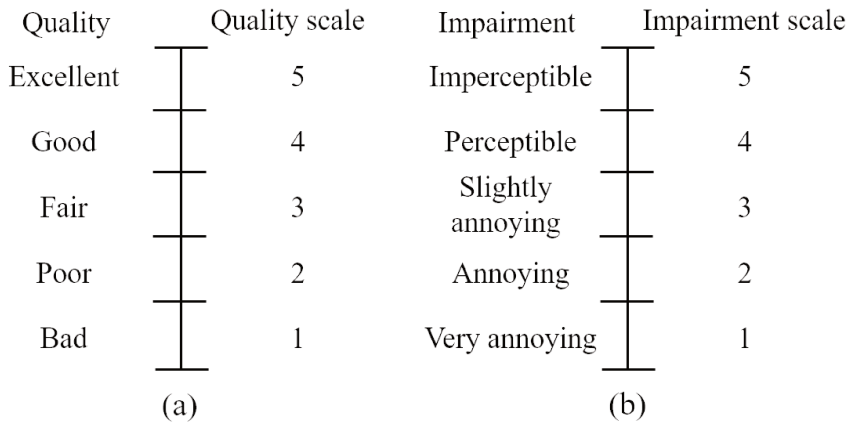
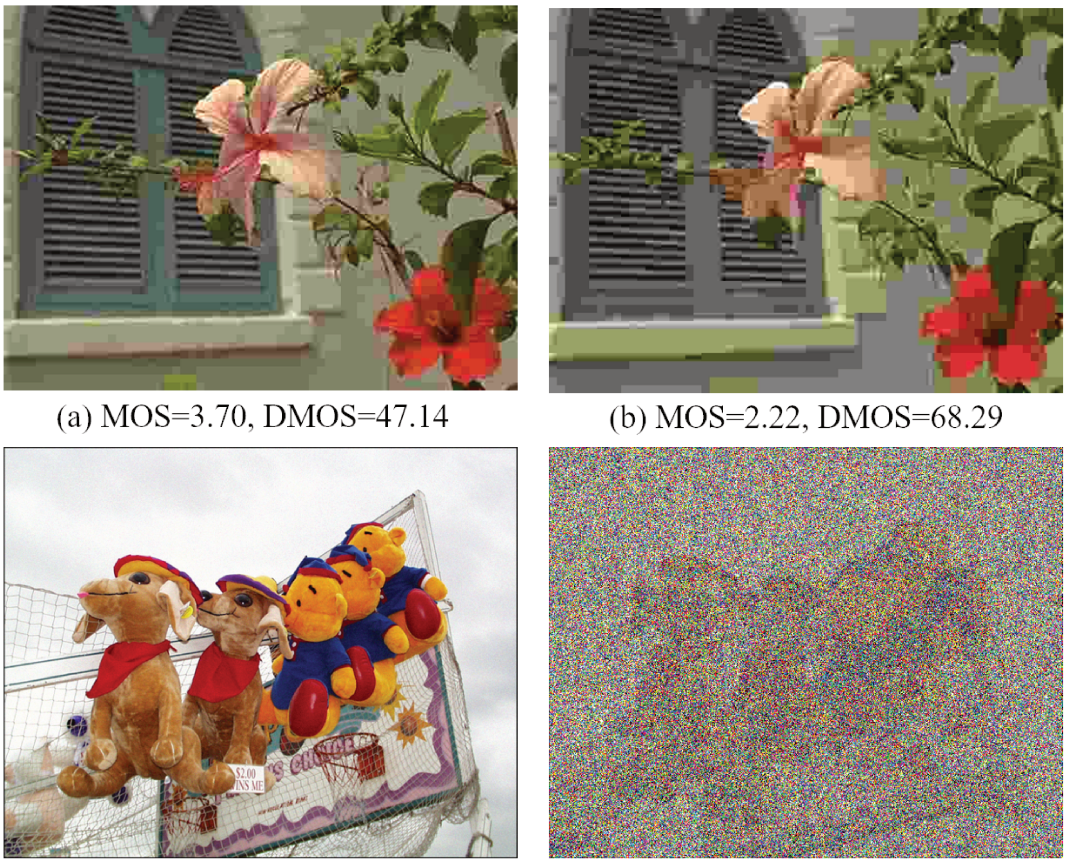


Fig. 2.1 Quality and impairment scales and corresponding scores: (a) Quality scale and (b) Impairment scale.

The main quantified scales are quality scale and impairment scale, according to which subjects can mark the image quality with quality score or difference score, as shown in Fig. 2.1. There are many subjective IQA methods [44] applied to different fields. The Mean Opinion Score (MOS) and the Difference MOS (DMOS) are the most representative

subjective evaluation methods [44]. These methods evaluate the image quality based on evaluations from a number of subjects. The MOS measures the similarity of the distorted image and reference image (larger MOS means less distortion) while the DMOS measures the difference (larger DMOS means larger distortion) as shown in Fig. 2.1.



(a) MOS=3.70, DMOS=47.14 (b) MOS=2.22, DMOS=68.29

(c) MOS=5.579, DMOS=20.30 (d) MOS=1.99, DMOS=71.46

Fig. 2.2 Two pairwise images evaluated with MOS and DMOS, where (a) and (b) are from TID2013 database, and (c) and (d) are from LIVE database 2, with MOS measuring the similarity between the distorted image and the reference one (range: scaled to 0-8, larger MOS means less distortion), and DMOS measuring the difference between the distorted image and the reference one (range: scaled to 0-100, larger DMOS means larger distortion).

Human subjective perception is the most reliable IQA method since the Human Visual System (HVS) is the final receiver or observer in many image applications. Subjective IQA is the standard method when formulating the relevant international standards (for example, standard of compression).

Subjective IQA method is the commonly used and most effective method for image quality evaluation. The advantages include the easily established quality scale, the convenience of implementation, and the full consideration of the HVS's perception and understanding of

the image. However, such a method is not only time-consuming and labor-intensive but also unable to be incorporated into the automatic image system. Therefore, it is desirable to design objective quality evaluation methods to assess the quality of images automatically. The subjective IQA methods currently are utilized to build the test database for studying the performance of the objective IQA methods.

2.1.2 Objective IQA methods and the state-of-the-art

2.1.2.1 Categories of objective IQA methods

Objective IQA methods usually assess the image degradation by the computer algorithm, automatically. Generally, the method quantifies the image degradation by comparing the visual-relevant features extracted from the reference image and the distorted one. According to the availability of a reference image, objective IQA methods can be classified into three subcategories: (1) Full Reference IQA (FR-IQA) methods [14, 30, 89, 111, 116], in which full information of the reference image should be available to evaluate the quality of distorted images, (2) Reduced Reference IQA (RR-IQA) methods [53, 54, 101, 108, 112, 118], in which only partial statistical information is required, and (3) No Reference IQA (NR-IQA) methods [52, 63, 64, 80], by which one can evaluate the image quality without the reference image.

In general, the FR-IQA has the best accuracy since it uses more information of the reference image than the other two. However, its application is limited to the occasion where a reference image can be obtained. RR/NR IQA has drawn more attention of researchers due to the often-impractical requirement of providing the full reference image by FR-IQA, particularly in communication applications [118]. We are quite interested in RR-IQA since we believe that with further research, RR-IQA may approach the state-of-the-art FR-IQA in view of the accuracy of image quality evaluation, particularly considering that the Human Visual System (HVS) has visual redundancy and is more sensitive to the macroscopic features of the image.

The RR-IQA methods are the practical and convenient tool for quality evaluation in real-time visual communication over wired or wireless networks. They achieve high accuracy with less information by extracting some reliable statistical features from the reference image as a basis for quality assessment. A framework for the deployment of RR-IQA in real-time visual communications is shown in Fig. 2.3. The system includes a feature extraction process at the transmitter side, and a feature extraction and analysis process at the receiver side. The extracted reduced-reference (RR) features have a much lower data rate than the image data and are typically transmitted to the receiver by an auxiliary and robust channel. Therefore,

it is often assumed that the auxiliary channel is error-free (for this, we use error-correcting code).

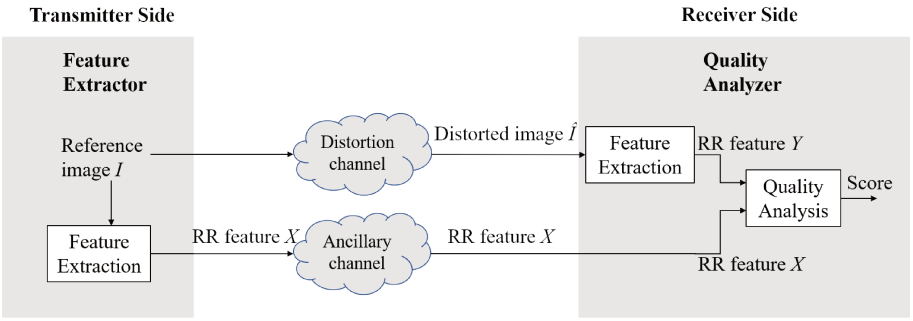


Fig. 2.3 Simplified framework for the deployment of RR-IQA systems.

The key issue of an RR-IQA method is to find the RR features which can summarize the image and are sensitive to various image distortions. Over the last decade, some researchers have contributed in RR-IQA and developed a number of algorithms. All the objective RR-IQA algorithms try to quantify the image quality by the quality-perceptual features and thus, according to different domains from which features are extracted, they can be classified into three categories: (1) the approaches aiming to evaluate the quality with spatial features, i.e., luminance, contrast and structural information in image pixel space [17, 18, 98], (2) the approaches based on features of transform domains, such as Discrete Cosine Transform (DCT) domain [57], log-Gabor transform domain [56] and wavelet domain [54, 55, 108, 112], and (3) the hybrid approaches based on integrated features extracted from spatial and transform domains [77, 118].

2.1.2.2 State-of-the-art of the RR-IQA methods

As introduced in the previous subsection, the objective RR-IQA methods can be divided into three categories. For the first class, the RR-IQA algorithms try to evaluate the image quality with spatial features such as the difference of pixel values (luminance), the standard deviation of neighboring areas (contrast) and the correlation coefficient of images (structural information). Based on the successful FR-IQA metric, i.e., SSIM [111], Tagliasacchi et al. designed an RR-IQA metric which mainly evaluates the image similarity by luminance, contrast and structure information [98]. In [17, 18], Carnec et al. proposed an RR-IQA metric based on luminance and contrast features and these features were adjusted by a Gamma function and Daly's contrast sensitivity function to pursue perceptual consistency with HVS. These methods attempted to capture the properties of the HVS in perceptions of luminance,

contrast and structure in the fixed scale of space but did not consider the influence of the distortions in different scales of space.

In the last three decades, scientists have found that neurons in the primary visual cortex are well-modeled localized multiscale bandpass oriented filters (loosely referred to as "wavelets") that decompose images into multiple visual channels [15, 23, 24, 82]. This provides a theoretical basis for the applications of multi-scale and multi-channel analysis in image quality assessment. The second category of methods is based on a prior statistical model in transform domains. The model parameters summarize the image information in an efficient way and this leads to RR-IQA algorithms with low RR data rate. These methods are considered more potential to be extended for general-purpose applications since the statistical and perceptual features being employed are not restricted to any specific distortion process [77]. In [108, 112], Wang et al. proposed a Wavelet-Based RR Method (WBRRM) utilizing a two-parameter General Gaussian Distribution (GGD) to summarize the image information, and the parameters served as RR features with a quite low data rate (around 0.05% of the reference image). This RR-IQA method has a good performance in evaluating single type of distortions but is of low accuracy and poor robustness in detecting mixed distortions. Inspired by [108], Ma et al. proposed a DCT domain method using city-block distance to measure the difference of the coefficients [57]. The algorithm reorganizes the DCT coefficients of an image in an analogous way of wavelet image representation [115, 119], that is to say, there are structural similarities between the sub-bands and the magnitude decays towards the high-frequency sub-bands [57]. This confirms the effectiveness of the multi-resolution of wavelet image representation. Further, Li et al. [53] proposed an improved method for WBRRM, based on a nonlinear Divisive Normalization Transform (DNT) but the improvement is limited (compared with the state-of-the-art FR-IQA). Moreover, the computational complexity increases significantly, due to the use of DNT.

With the development and deepening of wavelet transform research, complex wavelet transform has been proposed and proved to have advantages in many applications of image processing, such as image coding [84], denoising [106], motion estimation [59], quality measures [109], and image retrieval [105]. Several RR-IQA methods based on complex wavelet transform have been proposed and these methods have higher accuracy and better robustness compared with the DWT-based methods. In [54], Lin et al. presented an RR-IQA method based on the Average Direction Information (ADI) in complex wavelet domain and the method has been proved to outperform the DWT-based method WBRRM. The ADI is the mean of the magnitude of the inter-coefficient product which is a decimated pyramid of the wavelet coefficients [9]. Therefore, this RR-IQA method only uses the partial information of the coefficients' phase. Another RR-IQA metric based on the combination of relative

phase (phase difference of adjacent point) and ADI has been proposed [55]. The method makes full use of the phase information of complex wavelet transform and has been proved to have better performance than most other RR-IQA methods. However, another important information, e.g., the magnitude information has not been used. We assume that the method combining magnitude information with phase information will have better performance than the method only with phase information.

The third type of methods integrates features from spatial and transform domains and usually has high accuracy in image quality estimation because of the combination of multiple features. In [77], Rehman et al. proposed the RR-SSIM method based on the successful FR-IQA method SSIM [111]. According to fast wavelet-based image sharpness map and local standard deviation map, Zhang et al. [118] investigated the S4RR based on the FR method. These RR methods have a high accuracy but with a high RR data rate (for example 4.88% of the reference image for S4RR) which will lead to inconvenience for the communication system.

Thus, the previous approaches are often subject to two restrictions: either the distortions they can detect are few, or the accuracy is limited due to a small number of RR features. In this study, we focus on the method based on the features of complex wavelet transform domain to seek an RR-IQA method which can (1) detect various distortions, (2) be competitive with state-of-the-art FR-IQA methods, and (3) summarize the image information with a low data rate and a low computational complexity.

2.1.3 Evaluation of IQA method

2.1.3.1 Evaluation criteria

As more and more IQA algorithms have been proposed in last two decades, evaluation on the pros and cons of these algorithms naturally becomes an important issue. To validate the objective and the subjective methods of IQA and standardize the objective evaluation method, in 1997, the Video Quality Expert Group (VQEG) was established in Turin, Italy, consisting of a group of experts drawn from ITU-T and ITU-R Study Groups [78]. It mainly devotes to advance the field of video/image quality assessment by evaluating new subjective and objective methods for quality assessment, and recommends standard methods to the related standardization organization.

For subjective IQA methods, to obtain accurate and reliable MOS/DMOS, a large psycho-visual tests and a proper experimental methodology are necessary. Thus, the experimental methodology and the number of participated subjects and reference images determine the reliability of the database. The existing database like LIVE release 1/2 and TID2008/2013

based on large psychovisual tests typically serve as the benchmark to evaluate the objective IQA methods.

The performances of objective IQA methods are generally validated by some statistical criteria between the predicted MOS/DMOS and the real MOS/DMOS provided by the database. According to the recommendation of VQEG [44], the specific criteria are employed to evaluate the quality assessment methods. The first one, Pearson linear correlation coefficient (PLCC), is defined as

$$\text{PLCC} = \frac{\sum_{i=1}^{N}(S_i - \mu_S)(O_i - \mu_O)}{\sqrt{\sum_{i=1}^{N}(S_i - \mu_S)^2}\sqrt{\sum_{i=1}^{N}(O_i - \mu_O)^2}}, \tag{2.1}$$

where S_i, O_i are respectively the ith subjective and objective scores, μ_S, μ_O are the corresponding mean value, N is the number of the subjective and objective scores. This PLCC is employed to evaluate the prediction accuracy and a higher PLCC implies a higher prediction accuracy. As shown in Fig. 2.4 (a) and (b), (a) has a higher PLCC value than (b), which indicates the model of (a) has a higher prediciton accuracy than that of (b).

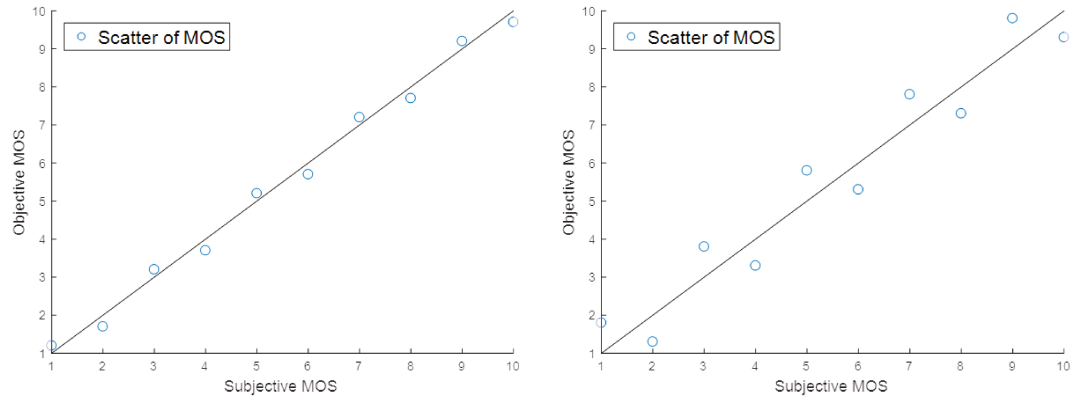


Fig. 2.4 Comparison of scatter plot with higher and lower PLCC: (a) A scatter plot with higher PLCC and (b) A scatter plot with lower PLCC.

The second one, Spearman rank-order correlation coefficient (SROCC), is utilized to assess the prediction monotonicity.

$$\text{SROCC} = 1 - \frac{6\sum_{i=1}^{N} d_i^2}{N(N^2-1)} \tag{2.2}$$

where N is the number of scores, d_i is the difference between the ith image's ranks in subjective and objectives. The higher the SROCC, the higher the prediction monotonicity, as

is illustrated in Fig. 2.5 (a) and (b), where (a) has a higher SROCC than (b) and that means the model of (a) has a higher prediciton monotonity than that of (b).

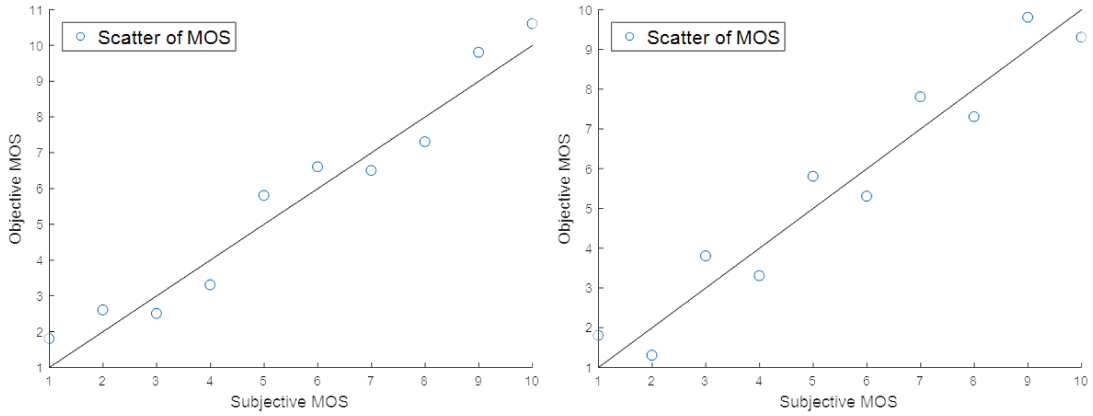


Fig. 2.5 Comparison of scatter plot with higher and lower SROCC value: (a) A scatter plot with higer SROCC and (b) A scatter plot with lower SROCC.

The last one, outlier ratio (OR), is the ratio between outlier-points and total sample points, which is employed to evaluate the prediction consistency.

$$\mathrm{OR} = \frac{N_o}{N} \tag{2.3}$$

where N_o is the total number of outlier points and N is the total number of all points. The outlier-points satisfy:$|E(i)| > 2\sigma_s$, where $E(i)$ is the prediction error of the i-th point, σ_s is the standard deviation of subjective scores. The lower the OR, the higher the prediction consistency, as is illustrated in Fig. 2.6 (a) and (b), where (a) has a lower OR than (b). This means the model of (a) has a higher prediciton consistency than that of (b).

In summary, the higher PLCC and SROCC values denote the higher prediction accuracy and prediction monotonicity, respectively, whereas the higher OR value reflects the lower prediction consistency. Thus, a good IQA method should have higher PLCC, higher SROCC and lower OR. We will evaluate the proposed image quality assessment metric with these criteria in Chapter 3.

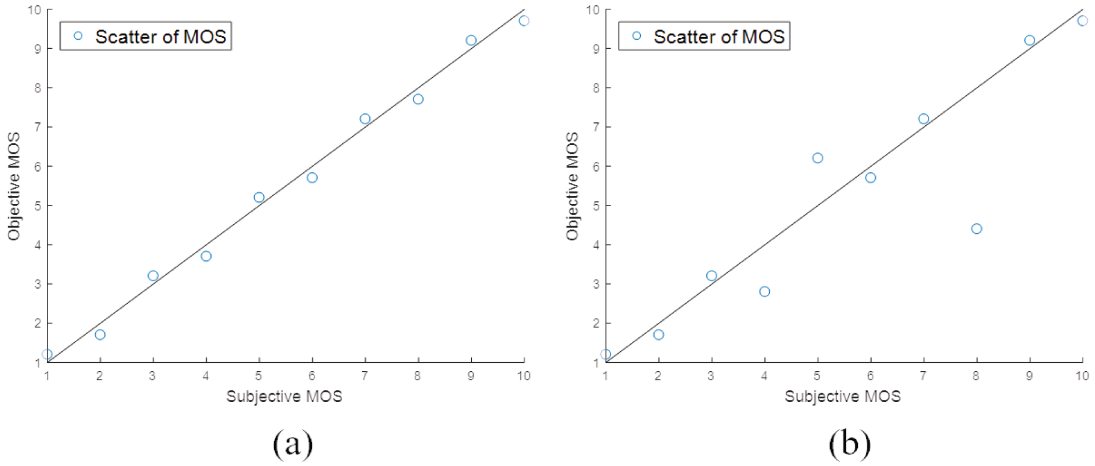


Fig. 2.6 Comparison of scatter plot with higher and lower OR values: (a) A scatter plot with lower OR and (b) A scatter plot with higher OR.

2.1.3.2 Representative IQA metrics

In addition to evaluating the proposed IQA metrics with subjective scores, some representative IQA indicators were used for comparisons, such as MSE, PSNR and SSIM [111].

The MSE metric indicates the power of the distortion, that is, the difference between the reference image and the distortion one. The MSE metric is defined as

$$\mathrm{MSE} = \frac{1}{MN}\sum_{m}^{M}\sum_{n}^{N}\left(I(m,n) - \hat{I}(m,n)\right)^2, \tag{2.4}$$

where I and $\hat{I}$ denote the reference and the distortion images respectively, m and n denote the row and the column of the specific pixel in the image, and M and N are numbers of rows and columns.

Another representative metric is the peak-signal-to-noise ratio (PSNR) which is a converted version of MSE in dB. PSNR is the ratio between the maximum power of a signal and the distortion power, i.e.,

$$\mathrm{PSNR} = 10\log_{10}\left(\frac{Q^2}{\mathrm{MSE}}\right), \tag{2.5}$$

where Q is the quantization maximum of pixel intensity, for example, for an 8 bits quantization $Q = 255$.

MSE and PSNR are the simple yet efficient metrics to detect the difference between the reference and the distortion images. However, in some cases, they cannot measure the image quality in consistent with the HVS. To measure the image quality more precisely, Wang et al. proposed an excellent index, named Structural SIMilarity (SSIM) [111].

The SSIM metric measures the image quality with the difference in luminance, contrast and structural similarity. Given reference image x and its corresponding distortion image y, the SSIM index is defined as

$$\begin{aligned} \mathrm{SSIM}(x,y) &= l(x,y)\cdot c(x,y)\cdot s(x,y) \\ &= \frac{2\mu_x\mu_y}{\mu_x^2+\mu_y^2}\cdot\frac{\sigma_x\sigma_y}{\sigma_x^2+\sigma_y^2}\cdot\frac{\sigma_{xy}}{\sigma_x\sigma_y} \\ &= \frac{(2\mu_x\mu_y+C_1)(\sigma_{xy}+C_2)}{(\mu_x^2+\mu_y^2+C_1)(\sigma_x^2+\sigma_y^2+C_2)}, \end{aligned} \tag{2.6}$$

where $l(x,y)$, $c(x,y)$ and $s(x,y)$ denote the luminance, contrast and structural comparisons, respectively, μ_* represents the mean value, σ_* denotes the variance, σ_{xy} is the covariance between images x and y, and C_1 and C_2 are two constants to avoid the denominator equal to 0. The range of the SSIM index is 0~1, the value 1 denotes that the image has a perfect quality, and the smaller the value, the worse the image quality.

2.1.3.3 Subjective IQA database

To verify the performance of objective IQA metrics, many subjective IQA databases have been established. In this thesis, we mainly introduce five extensively used IQA databases including LIVE 2 [90], CSIQ [50], TID2008 [72], TID2013 [71] and Toyama [100].

The LIVE 2 database comprises 29 reference images and 779 distortion images. The types of distortions are: white noise, Gaussian blur, JPEG compression, JPEG 2000 compression and bit errors in JPEG 2000 bit stream (fast fading channel model). The database utilizes DMOS to evaluate the image quality and the range is scaled to 0~100.

The CSIQ database consists of 30 reference images and 866 distortion images. The database includes six distortions: additive white noise, additive pink Gaussian noise, JPEG compression, JPEG2000 compression, Gaussian blur and global contrast decrements. The subjective score is in the form of DMOS and the range is scaled to 0~1.

The TID2008 database contains 25 reference images and 1700 distortion images. The distortion images generated from the reference images with 17 types of distortions at four distortion levels. One can refer to [72] for the specific distortion types. The database utilizes MOS to assess the image quality and the range is scaled to 0~8.

The TID2013 database is the updated version of TID2008, including 25 reference images and 3000 distortion images. The distortion types are increased to 24 and each type has five distorted levels. The database also adopts MOS to estimate the image quality and the range is scaled to 0~8.

The Toyama database comprises 14 reference images and 196 distortion images. The database has two types of distortions: JPEG compression and JPEG2000 compression. The quality ratings of images are quantified with MOS (range 1~5).

To form a comparison, Table 2.1 intuitively shows the main information of the five databases. We will evaluate the new proposed RR-IQA metric based on these databases in Chapter 3. However, they are not suitable for the classification problem in Chapter 4. The reasons are two-fold: 1) all these databases are about the prediction of subjective scores and this is a regression problem rather than a classification problem, 2) all the distortion images of these databases are obtained in a simulated environment, not in a realistic environment. Thus, it is necessary to construct a new database to provide a classification benchmark for decoding strategy optimization.

Table 2.1 Information of five extensively used image databases

Database	Reference images	Distortion images	Distortion types	Distortion levels	Subjective evaluation	
					Score type	Scale
LIVE 2	29	779	5	5-9	DMOS	0~100
CSIQ	30	866	6	4-5	DMOS	0~1
TID2008	25	1700	17	4	MOS	0~8
TID2013	25	3000	24	5	MOS	0~8
Toyama	14	196	2	6	MOS	1~5

2.2 Wavelet transform

2.2.1 Discrete Wavelet Transform

Since its emergence 30 years ago, the wavelet transform has been extremely successful in a variety of signal processing applications, in the process, often redefining the state-of-the-art performance. [104, 85, 96]. The Discrete Wavelet Transform (DWT) replaces the infinitely oscillating sinusoidal-basis functions of the Fourier transform with a set of locally oscillating basis functions called wavelets [85]. In the classical setting, the wavelet is a stretched and shifted version of a basic real-valued bandpass wavelet $\psi(t)$ [85]. When carefully selected and combined with shifts of a real-valued low-pass scaling function $\phi(t)$, they form an orthonormal basis expansion for one-dimensional (1-D) real-valued continuous-time signals [26, 85]. Thus, given a finite-energy analog signal $x(t)$, it can be represented by wavelets and

scaling functions as

$$x(t) = \sum_{m=-\infty}^{\infty} s(m)\phi(t-m) + \sum_{j=0}^{\infty}\sum_{m=-\infty}^{\infty} w(j,m)2^{j/2}\psi(2^j t - m). \tag{2.7}$$

The wavelet coefficients, $w(j,m)$, and scaling coefficients, $s(m)$, can be computed by the inner products

$$w(j,m) = 2^{j/2}\int_{-\infty}^{\infty} x(t)\psi(2^j t - m)\mathrm{d}t, \tag{2.8}$$

$$s(m) = \int_{-\infty}^{\infty} x(t)\phi(t-m)\mathrm{d}t, \tag{2.9}$$

where j is the scale factor and m is the translation. The DWT provides a time-frequency analysis with wavelet coefficients and scaling coefficients. With well-designed filter banks, the coefficients $w(j,m)$ and $s(m)$ will be efficiently computed with recursive filtering and downsampling operations. For the 2D analysis, the decomposition of two dimensions will form an approximation sub-band and three other sub-bands that are respectively horizontal, vertical and diagonal sub-bands. In the last three decades, scientists have found that neurons in the primary visual cortex are well-modeled localized multiscale bandpass oriented filters (loosely referred to as "wavelets") that decompose images into multiple visual channels [15, 23, 24, 82]. With the localized multiscale bandpass oriented filter banks, the discrete wavelet transform has very great advantages in signal and image processing.

However, the DWT suffers two main drawbacks: Lack of shift invariance and poor direction selectivity [47]. This leads to the constraint within some applications. To overcome the drawbacks of DWT, the Dual-Tree Complex Wavelet Transform has been proposed [46], with advantages in approximate shift invariance, good direction selectivity, limited redundancy and perfect reconstruction [47]. It was widely used in many applications such as denoising, coding and quality assessment. This section mainly introduces the structure and properties of the DT-CWT.

2.2.2 Dual-Tree Complex Wavelet Transform

The DT-CWT proposed by Kingsbury [46, 47] is an enhancement of the DWT. Its filters employed in the two trees are designed in such a way that the aliasing in one branch in the first tree is approximately cancelled by the corresponding branch in the second tree [105]. In [47], Kingsbury focused on designing of a dual-tree filter bank (FB) making filters in tree a satisfying the half-sample phase delay condition with respect to filters in tree b, as shown in Fig. 2.7, where filters $h(n)$ and $g(n)$ have the approximate Hilbert transform relation (90° out of phase from each other). The Hilbert transform defined in the time domain is a

convolution between the Hilbert transformer $1/(\pi t)$ and a function $f(t)$. Given $\widehat{f(t)}$ as the Hilbert transform of $f(t)$, it is defined as

$$\widehat{f(t)} = \frac{1}{\pi t} * f(t) = \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{f(\tau)}{t-\tau} \mathrm{d}\tau, \tag{2.10}$$

where $*$ denotes the convolution and the integral is Cauchy principal value (ignoring the singularity at $\tau = t$ and $\tau = \pm\infty$).

The Hilbert transform will lead to the complex wavelet transform having the same advantages of Fourier transform such as no oscillating, shift invariance and non-aliasing. The scaling function and the wavelet described in Eq. (2.7)-(2.9) can be represented as

$$\psi_c(t) = \psi_r(t) + \mathrm{j}\psi_i(t), \tag{2.11}$$

$$\phi_c(t) = \phi_r(t) + \mathrm{j}\phi_i(t), \tag{2.12}$$

where $\mathrm{j} = \sqrt{-1}$, $\psi_r(t)$ and $\phi_r(t)$ are the real components, and $\psi_i(t)$ and $\phi_i(t)$ are the imaginary components. Moreover, $\psi_r(t)$ and $\psi_i(t)$, $\phi_r(t)$ and $\phi_i(t)$, respectively, are Hilbert transform pairs, while $\psi_c(t)$ and $\phi_c(t)$ are analytic signals and supported on positive of the frequency axis [85]. The analytic signal at different scales could be a good strategy to model the degradation in consistent with the HVS since scientists have found that neurons in the primary visual cortex are well-modeled localized multiscale bandpass oriented filters [15, 23, 24, 82].

Then the complex wavelet coefficients can be computed by

$$w_c(j,m) = w_r(j,m) + \mathrm{j}w_i(j,m). \tag{2.13}$$

Thus, the magnitude and the phase are respectively

$$|w_c(j,m)| = \sqrt{[w_r(j,m)]^2 + [w_i(j,m)]^2}, \tag{2.14}$$

$$\angle w_c(j,m) = \arctan \frac{w_i(j,m)}{w_r(j,m)}, \tag{2.15}$$

when $w_r(j,m) > 0$.

With the structure of the dual-tree FB, the DT-CWT has advantages of less aliasing for sub-band, better shift invariance and better direction selectivity than the DWT. Directional selectivity is achieved in the 2-D case by combining the outputs of the FB in such a way that the equivalent complex filters have supports in only one quadrant of the frequency plane [46].

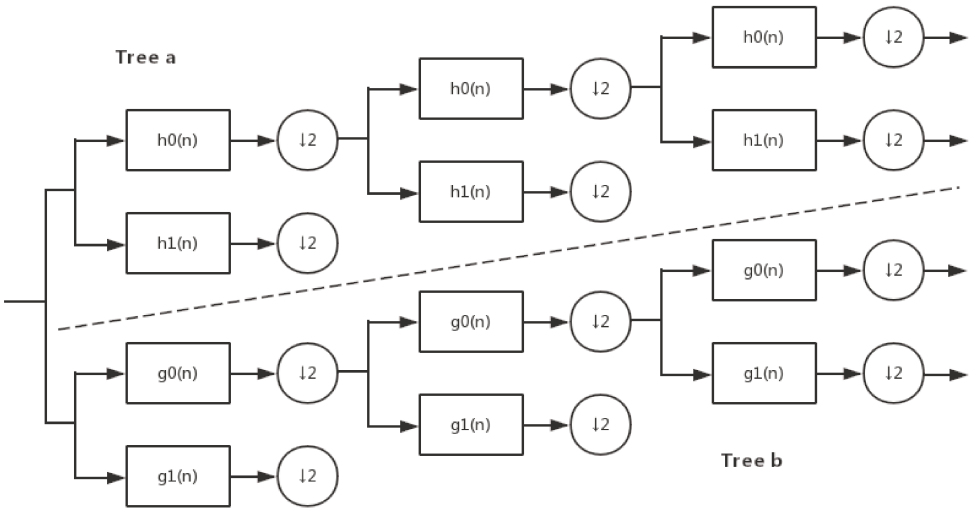


Fig. 2.7 Analysis filter bank of the DT-CWT.

The wavelet functions from two trees play the role of the real and imaginary parts of a complex analytic wavelet. In this way, the 2-D DT-CWT has six directions at approximately $-15°$,$-45°$,$-75°$,$+75°$,$+45°$,$+15°$, as shown in Fig. 2.8.

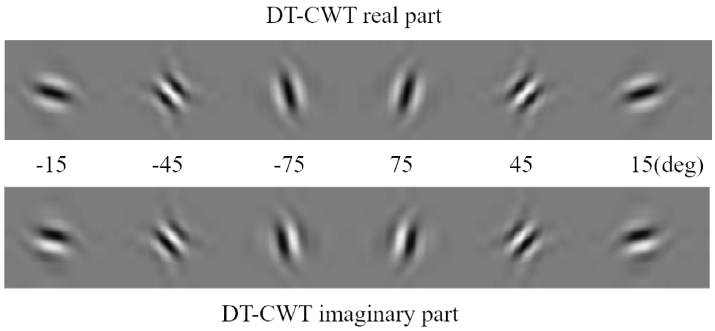


Fig. 2.8 Six orientation sub-bands of DT-CWT with real and imaginary parts.

Given a mother wavelet $\psi(\cdot)$, a scale wavelet $\phi(\cdot)$, and x, y the two dimensions. The 2-D wavelet can be represented as $\psi(x,y) = \psi(x)\psi(y)$, where $\psi(x)$ is a complex wavelet, i.e., $\psi(x) = \psi_h(x) + \mathrm{j}\psi_g(x)$. Then, the expression of $\psi(x,y)$ can be computed as

$$\begin{aligned}\psi(x,y) = \psi(x)\psi(y) &= [\psi_h(x) + \mathrm{j}\psi_g(x)][\psi_h(y) + \mathrm{j}\psi_g(y)] \\ &= [\psi_h(x)\psi_h(y) - \psi_g(x)\psi_g(y)] + \mathrm{j}[\psi_g(x)\psi_h(y) + \psi_h(x)\psi_g(y)].\end{aligned} \tag{2.16}$$

Considering the support of the Fourier spectrum and the effect of Hilbert transform, we can judge that this wavelet is orientated at -45°.

The other directional wavelets have the similar expressions, therefore six real wavelets can be formulated as [85]

$$\psi_i(x,y) = \frac{1}{\sqrt{2}}(\psi_{1,i}(x,y) - \psi_{2,i}(x,y)), \tag{2.17}$$

$$\psi_{i+3}(x,y) = \frac{1}{\sqrt{2}}(\psi_{1,i}(x,y) + \psi_{2,i}(x,y)), \tag{2.18}$$

for $i = 1,2,3$, where the two separable 2-D wavelet bases are defined in a usual manner [85]:

$$\psi_{1,1}(x,y) = \phi_h(x)\psi_h(y), \ \ \psi_{2,1}(x,y) = \phi_g(x)\psi_g(y), \tag{2.19}$$

$$\psi_{1,2}(x,y) = \psi_h(x)\phi_h(y), \ \ \psi_{2,2}(x,y) = \psi_g(x)\phi_g(y), \tag{2.20}$$

$$\psi_{1,3}(x,y) = \psi_h(x)\psi_h(y), \ \ \psi_{2,3}(x,y) = \psi_g(x)\psi_g(y). \tag{2.21}$$

The normalization operator $1/\sqrt{2}$ is employed to constitute an orthonormal operation for sum and difference. The imaginary wavelet can be obtained in the same way. Thus, with the real and imaginary wavelets, we can get complex wavelets of DT-CWT as well as magnitude and phase information of each wavelet.

With six 2-D analytic wavelet sub-bands, for one scale, the DT-CWT has six pairwise magnitude and phase information as features. They are widely applied in signal and image processing, such as image denoising, compression, segmentation, texture retrieval and quality evaluation. For 2-D image processing, magnitude information indicates the extent of discontinuities, while phase information provides the locations of features [105]. The combination of magnitude and phase in complex wavelet domain can represent the image features better than the single magnitude in discrete wavelet domain. Thus, in this thesis, we will use features both from magnitude and phase of DT-CWT to evaluate image quality.

2.2.2.1 Magnitude

The magnitude of DT-CWT wavelet is a crucial feature of the wavelet. The magnitude information indicates the discontinuities and the mutability of the signal. As shown in Fig. 2.9 (a), the edges and textures of the image with the discontinuities and mutability can be seen obviously.

For classic DWT, the real wavelet exhibits a certain statistical law that the marginal distribution of the coefficients in individual wavelet sub-bands satisfies a generalized Gaussian

distribution [108, 112]. The property has been widely used in image processing, including texture retrieval, segmentation and quality assessment [53, 77, 105, 106]. Although the shape of the histogram is different from that of the DWT, the DT-CWT sub-band also has a fixed distribution, as shown in Fig. 2.9 (b). For example, Shaffery et al. modelled the magnitude distribution with a Mixture Rayleigh Distribution (MRD) [87], as shown in Fig. 2.9 (c), and Rakvongthai et al. modelled the magnitude distribution with a multivariate generalized Gaussian distribution [75]. Thus, we believe that the statistical feature of DT-CWT magnitude can be applied in image quality assessment by selecting an appropriate model and a proper strategy to reduce the model error.

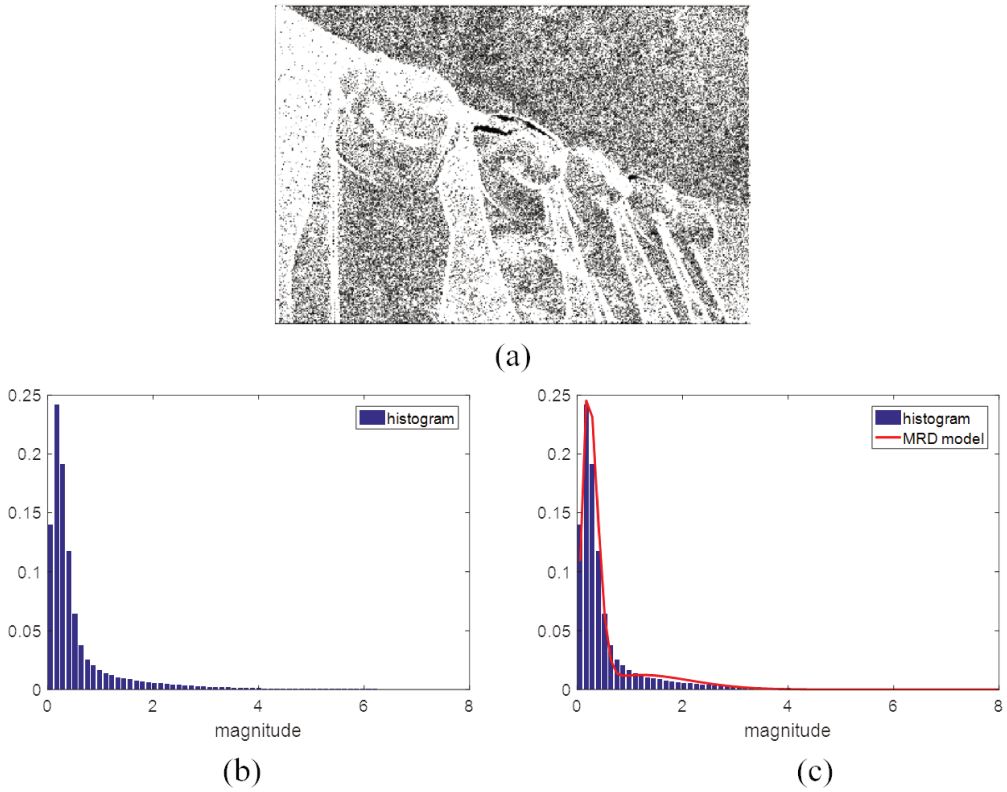


Fig. 2.9 The magnitude of DT-CWT coefficients and its histogram: (a) The magnitude of image "caps", (b) The histogram, and (c) The approximation of histogram.

2.2.2.2 Phase and relative phase

The phase of DT-CWT coefficient is another important feature, which can provide structural information and has been used in many image applications including edge detection [79], texture retrieval [105], segmentation [82] and pattern regcognition [38]. Furthermore, phase contains more structural information than magnitude in typical natural images [69], and a

rigid translation of image structures leads to a consistent phase shift [82]. Fig. 2.10. illustrates the phase of DT-CWT coefficients can provide valuable information of image structure, for example, the edge or the ridge information. However, it is difficult to obtain the statistical property of phase directly since the Probability Density Function (PDF) of phase in some decomposition scales is unregular [74]. The phase histogram tends to be unregular when the decomposition scale larger than two, as illustrated in Fig. 2.10 (c). To make full use of the features of phase, An [105] advised to extract the statistical property from relative phase instead of from phase directly.

Definition 1 The relative phase is a phase difference of two adjacent complex wavelet coefficients. For a spatial location (i, j) within a particular complex sub-band, the relative phase is expressed as [105]

$$\theta(i,j) = \angle y(i,j) - \angle y(i,j+1) \text{ or } \theta(i,j) = \angle y(i,j) - \angle y(i+1,j), \tag{2.22}$$

where $\angle y(i,j)$ is a phase value at position (i,j).

Property 1 The featured orientation α of an edge in the supported region of a sub-band is linearly proportional to the relative phase of two adjacent complex wavelet coefficients, which are located in the vicinity of the edge as [105]

$$\alpha(i,j) \approx a + b\theta(i,j), \tag{2.23}$$

where a and b are constant and can be estimated for each sub-band.

With this property, the relative phase can represent the orientation feature very well. An [105] applied the relative phase in texture retrieval and segmentation. Since some degradations disturb the structure of the image information or the orientation of the details, the relative phase can measure these degradations. In this thesis, the relative phase as a feature is employed to evaluate image quality.

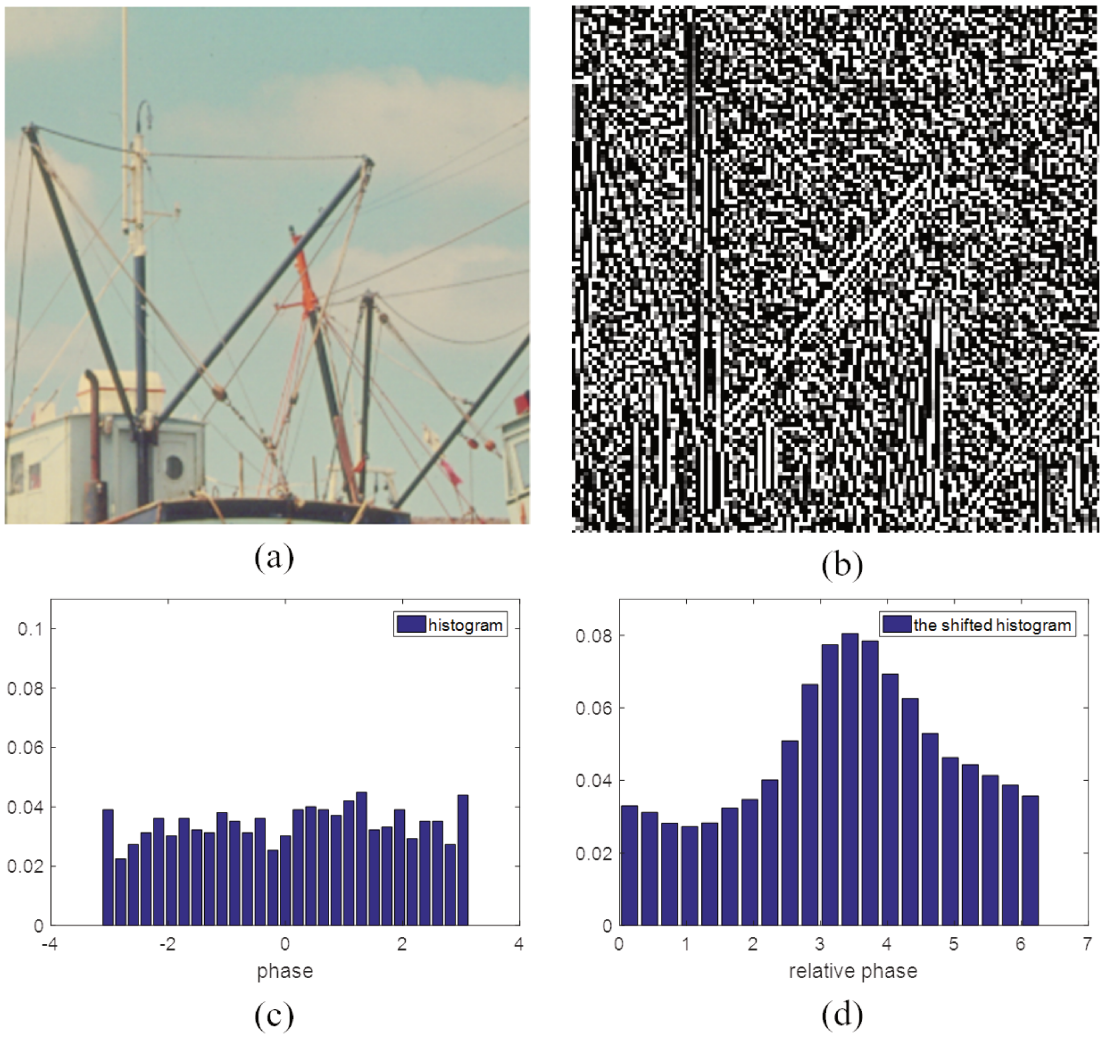


Fig. 2.10 The phase of DT-CWT coefficients and its histogram: (a) Section of image "Boats", (b) Phase of the DT-CWT coefficients, (c) The corresponding histogram, and (d) The histogram of relative phase (π shifted histogram).

2.3 Wireless channel

Over the last decades, with the development of communication and information technologies, mobile communication and applications have become more important and popular. The users' demand on high-quality and high-definition image or video constantly increases with the development of mobile applications. However, it conflicts with the well-known constraints in wireless transmission such as multipath frequency selective fading due to multipath transmission, Doppler effect because of the mobility of the users, and limited bandwidth. In this section, we will introduce the properties of the wireless channel and the corresponding technologies used to improve the transmission efficiency.

2.3.1 Channel properties

The major property of the wireless channel is that the channel strength varies with time and frequency. The variation can be divided into three classes:

1. Path loss, also known as attenuation, referring to the loss caused by the propagation of electromagnetic waves in space.

2. Large-scale fading, caused by shadows of large objects such as buildings and hills.

3. Small-scale fading, mainly produced by multipath propagation. The multipath propagation is a mixed propagation of different paths such as direct, reflection, diffraction and dispersion paths, as shown in Fig. 2.11.

The large-scale fadings and the small-fadings can be treated with the strategies such as communication link design and cell-site planning. The fast fadings consist of time, frequency, and space selective fadings. These fadings may form inter-symbol interference and lead to degradation of communication system performance. How to overcome these fadings and design a reliable and effective communication system has become an important research focus of communication community.

Another important property of the wireless channel is the mobility of the terminal users, which leads to the change in the frequency or wavelength of the electromagnetic wave relative to the mobile users, namely the Doppler effect. The Doppler effect and fast fading effect cause the time-varying characteristics of wireless channels. In transmission, the existence of time-varying characteristics means time delay for the received signal in different paths, and this may cause one waveform symbol to extend to another waveform symbol, resulting in an Inter-Symbol Interference (ISI).

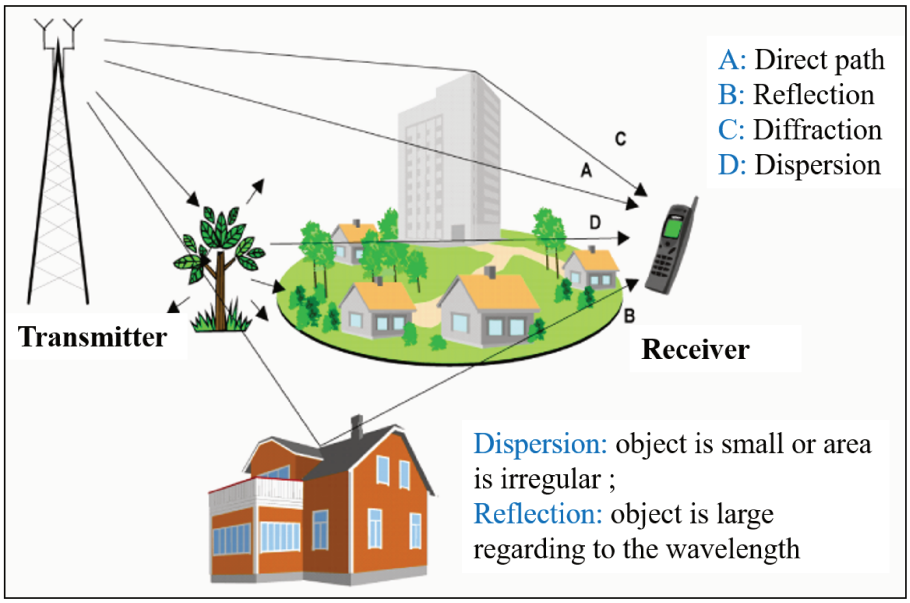


Fig. 2.11 Multipath propagation in wireless channels.

To avoid ISI, the symbol width is required to be much larger than the maximal delay spread of the channel. However, for the wireless communication system, the width of the single information symbol is much lower than the delay spread. Consequently, ISI will be a serious problem and lead to a high error rate, if the corresponding techniques have not been employed. Next, some crucial technologies will be introduced, which are exploited to handle the relevant problems caused by channel properties.

2.3.2 Stochastic models

To date, there have been many types of radio channel models [10, 70, 93, 114]. For example, based on the bandwidth, the channel models can be classified into wideband and narrowband models. Further, the narrowband models comprise path loss, large-scale fading and small-scale fading models, while the wideband models include power delay profile and tapped delay line models, and models for the Arrival Times of Rays and Clusters [102, 114]. According to the modeling methods, the channel models can be divided into empirical, stochastic, deterministic and semi-deterministic models. Since the thesis mainly focuses on the image quality optimization of a wireless image communication system, here, we only introduce stochastic models.

The stochastic model indicates the uncertainty of the channel fading but its statistical properties follow a certain law, i.e., the PDF of the signal transmitted in channels can be estimated. A stochastic model is a tool to estimate probability distributions of potential

outcomes by allowing random variation in one or more inputs over time. Generally, a stochastic model utilizes one or more variables to estimate the potential responses of the radio channel. It is usually suitable to model all fadings like the small-scale and large-scale fadings. In the wireless local area networks, the small-scale fadings of the Non-Line-Of-Sight (NLOS) and Line-Of-Sight (LOS) transmission have been demonstrated to follow Rayleigh distribution and Rice distribution, respectively [93]. While the large-scale fading follows log-normal distributions [10, 114].

2.3.2.1 Rayleigh fading model

For the NLOS transmission path, the Rayleigh fading model has generally been exploited to model multipath fading. For the NLOS case, given a received signal amplitude μ_l, it satisfies the Rayleigh distribution at any time [22]:

$$P_l(\mu_l) = \frac{\mu_l}{\sigma_l^2} \exp\left(-\frac{\mu_l}{2\sigma_l^2}\right), \quad 0 \leq \mu_l < \infty, \tag{2.24}$$

where σ_l is the scale parameter.

The Rayleigh distribution is usually employed to simulate the degradation of a signal within transmission channels, for example, in LIVE 2 database, the JPEG2000 transmission error distortion is simulated by a Rayleigh distribution.

2.3.2.2 Rice fading model

For the direct LOS transmission path, the envelope of the channel complies with the Rice fading model:

$$P_c(\mu_c) = \frac{\mu_c}{\sigma_c^2} \exp\left(-\frac{\mu_c^2 + \rho^2}{2\sigma_c^2}\right) I_0\left(\frac{\mu_c \rho}{\sigma_c^2}\right), \quad 0 \leq \mu_l < \infty, \tag{2.25}$$

where μ_c is the magnitude of envelope, σ_c is the scale parameter, ρ is the maximal value of the main component (LOS component) of signal, and $I_0(\cdot)$ is the Bessel function of the first kind.

The Rice factor is defined as $K = \rho^2/(2\sigma^2)$, i.e., the ratio of power of the LOS component to the variance of all NLOS components. If $\rho \to 0$ and $K \to 0$, the Rice distribution gradually approaches the Rayleigh distribution. Thus, the Rayleigh distribution can be regarded as the Rice distribution without LOS component.

2.3.2.3 Nakagami-m fading model

The Nakagami-*m* fading model is defined as [117]

$$P_n(\mu_n) = \frac{m^m \mu_n^{2m-1}}{\Omega_n^m \Gamma(m)} \exp\left(-\frac{m\mu_n^2}{\Omega_n}\right), \quad 0 \leq \mu_n < \infty, \tag{2.26}$$

where μ_n is the magnitude of the received signal, $\Omega_n = \overline{\mu_n^2}$ is the average power of fading, μ_n is the scale parameter of Nakagami-*m* fading, and $\Gamma(\cdot)$ is the Gamma function. The parameter $m \in [1/2, \infty)$ represents the extent of the fading. A smaller m indicates a severer fading and vice versa. The Nakagami-*m* distribution will reduce to the Rayleigh and the one-side Gaussian distributions when m respectively is 1 and 2. If $m \to \infty$, the Nakagami-m fading channel approximates a nonfading addative Gaussian noise channel[70, 114]. If $m > 1$, the distribution can be mapped to a Rice distribution with their parmaters:

$$m = \frac{(1+K)^2}{1+2K}, \quad 0 \leq K \leq \infty, \tag{2.27}$$

where K is the Rice factor. Therefore, the Nakagami-*m* distribution can be employed to model an extensive range of multipath fading.

2.3.2.4 Log-normal fading model

It has been revealed the large-scale fading can be approximated by the Log-normal fading model [88, 70]. Considering the mean power of the received signal is measured in dBW $\nu_d = 10 \log \mathrm{E}(\mu^2)$, the distribution can be defined as

$$P_{\nu_d}(\nu_d) = \frac{1}{\sqrt{2\pi}\sigma_{\nu_d}} \exp\left[-\frac{(\nu_d - \mu_{\nu_d})^2}{2\sigma_{\nu_d}^2}\right], \tag{2.28}$$

where μ_{ν_d} is the mean of ν_d and σ_{ν_d} is the standard deviation, correspondingly. ν_d can be derived from the mean path loss.

With the channel model, we can know the degradation condition of the transmitted signal. There are also some technologies to improve the transmission efficiency and cope with the degradation.

2.3.3 Deterministic models

The deterministic models simulate the physical propagation process such as reflection, diffraction and refraction. They are usually employed to generate both the wideband and the

narrowband signals and have high precision. The deterministic models can be classified into ray-optical models and finite difference time domain models. Here, we mainly introduce the ray-optical models consisting of the ray-launching model and the ray-tracing model.

2.3.3.1 Ray-launching model

Based on the geometrical optics, the ray-launching model simulates the radio wave propagation by direct (LOS), reflected, transmitted, diffracted, scattered, and some combined rays [51, 114]. The transmitter (Tx) launches rays with small angles, and thus ensures that the rays are transmitted in different directions. Then, the obstacles encountered in different propagation directions lead to different transmission ways such as reflection, diffraction and refraction, as shown in Fig. 2.12. The ray propagation will be terminated when the receiver (Rx) is reached or when the ray power attenuates to a particular threshold. The inherent drawback of this model is that the Rx far away from the Tx may be missed even if the interval angle between the rays is very small, as illustrated in Fig. 2.13, where the Rx2 far away from the Tx is missed. Ray launching is often referred to as the brute force method because of its large number of rays launching at very similar angles [8, 41].

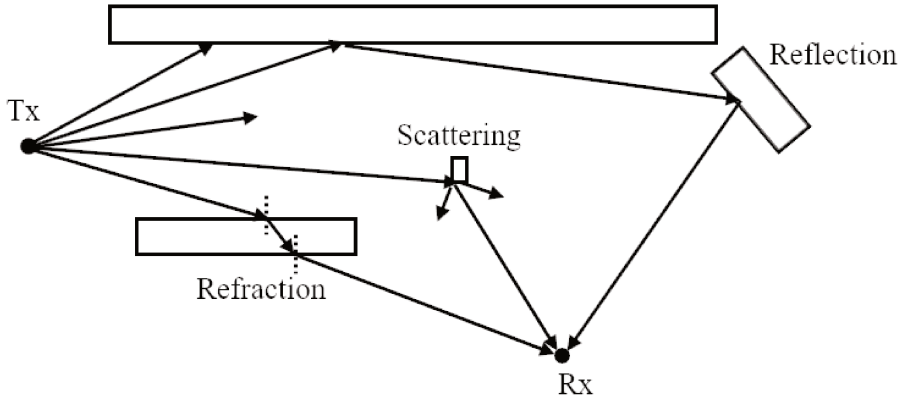


Fig. 2.12 Illustration of ray-launching model.

2.3.3.2 Ray-tracing model

The ray-tracing model, also named image method, traces rays backwards unlike the ray-launching method that traces rays forwards [114]. The basic idea of the ray-tracing model is to treat all objects in the transmission environments as reflectors. Then, according to the principle of the geometrical optics to determine the location of the image of Tx, as illustrated in Fig. 2.14. The ray-tracing model is a simple yet accurate method to determine the trajectory between Tx and Rx [41]. For the ray-launching model, the Tx has to launch a

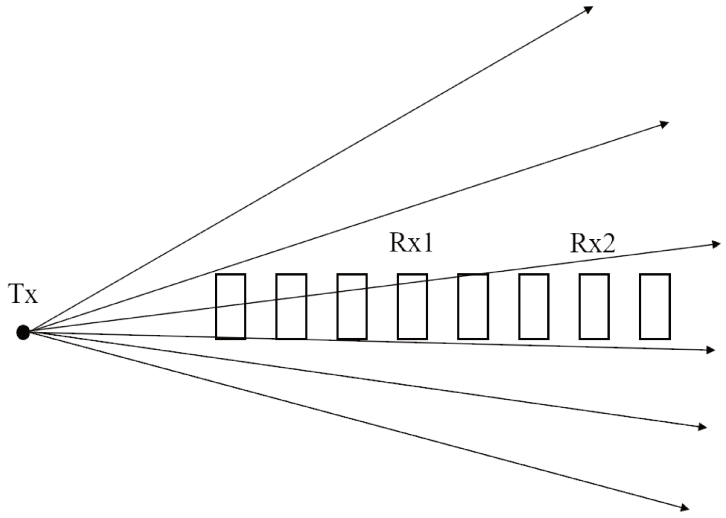


Fig. 2.13 Rx far away from Tx may be missed by rays.

large number of rays but just few of them will reach the Rx and this results in power waste. Moreover, it has to test the intersection of a ray with all objects which is time-consuming and inefficient [41]. For the ray-tracing model, it only considers the paths actual existing between Tx and Rx and hence requires much less computation time than that of ray-launching. However, its drawback is the computation time will increase exponentially with the reflection times. By combining the advantages of two models, a hybrid model was proposed [99] via utilizing the ray-launching model to identify the possible ray trajectory and the ray-tracing model to calculate the reflection positions.

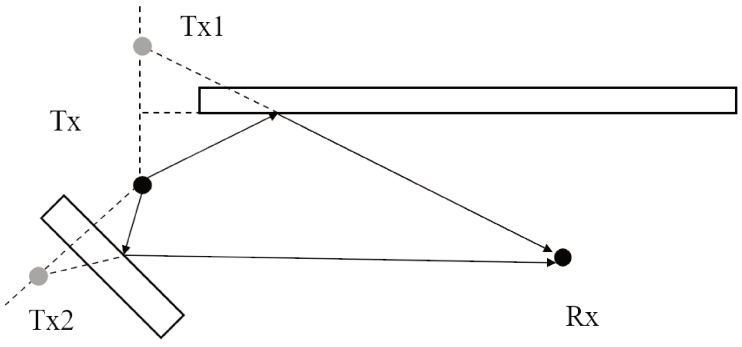


Fig. 2.14 Illustration of ray-tracing model.

2.3.4 Adaptive strategies to guarantee QoS

Over the last two decades, the rapid development of multimedia devices such as digital cameras, computers and mobile smart-phones as well as the growth of networking facilities has enabled a wide application of image and video communication. In turn, the popularity of high-resolution multimedia devices are demanding higher-resolution images and better QoS. However, it conflicts with the well-known constraints in wireless transmission system introduced before. This brings the research boom of communication technologies including kinds of methods to exploit the diversity in time, space and frequency domains.

The diversity in time domain can be exploited to conduct the redundant coding, for example, Error-Correcting Code (ECC). ECC efficacy has been verified for a long duration [58] and is involved in numerous telecommunication standards. The Orthogonal Frequency Division Multiplexing (OFDM) modulation employs frequency diversity to overcome the multipath phenomenon [48, 113]. Based on the multiple antennas at both transmitter and receiver, the Multiple-Input-Multiple-Output (MIMO) technology efficiently exploits the spatial diversity [2, 3, 80]. The combination of MIMO with OFDM (MIMO-OFDM) has been demonstrated to improve the robustness greatly [97]. The MIMO-OFDM has been adopted in serval telecommunication standards such as Wi-MAX IEEE802.11e [40], Wifi IEEE802.11n [39] and Long-Term Evolution (LTE) [86].

For the image transmission in wireless channels, the key point is to ensure the QoS while improving the QoE. To ensure the image quality service, the international standards of image compression and transmission, e.g., JPEG 2000 [43] and JPWL [42] have specified, in an extensible way, a set of methods and syntaxes for error protection and signaling such as unequal error protection, forward error correction code, data partitioning and interleaving [32].

Recently, to adjust the system parameters adaptively, the quality optimization strategies of transmission systems have become sophisticated. An efficient way is to exploite the hierarchy of image compression to protect more important data from distortion. As recommended by the JPWL standard, the Unequal Error Protection (UEP) scheme was promptly applied in wireless communication systems [4, 60] and experimental results demonstrated that these schemes have a good improvement in QoS, compared with the equal error protection.

In a similar way, the Unequal Power Allocation (UPA) scheme has been proposed [11], which divided the JPEG 2000 codestream into hierarchical packages transmitted with different rates and powers. In contrast with the equal power allocation scheme, the scheme obtained about 4 dB PSNR gain at low SNR of Rayleigh and an additive Gaussian white noise fading channel. The advantages of MIMO technology and OFDM modulation technology can be well used in UPA scheme. Sarbi proposed a UPA scheme for JPEG transmission over

MIMO system [81], in which the images are classified into different-importance codestreams. The codestreams are simultaneously transmitted by different antennas of different powers with spatial multiplexing. The main idea is to reduce the distortions of more important data at the cost of more distortions for less important data, thus reducing the overall distortions [81]. The scheme has a PSNR gain of 14 dB under low SNR condition, as contrasted with the equal power allocation scheme.

The Closed Loop (CL) MIMO system, based on Channel State Information (CSI), provides a possibility for adaptive parameter adjustment. A series of CL-MIMO-OFDM schemes with precoder designs have been proposed for the JPWL transmission over a realistic MIMO channel [2, 3]. These adaptive schemes, by considering channel condition, adjust the precoding coefficients and other parameters such as channel SNR, power allocation and channel coding, leading to an improvement of robustness and QoS. However, these schemes only take into account the specific system parameters to optimize transmission and improve QoS, but does not incorporate the QoE of the users. QoE based on the perception of the HVS can indicate the image quality more accurately than QoS. Thus, we propose to apply the RR-IQA quality metric to improve the QoE for JPWL transmission. In Chapter 3, we will propose an RR-IQA image quality metric and then apply it to improve the QoE of a wireless transmission system, in Chapter 4.

2.4 Chapter summary

In this chapter, we first summarise the advantages and disadvantages of subjective and objective IQA, and the validation methods. It is explained why we are interested in the RR-IQA method. Then, we describe the wavelet transform and the extension to complex wavelet transform, which is the foundation of statistical the RR-IQA methods. Finally, we introduce the properties, indicators and relevant technologies of the wireless channel, preparing a basis for the application of RR-IQA in wireless communication systems. The relevant content introduced in this chapter provides theoretical support for the subsequent chapters and also paves the way for the content-extension of the thesis.

Chapter 3

Proposed metric for image quality assessment

As introduced in Chapter 2, we focus on designing a statistical feature-based RR-IQA metric in the complex wavelet domain to evaluate the image quality. The RR-IQA method makes a good trade-off between the amount of reference information and the prediction accuracy, in comparison with FR-IQA and NR-IQA methods. The combination of magnitude and phase features in the complex wavelet domain is suitable to detect multiple distortions frequently encountered in communication systems. Thus, in this chapter, we propose a new RR-IQA metric based on the statistical model in the dual-tree complex wavelet transform domain.

3.1 Motivation

Over the last few decades, the DWT has been used as one of the important decomposition tools in signal and image processing. However, in recent years, it has been demonstrated that the Complex Wavelet Transform (CWT) has a better performance than DWT, especially in image processing [47, 85, 105]. In this section, we dedicate to devising an RR metric based on statistical models in the CWT domain. The main advantages of CWT over DWT are its shift invariant property and good directional selectivity [85]. The CWT is suitable for many applications, such as image coding [84], image denoising [106], motion estimation [59], quality measures [109], and image retrieval [105]. Some approaches to deploying CWT have been proposed in the literature, for example Dual-Tree Complex Wavelet Transform (DT-CWT) [46, 47], Pyramidal dual-tree directional filter banks [68], and uniform discrete curvelet transform [67]. We choose the DT-CWT as a decomposition tool, which is a very efficient numerical algorithm. Hereafter, we briefly introduce the DT-CWT.

Although several RR-IQA methods based on DT-CWT have been proposed, these methods only used partial information, for example, the relative phase information or average directional information i.e., mean magnitude of the inter-coefficient product. The comprehensive method based on the fusion of the magnitude and the phase information of the DT-CWT coefficients has not been proposed yet. It is mainly because the uniform distribution of phase prevent us from extracting valuable features, and the magnitude distribution is difficult to model with a low error. In [105] , it is indicated that the relative phase can be modelled and applied to texture image segmentation and retrieval. The relative phase is a phase difference with a clearer statistical characteristic than the phase itself. Inspired by [105], two RR-IQA methods based on the relative phase [55] and the average directional information [54], respectively, were proposed. However, another important information, the magnitude of the DT-CWT was not considered in these methods. In our method, the magnitude of complex wavelet coefficients will be used as a complement to the relative phase. Our conjecture is because the magnitude is sensitive to some distortions (like white noise, Gaussian blur and JPEG compression distortion) while the relative phase is sensitive to other distortions, such as fast-fading distortion and JPEG 2000 compression distortion. Moreover, to reduce the modelling error, we adopt the Information criterion (IC) to obtain a better model of the information. The IC is used to build histogram to select the best PDF.

Therefore, the main contribution of the thesis is the propositions of an RR-IQA metric based on the magnitude and the relative phase of DT-CWT coefficients, and the IC permits one to propose a strategy to estimate the Kullback-Leibler Divergence (KLD). For this, we first analyze how the magnitude and the relative phase information change with image degradations. Secondly, we design a strategy based on the IC to optimally approximate the distribution of the information. The strategy helps to reduce the modelling error and also can be used in other RR-IQA metrics based on statistical models. Thirdly, we compare the candidate models and select the appropriate models for the magnitude and the relative phase. The models also can be applied to other image applications including texture segmentation, image retrieval, and pattern recognition.

3.2 Framework of the RR-IQA metric FMRP

The new RR-IQA method is based on the statistical model in the complex wavelet domain. We name it as Feature of Magnitude and Relative Phase (FMRP) since it extracts features from both the magnitude and the relative phase of the complex wavelet coefficients.

As illustrated in Fig. 3.1, the reference image I is first decomposed by the DT-CWT. Secondly, the magnitude and the relative phase for each scale are modelled, and the IC is

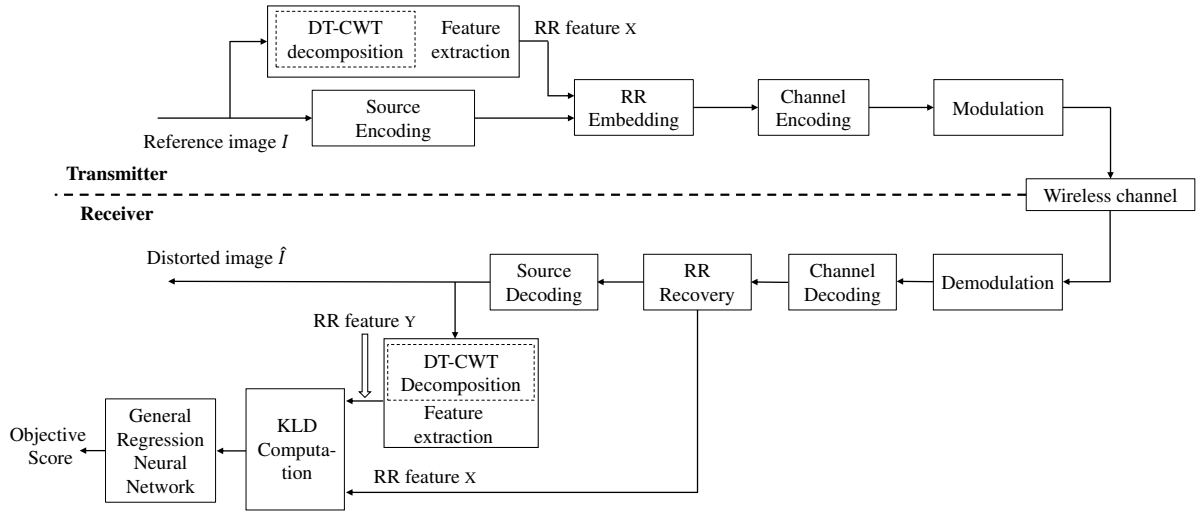


Fig. 3.1 Framework of the proposed RR-IQA method for a wireless system.

employed to reduce the error between the model and the distribution. The parameters of the PDF of the distribution serve as the crucial feature, i.e., RR feature X. Next, the image I and the RR features X will be transmitted to the receiver via a wireless transmission channel and an auxiliary channel, respectively. Generally, the transmission channel introduces distortions and the auxiliary channel is assumed to be distortion-free (for this, we can use error-correcting code and low order modulation). At the receiver, the feature Y is extracted from the received image $\hat{I}$ with the same processing as used in the reference image. Finally, the features X and Y are compared to quantify the distortion and a Generalized Regression Neural Network (GRNN) is utilized to map the difference of two RR features into the objective score.

In the next sections, we will demonstrate how the magnitude and the relative phase allow one to detect image degradation, introduce the IC to build the optimal histogram of the DT-CWT coefficients, and detail each process of the FMRP metric. Moreover, we will give some examples to demonstrate some phenomena and processes. All the test images, as shown in Fig. 3.2, are from the LIVE database 2.



Fig. 3.2 Test images from LIVE database 2.

3.3 The crucial features

As introduced in Section 3.1, our proposition is based on the DT-CWT. We decompose the reference image or the distorted one into 18 sub-bands by a three-scale and six-orientation DT-CWT. The reason why selecting three scales is that the influences of the degradation mainly lie on these scales. The KLDs between the distributions of the sub-bands of the distorted image and the reference one are used as RR feature. In order to describe the RR features more clearly, we review the KLD here.

Definition 2 The KLD is the Kullback-Leibler Divergence between the sub-band coefficients of the distorted image and the reference one. Let $\mathbf{x} = \{x_1, ..., x_N\}$ be a set of N selected coefficients in the DT-CWT domain. Let $P(x)$ and $\hat{P}(x)$ denote the PDF of the DT-CWT coefficients in the same sub-band of the reference image and the distorted one, respectively. The log-likelihoods of x being calculated via $P(x)$ and $\hat{P}(x)$ are

$$l(P) = \frac{1}{N}\sum_{n=1}^{N} \log P(x_n) \text{ and } l(\hat{P}) = \frac{1}{N}\sum_{n=1}^{N} \log \hat{P}(x_n) \tag{3.1}$$

respectively. Based on the law of large numbers, if N is large enough, the difference of the log-likelihood-ratio between $P(x)$ and $\hat{P}(x)$ asymptotically approaches the KLD between

$P(x)$ and $\hat{P}(x)$

$$d(P\|\hat{P}) = \int P(x) \log \frac{P(x)}{\hat{P}(x)} \mathrm{d}x \tag{3.2}$$

The relation between KLD and log-likelihood function has been found and was used to compare images in the applications of image classification, image retrieval and image evaluation [27, 31, 112]. As described above, the KLD between the sub-band coefficients of the distorted image and the reference one is computed as the basis of objective score.

The magnitudes and the relative phases are the extraction sources of RR features. The way is to fit the PDFs with the parameterized models and the parameters serve as the RR features. However, the magnitude has some near-zero values, after the filtering operation of filter banks in DT-CWT. These near-zero coefficients are caused by the uniform area and may lead to difficulty and error in modelling PDF (because the numerical estimation of the phase value is inconsistent). It occurs frequently for natural scenery images. These images often contain some scenes with the same pixel intensity such as the sky, grass and sea. As shown in Fig. 3.3, a natural scene image in (a) has uniform areas such as sky and grass ((b) and (c) are respectively their histograms of the magnitude and the relative phase of DT-CWT). From the histograms, it can be seen an abrupt change located at the near-zero value of the horizontal axis, which is caused by the uniform area (sky and grass) in the original image. Considering that this abrupt change may lead to difficulty in modelling, we should remove these near-zero values.

Therefore, a thresholding process is designed to eliminate these coefficients in magnitude and relative phase. The purpose of the thresholding is to remove the near-zero coefficients while keeping the image information as much as possible. Therefore, we studied the coefficients of all the images in the LIVE database 2 and found that the threshold increases with the decomposition scales. For example, for the 1-scale, 2-scale and 3-scale decompositions, the thresholds are respectively around 10^{-6}, 10^{-4} and 10^{-2} since the DC gain factor of low filter bank is larger than one. The process to determine the threshold is as follows:

1. Dividing the images into two groups: images with and without uniform areas.
2. Performing the DT-CWT and sorting all the elements of the coefficients.
3. After sorting, the near-zero coefficients will cluster together and it is easy to determine the threshold.

For the uniform areas, the relative phase values of the DT-CWT coefficients are equal to zeros. Thus, it is easy to select zero as the threshold of the relative phase. We just need to remove the zeros.

After the thresholding operation, the histograms of the magnitude and the relative phase are easy to model by a specific distribution, as shown in Figs. 3.3 (e) and (f). Next, the modelling of the magnitude and the relative phase can be described with a specific PDF.

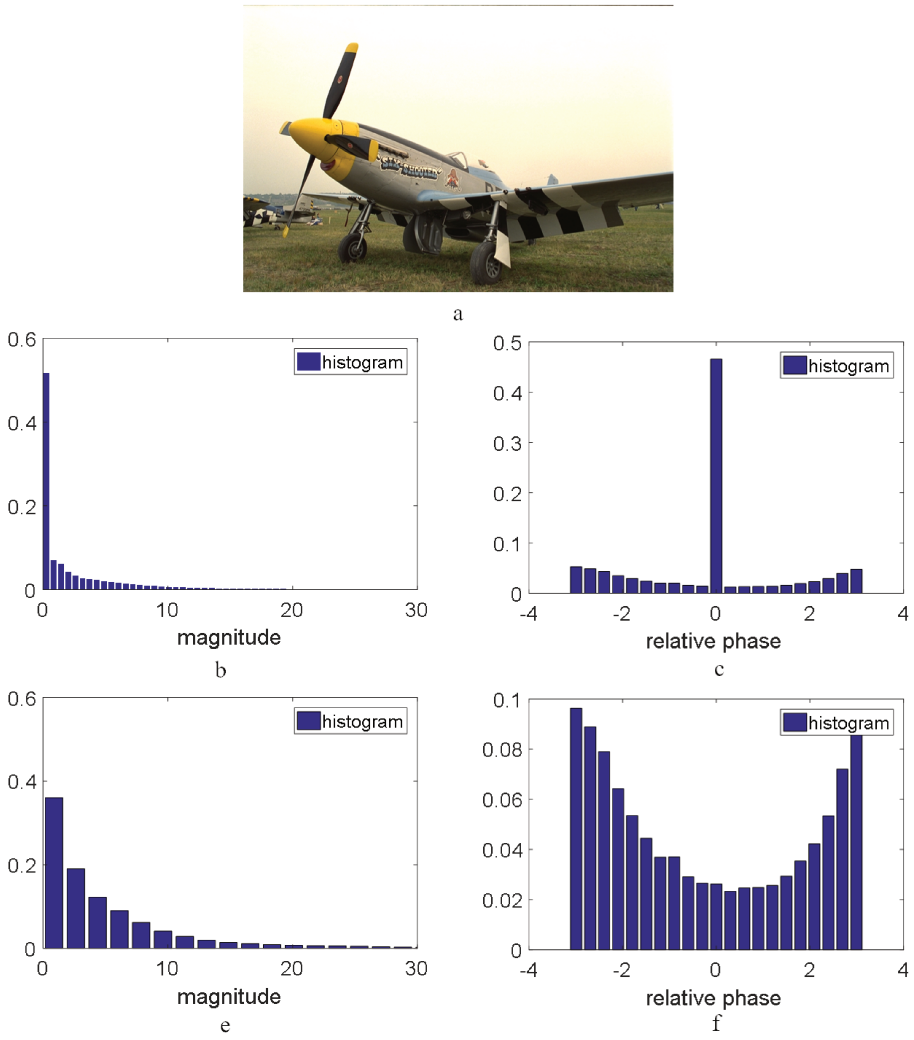


Fig. 3.3 Natural scene image "Plane" and its histograms of magnitude and relative phase in the DT-CWT domain. (a) Original image, (b) Histogram of magnitude, (c) Histogram of relative phase, (e) Histogram of magnitude after thresholding, and (f) Histogram of relative phase after thresholding.

3.4 Modelization

The objective of this section is to find a statistical model to capture the magnitude or the relative phase information in the complex wavelet domain. To reduce the model error between

the parameterized model and the real distribution, the information criteria (IC) is exploited to obtain the optimal histogram and then some candidate models are compared. The parameters of the selected models will be regarded as RR features.

3.4.1 Optimizing the histogram by IC

The IC is introduced to build an optimal histogram for the distribution and reduce the model error. Since the maximum likelihood criterion does not converge, it would lead to an overestimation of the number of free parameters of models [25]. Motivated by literature [13, 101], we use the IC to obtain the optimal number of bins of histograms. The purpose is to find a histogram that best summarizes the coefficients' PDF regarding IC. With a histogram of equal-width bins, the criterion is expressed as [101]

$$\mathrm{IC}(k) = \sum_{j=1}^{k} n_j \log n_j - n \log k + k c_n, \tag{3.3}$$

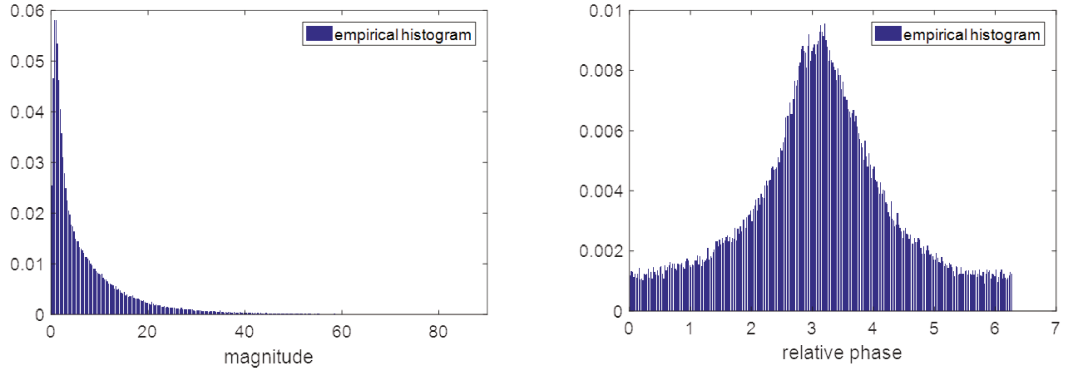
$$\hat{k} = \arg\min \mathrm{IC}(k), \text{ with } k = 1, ..., (2\sqrt{n} - 1), \tag{3.4}$$

where $n_j, j = 1, ..., k$ is the number of realizations of coefficients into the corresponding bin j, k is the number of histogram bins, n is the number of data coefficients by sub-band, c_n is the penalty function, and $\hat{k}$ is the estimation of the number of histogram bins in the sense of information criteria.

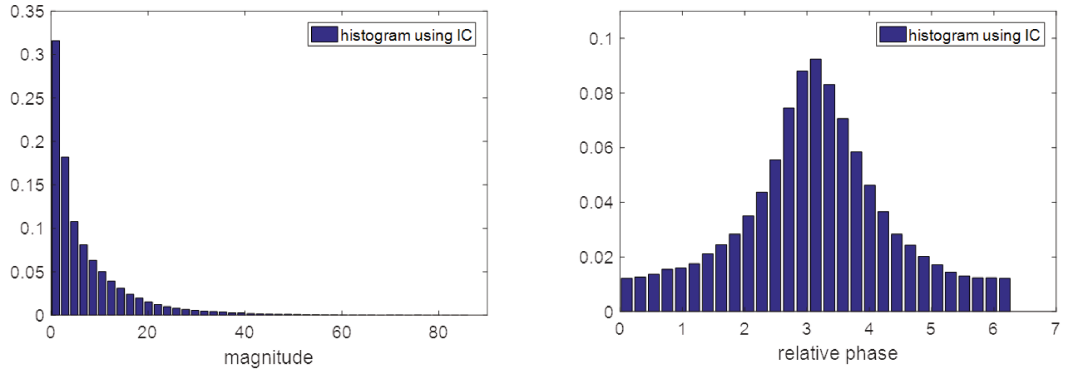
The first two terms in Eq. (3.3) are the log-likelihood terms. It is also the reason why IC is regarded as a penalized log-likelihood criterion. Different c_n values mean different IC, for example $c_n = 2$ defines the AIC [5] which is the earliest IC. This criterion will bring about an over-parameterization. Two other criteria respectively named BIC and φ_β are developed to overcome the inconsistency of AIC. The BIC criterion proposed by Schwarz [83] is strongly consistent, which is based on the Bayesian justification with $c_n = \log n$. The φ_β criterion was proposed by Matouat and Hallin [34], where $c_n = n^\beta \log\log n$. It is a strongly consistent criterion when $0 < \beta < 1$. On the basis of the φ_β penalty it is possible to obtain the penalties associated with the BIC and the AIC criteria: β_{BIC} is such that $n^{\beta_{\mathrm{BIC}}} \log\log n = \log n$, and β_{AIC} is such that $n^{\beta_{\mathrm{AIC}}} \log\log n = 2$ [6]. In addition, a refined condition: $\beta_{\min} = \frac{\log\log n}{\log n} < \beta < 1$ allows to adjust β according to the number of data coefficients by sub-band [6, 45]. Furthermore, it is noted that the optimal values of β must be between $\beta_{\min}$ and 1, and the values of β_{AIC} are not in the interval $[\beta_{\min}, 1]$ [87, 101]. It confirms that the inconsistency of AIC is contrary to BIC with β_{BIC} in $[\beta_{\min}, 1]$.

Motivated by [101], we utilize the criterion φ_β with $\beta = \beta_{\mathrm{BIC}}$ to estimate the optimal number of bins of histograms. Fig. 3.4 shows a comparison of the histograms using and

without using IC, where (a), (b) are the empirical histograms without using IC, and (c), (d) are the optimal histograms using IC. The KLD between the model and the distribution is used as a quantification criterion. It can be seen that a better estimation can be acheived with IC while minimizing the number of bins. These distribution shapes will contribute to the selection of the PDF.



(a) k=300, KLD=0.0403 (b) k=295, KLD=0.0020



(c) k=45, KLD=0.0216 (d) k=29, KLD=0.0006

Fig. 3.4 Comparison of the histograms using and without using IC (Histograms taken from the image "Bikes", the decomposition scale is 1, and the orientation is 15 degree): (a) Empirical histogram of the magnitude without using IC, (b) Empirical histogram of the relative phase without using IC, (c) Optimal histogram of the magnitude using IC and (d) Optimal histogram of the relative phase using IC.

3.4.2 Model of magnitude

Since the statistical property of the magnitude can be extracted directly, it is important to select an appropriate model. In [87], the Mixture Rayleigh Distribution (MRD) is selected to model magnitude. The MRD mixed by two Rayleigh distributions can obtain a better approximation than a single Rayleigh distribution. However, the fitting error is still a bit

large and there are too many parameters. We propose to model the magnitude by the Inverse Gaussian Distribution (IGD) [21] which has smaller errors and fewer parameters than MRD. The comparison between MRD and IGD is performed by an experimental method. The mathematical expressions of MRD and IGD are respectively,

$$R(x) = k_1 \frac{x}{\sigma_1^2} \exp\left(\frac{-x^2}{2\sigma_1^2}\right) + k_2 \frac{x}{\sigma_2^2} \exp\left(\frac{-x^2}{2\sigma_2^2}\right), \tag{3.5}$$

where σ_1, σ_2 are the scale parameters of the MRD, and k_1, k_2 are the weighting factor,

$$I(x) = \left(\frac{\lambda}{2\pi x^3}\right)^{\frac{1}{2}} \exp\left[-\frac{\lambda(x-\mu_v)^2}{2\mu_v^2 x}\right], \tag{3.6}$$

where λ is the sharp parameter, and μ_v is the mean (v is a subscript to distinguish other mean values).

To obtain the optimal parameters of models, we first use the maximum likelihood criterion to construct the cost function, namely, the KLD between the model and the wavelet sub-bands' PDF of the reference image. Then, the Nelder-Mead simplex method [49] is employed to minimize the cost function and determine the optimal parameters.

To find a better model, we compare the KLDs between the original distribution and two models. The model with smaller KLD will be selected. For this purpose, we take all the images in LIVE database 2 and calculate the KLDs for the two models. The results are compiled in Table 3.1, where the KLD of 5 single images and average KLD of all images are listed. Furthermore, Fig. 3.5 intuitively shows the fitting performance of the two candidate models, taking the image "Bikes" as an example.

Table 3.1 Comparison of KLD of the models MRD and IGD.

Image	Size	Sub-bands	MRD	IGD
Bikes	768×512	18	0.0466	0.0110
Caps	768×512	18	0.0257	0.0076
Cemetry	627×482	18	0.0463	0.0037
Buildings2	640×512	18	0.0296	0.0239
Womanhat	480×720	18	0.0256	0.0017
...	...	...	...	...
Average(all images)		18	0.0414	0.0157

The experimental results show that the average KLD of MRD (a four-parameter model) is around 0.04 and the average KLD of IGD (a two-parameter model) is about 0.01. Thus, we chose IGD as the model of magnitude.

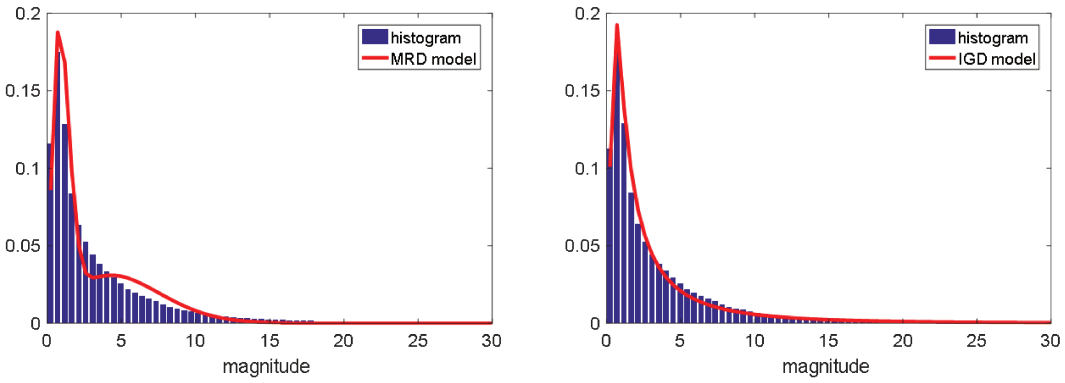


Fig. 3.5 Comparison of two PDF candidates (Histogram taken from the image "Bikes", the decomposition scale is 1, and the orientation is 45 degree). (a) Curve fitting of MRD and (b) Curve fitting of IGD.

3.4.3 Model of relative phase

Compared with the phase of the DT-CWT coefficients, the relative phase has a clearer statistical characteristic. An [105] found that the Wrapped Cauchy Distribution (WCD) can fit the PDF of relative phase well. From a preliminary study of the histograms of the relative phase in LIVE database 2, we found that the Generalized Gaussian Distribution (GGD) also can fit the PDF well. Thus, we compare the two models in an experimental way in this section.

The WCD is obtained by wrapping the Cauchy distribution on the real line with density $f(x)$ around the circle, where

$$f(x) = \frac{\sigma_w}{\pi[\sigma_w^2 + (x-\mu_w)^2]}, \quad -\infty < x < \infty. \tag{3.7}$$

It has the probability density function [61]

$$p(\theta) = \frac{1}{2\pi}\frac{1-\rho^2}{1+\rho^2-2\rho\cos(\theta-\mu_w)}, \quad -\pi \le \theta \le \pi, \tag{3.8}$$

where $\rho = e^{-\sigma_w}$ is the sharp parameter, σ_w is the scale factor, and μ_w is the mean (w is a subscript to distinguish other similar symbols).

The GGD model with function is defined as

$$G(x) = A_2 \exp[-(|A_1(x-\mu_g)|^{\alpha})], \tag{3.9}$$

where $A_1 = \frac{1}{\sigma_g}\sqrt{\frac{\Gamma(3/\alpha)}{\Gamma(1/\alpha)}}$ and $A_2 = \frac{A_1\alpha}{2\Gamma(1/\alpha)}$, $\alpha > 0$ is the sharp parameter, σ_g is the standard deviation, μ_g is the mean, and $\Gamma(\cdot)$ is the Gamma function.

Note that the histogram of the relative phase of a natural image is usually a concave curve, not a bell curve, and the relative phase values are defined between $-\pi$ and π. Thus, by considering that the GGD model is aperiodic symmetry, it is necessary to shift before building the histogram. Considering that the relative phase is periodic and the range of the relative phase value is $[-\pi,\pi]$, we can add 2π to the data in range $[-\pi,0]$ and then obtain the relative phase in range $[0,2\pi]$.

The shifting processing and a curve fitting example for image "Bikes" are shown in Fig. 3.6, where (a) is the original histogram and the curve fitting of WCD model (WCD model is periodic symmetry), (b) and (c) are the π-shift histograms, and respectively fitting by WCD and GGD models.

Table 3.2 Comparison of KLD of the GGD and WCD models.

Image	size	sub-bands	GGD	WCD
Bikes	768×512	18	0.0055	0.0048
Caps	768×512	18	0.0049	0.0045
Cemetry	627×482	18	0.0046	0.0020
Buildings2	640×512	18	0.0038	0.0012
Womanhat	480×720	18	0.0042	0.0024
...	...	...	...	...
Average (all images)		18	0.0042	0.0025

To select a more suitable model, we compare the two models in the same way as in Subection 3.4.2. The experimental results are shown in Table 3.2, where the average KLD of WCD is smaller than that of GGD. Thus, we will select WCD as the model of the relative phase.

In summary, we have selected the IGD model for the magnitude and the WCD model for the relative phase, respectively. The IGD model has two parameters, i.e., the sharp parameter λ and the mean μ_v. The WCD model also has two parameters: the sharp parameter ρ and the mean μ_w. Next, we will demonstrate the influence of image degradation on RR features and the implementation of FMRP metric.

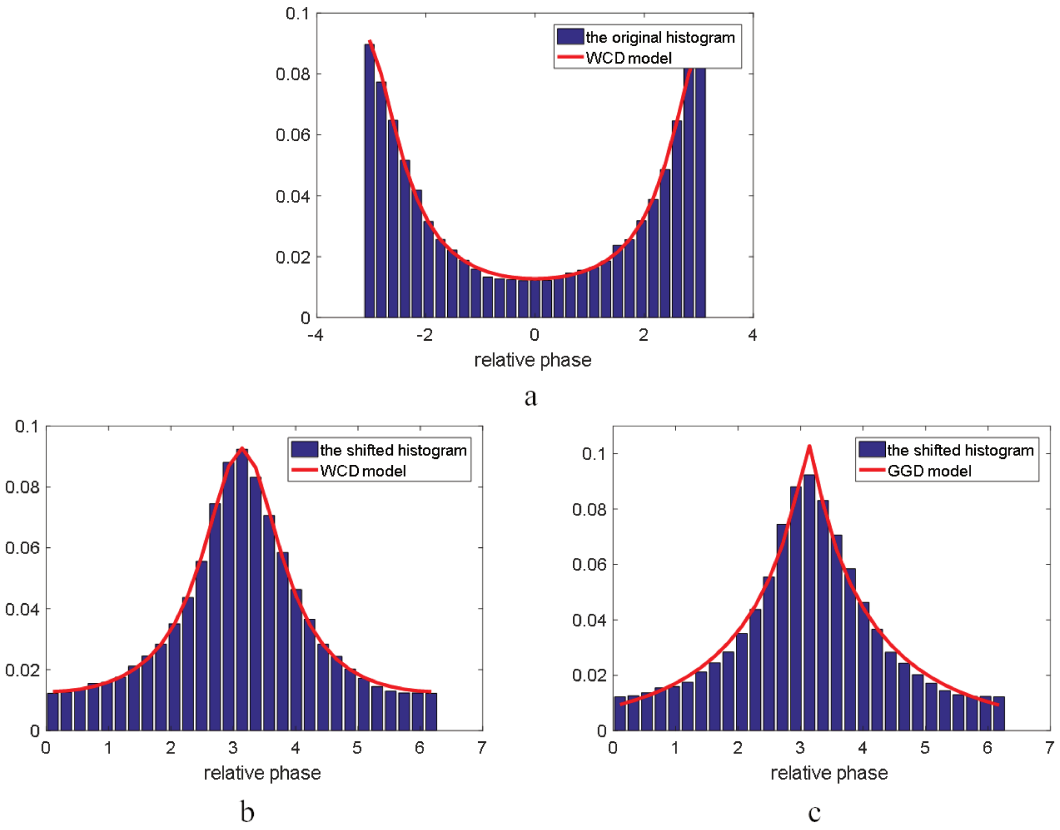


Fig. 3.6 Comparison of two PDF candidates (Histogram taken from the image "Bikes", the decomposition scale is 1, and the orientation is 15 degree). (a) Curve fitting of WCD, (b) Curve fitting of WCD, histogram after a phase shift of π, and (c) Curve fitting of GGD, histogram after a phase shift of π.

3.5 Influence of image degradation on RR features

To show the influence of image degradation on RR features, we extract the features from a set of images with different types of distortions and observe how these features vary with the distortions. The results are illustrated in Fig. 3.7. As described before, the KLD between the PDFs of the original image and the distorted one is exploited to evaluate the degree of the distortion. It is easy to observe that the PDFs of the magnitude and the relative phase vary in different ways for different types of image distortions. For the magnitude, the variation of the PDF due to the white Gaussian noise and Gaussian blur distortion is larger than other distortions, and this causes the magnitude feature more sensitive to these types of distortions (random error or information removal). The relative phase can detect local structural attacks like JPEG 2000, JPEG, and fast fading distortions.

The similar variation of PDF in the DWT-decomposed image has previously been observed and used for DWT-based quality assessment [108, 112]. Thus, we extract the features from the distribution models of the magnitude and the relative phase, and the parameters of the models are transmitted as the RR features in the real-time communication system.

As defined in Section 3.3, the KLD between the sub-band distributions of the distorted image and the reference one serve as the basis of the image quality evaluation. The KLD is symmetric and always positive. To make an effective estimation, the histograms of the DT-CWT coefficients for both images must be available. The histogram, $\hat{P}(x)$, is easy to compute from the distorted image but it is impossible to obtain the histogram of the reference image at the receiver side. However, we have verified that the histograms of the DT-CWT coefficients of the reference image can be estimated from the parameters of the well-selected models (namely IGD and WCD). Thus, we just need to transmit the parameters of the models to the receiver side. The parameters serve as RR features for the image quality evaluation.

Motivated by [7], we compute the KLD between the PDFs of the original image and that of the distorted image:

$$\hat{d}(P\|\hat{P}) = \int P_m(x) \log \frac{P(x)}{\hat{P}(x)} \mathrm{d}x = d(P_m\|\hat{P}) - d(P_m\|P), \tag{3.10}$$

where $P_m(x)$ is the density function model, $d(P_m\|\hat{P})$ is the KLD between the model and the distortion image, and $d(P_m\|P)$ is the KLD between the model and the original image, namely modelling error calculated at the transmitter side. The modelling error should be sent to the receiver side as the RR features to define a more precise estimation. In this thesis, we don't exploit the direct expression of the KLD for the IGD distributions or the WCD distributions but we propose to estimate it with a sampling of the KLD computation. To define a "sampling" rate, we propose to take the IC concept. Thus, $d(P_m\|P)$ is computed in a discretization form as

$$d(P_m\|P) = \sum_{n=1}^{N} P_m(n) \log \frac{P_m(n)}{P(n)}, \tag{3.11}$$

where n is the number of histogram bins of the PDF and N is the maximum number of bins. N can be optimized with the IC introduced in Subsection 3.4.1. $d(P_m\|\hat{P})$ is computed in the same way.

Finally, the objective scores are computed by a machine learning regression method according to the features of the distorted images. To obtain an appropriate range of values for the training model, we normalize the value of the features with a logarithmic function.

$$M_i \text{ or } R_i = \log_{10}(1 + k\hat{d}_i(P_i\|\hat{P}_i)) \tag{3.12}$$

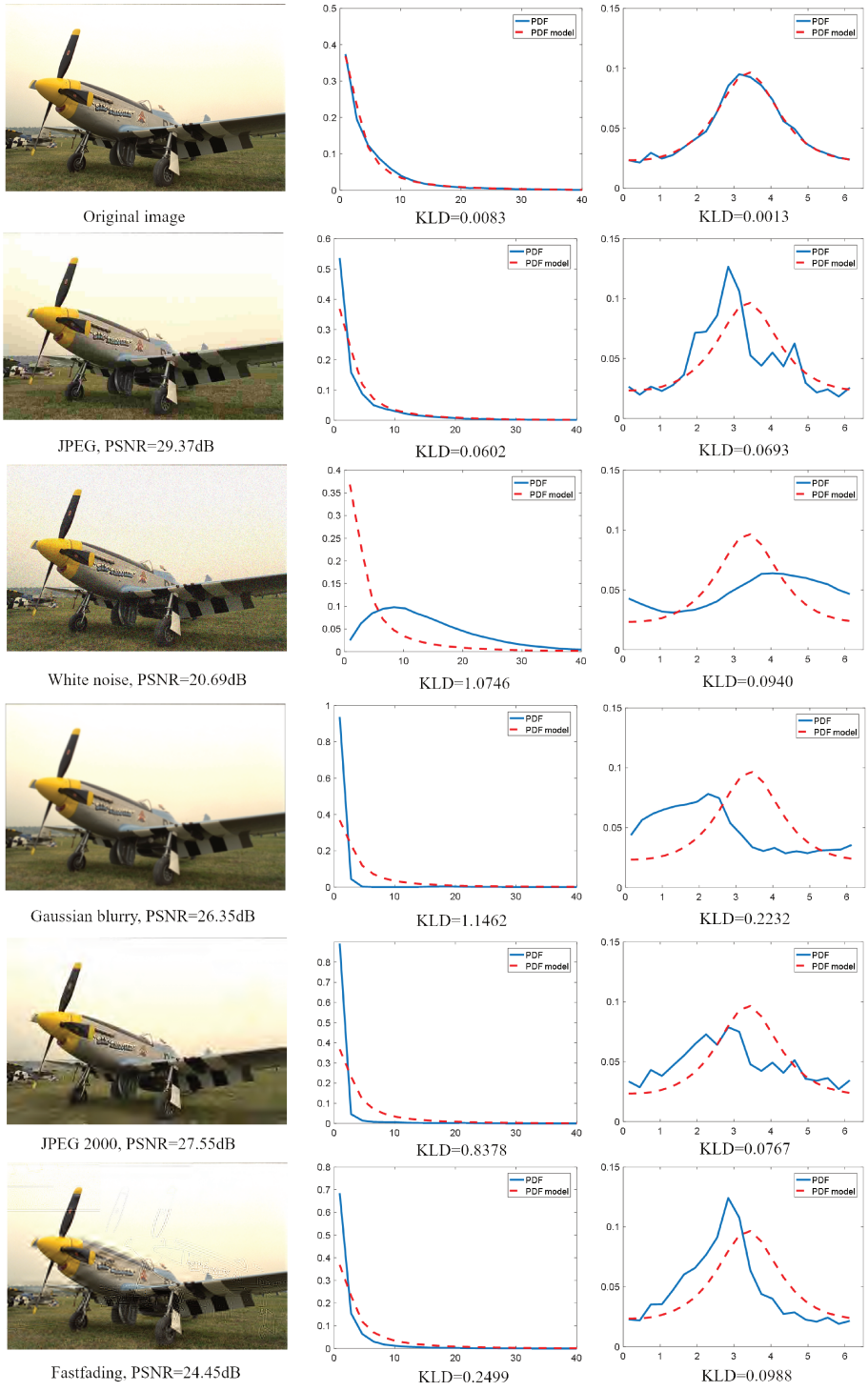


Fig. 3.7 Comparison of the PDFs of magnitude and relative phase from different distortions. The first column shows the original image and the distorted images, the second column shows the PDF of the magnitude, and the third column shows the PDF of the relative phase. (The red dashed curve and the blue solid curve are the PDF of the model and that of the distribution, respectively.)

where $i = 1, ..., 18$ is the index of the features, M_i and R_i are respectively the magnitude and the relative phase of the sub-bands of the distorted image, $\hat{d}_i(P_i \| \hat{P}_i)$ is calculated according to Eq. (3.10), and k taking 1000 is a constant to control the scale of the distortion metric. The image is decomposed with a three-scale six-orientation DT-CWT. Therefore, we can extract 36 features from 18 sub-bands, including 18 magnitude and 18 relative phase features.

3.6 Machine learning method for regression

After obtaining the features from the magnitude and the relative phase, the next step is to map the features into subjective scores. The goal is to find a function $f(\cdot)$ that uses the calculated indices as inputs and predicts an objective score S_0 for each image. The function $f(\cdot)$ expressed:

$$S_0 = f(M_1, R_1; M_2, R_2; ...; M_{18}, R_{18}) \tag{3.13}$$

where M_i and R_i are respectively KLDs of the magnitude and the relative phase between the distortion image and the reference one for i-th sub-bands, referring to section 3.5 and Eq. (3.12).

In recent years, researchers have tended to employ machine learning methods to develop regression function, such as Support Vector Regression (SVR) [16, 30] and general regression neural network (GRNN) [52, 54, 55, 73]. Considering the time consumption and the complexity of implementation, we select GRNN as a regression tool.

GRNN is a kind of Radial Basis Function (RBF) neural network that is often used for function approximation based on a linear or non-linear regression. Its main function can be represented as [95]

$$\hat{Y}(x) = \frac{\sum_{i=1}^{n} Y_i \exp\left(\frac{D_i^2}{2\sigma^2}\right)}{\sum_{i=1}^{n} \exp\left(-\frac{D_i^2}{2\sigma^2}\right)} \tag{3.14}$$

where $D_i^2 = (X - X^i)^T (X - X^i)$, D_i^2 is the squared Euclidean distance between the measured value X and the sample observation X^i, Y^i is the score associated with X^i, n is the number of sample observations of training database, and σ is the standard deviation of radial kernal (a larger σ may bring a smoother regression surface but less accuracy while a smaller one may provide higher approximation accuracy but a worse smooth surface [94]).

We implement the GRNN by the Matlab function newgrnn(). The only parameter involved is the spread parameter $p \in [0, 2]$, linked with the standard deviation σ described previously. To set this parameter, we optimize the value of p with the 10-fold method. The process of the 10-fold method is as follows:

1. Dividing the database into 10 sub-sets according to the 10-fold method.
2. Computing the prediction score with $p = 0 : 0.1 : 2$.
3. Computing the correlation coefficients between the prediction score and subjective one, and comparing the mean values of 10 sub-sets for different spread parameters.
4. Obtaining the spread parameter that maximizes the correlation coefficient as the optimal value.

The experimental results show that the optimal p is 0.8. Thus, we set the spread parameter to 0.8 in the next step.

Another important issue is to design the training and testing subsets. Four classical databases are utilized to test the performance of the proposed image quality metric and two commonly adopted strategies are employed to design training and testing subsets, as detailed in Section 3.7.

Since the proposed metric maps the RR features into the objective score with GRNN, the objective score will vary with the subjective score provided by the database. If the subjective score is MOS, the objective score will also be MOS. In addition, if the subjective score is DMOS, the objective score will be DMOS. The MOS measures the similarity of the distorted image and the reference one (larger MOS means less distortion) while the DMOS measures the difference (larger DMOS means larger distortion). Suppose we take the LIVE database 2 (DMOS as subjective score) as the training set, the objective score will be DMOS. Fig. 3.8 shows the objective scores of FMRP of image "Buildings" in LIVE database 2, where DMOS is the subjective score of the database, and FMRP is the objective score of FMRP metric.



DMOS=0, FMRP=0



DMOS=49.58, FMRP=46.17, PSNR=25.22



DMOS=61.18, FMRP=57.62, PSNR=20.92



DMOS=61.41, FMRP=58.50, PSNR=22.45

Fig. 3.8 The objective scores of FMRP of image "Buildings" in LIVE database 2.

3.7 Test results

3.7.1 Test databases and evaluation criteria

To evaluate the performance of the proposed image quality metric, four publicly available image databases, including LIVE database 2 [91], CSIQ [50], TID2013 [71] and Toyama [100], are selected. The characteristics of these four databases are shown in Table 3.3. Note that CSIQ and TID2013 include some distortions, such as contrast change, change of colour saturation and intensity shift, which are designed to test the FR-IQA metrics and usually do not occur in communication systems. Thus, we remove these types of distortions from CSIQ and TID2013, remaining 5 types of distortions in CSIQ (6 in total) and 18 types of distortions in TID2013 (24 in total), and call them CSIQ*[1], TID2013*[2], respectively.

Table 3.3 Information of the image databases.

Database	Distortion types	Reference images	Distorted images	Distortion levels	Subjective evaluation	
					Score type	Scale
LIVE database 2	5	29	982	5-9	DMOS	[0, 100]
CSIQ*	5	30	750	5	DMOS	[0, 1]
TID2013*	18	25	2250	5	MOS	[0, 8]
Toyama	2	14	196	6	MOS	[1, 5]

According to the suggestion given by the Vide Quality Experts Group [44], we employ three evaluation criteria to quantify the performance of our metric. The first one is the Pearson linear correlation coefficients (PLCC) used to evaluate the prediction accuracy. The second one is the Spearman rank-order correlation coefficients (SROCC) used to evaluate the prediction monotonicity. The last one is the outlier ratio (OR) to evaluate the prediction consistency. The definition of the three evaluation criteria is introduced in Section 2.1.

3.7.2 Decomposition scales and the features

As introduced in Section 3.5. the influences of the image degradation mainly lie on the first three decomposition scales of the DT-CWT. Table 3.4 lists the prediction accuracy of different decomposition scales for all images of the LIVE database 2. Obviously, a 3-scale decomposition is better than a 2-scale one and almost with the same accuracy as a 4-scale one.

[1]CSIQ* includes 5 types of distortions in CSIQ with the "Contrast" distortion removed.

[2]TID2013* is generated from TID2013 by removing 6 kinds of distortions including "Non eccentricity pattern noise", "Local block-wise distortions of different intensity", "Mean shift", "Contrast change", "Change of colour saturation", "Chromatic aberrations" distortions.

Thus, the 3-scale DT-CWT is selected to decompose the image, which is the best trade-off between the accuracy and the size of transmitted information.

Table 3.4 Prediction accuracy of 2, 3 and 4 decomposition scales.

Scales	PLCC	SROCC	OR
2 scales	0.9498	0.9308	0.0096
3 scales	0.9731	0.9638	0.0035
4 scales	0.9738	0.9640	0.0045

To analyze the reliability of RR features on detecting the image distortion, the RR features from a group of images with different distortion levels (JPEG 2000 compression distortion) are compared. The results are shown in Tables 3.5 and 3.6, where "Distortion 1,...,4" denotes 4 different levels of distortions of JPEG 2000 compression sorted in increasing DMOS, and KLD is the Kullback-Leibler Divergence between the distributions of the reference image and the distorted one. One can see that the KLDs in three scales vary significantly with different levels of degradations, and the variation decreases with the increase in the number of scales. The observation illustrates that the features can detect the distortion very well and the image degradation usually affects the high frequency signals (scale 1), which holds information of textures and edges in an image. With the increase of the decomposition levels, the influence of degradation on the features of the image decreases progressively and consequently, the variation of KLDs and the information of textures and edges decrease correspondingly. This confirms a decomposition just with three scales is acceptable.

Table 3.5 KLD of the sub-bands of the magnitude (from image "Building2" of LIVE database 2).

	Distortion 1	Distortion 2	Distortion 3	Distortion 4
	25.4774	31.5596	48.0727	69.2985
Sub-band	KLD			
Scale 1, 75 deg.	0.0347	0.0692	0.1820	0.3884
Scale 1, 15 deg.	0.0389	0.0751	0.1170	0.2665
Scale 1, 45 deg.	0.0356	0.1151	0.3753	0.2792
Scale 2, 75 deg.	0.0054	0.0157	0.0496	0.1842
Scale 2, 15 deg.	0.0068	0.0107	0.0514	0.1874
Scale 2, 45 deg.	0.0096	0.0183	0.0581	0.2679
Scale 3, 75 deg.	0.0010	0.0005	0.0127	0.0457
Scale 3, 15 deg.	0.0012	0.0011	0.0137	0.0435
Scale 3, 45 deg.	0.0004	0.0028	0.0115	0.0369

Table 3.6 KLD of the sub-bands of the relative phase (from image "Buildings2" of LIVE database 2).

	Distortion 1	Distortion 2	Distortion 3	Distortion 4
	25.4774	31.5596	48.0727	69.2985
Sub-band	KLD			
Scale 1, 75 deg.	0.0167	0.0502	0.1961	0.2400
Scale 1, 15 deg.	0.0164	0.0583	0.1747	0.2188
Scale 1, 45 deg.	0.0041	0.0069	0.0246	0.0416
Scale 2, 75 deg.	0.0017	0.0075	0.0672	0.1436
Scale 2, 15 deg.	0.0022	0.0108	0.0802	0.1936
Scale 2, 45 deg.	0.0013	0.0095	0.0787	0.1174
Scale 3, 75 deg.	0.0005	0.0009	0.0073	0.0298
Scale 3, 15 deg.	0.0005	0.0018	0.0074	0.0692
Scale 3, 45 deg.	0.0006	0.0004	0.0020	0.0216

To further analyze how the magnitude and the relative phase features vary with distortion types, we take magnitude and relative phase as independent metrics respectively and compare them with the features of DWT [112]. The experimental results are illustrated in Table 3.7, where we utilize LIVE database 2 as test database, including JPEG compression, JPEG2000 compression, Gaussian blur, white noise and fast fading distortions. The PLCC of the prediction scores and the subjective scores serve as a quantifiable criteria for performance. One can see that the features of magnitude and relative phase of DT-CWT have a better performance than those of the DWT. It mainly because of the advantages of the DT-CWT, including shift invariance, non-oscillations, multi-directions and phase information. Another reason is the use of GRNN in the mapping between RR features and objective scores, which performs better than the logistic function. Moreover, Table 3.7 shows that the magnitude feature is sensitive to JPEG compression, white noise and Gaussian blur distortions, and the relative phase feature is sensitive to JPEG 2000 compression and fast-fading distortions. Therefore, to take full advantage of the two features, we combine them to form a new image quality metric, i.e., FMRP. The FMRP metric has the highest accuracy in detecting five distortions of LIVE database 2 and the best robustness in detecting all mixed distortions.

Table 3.7 Comparison of the features of DWT and DT-CWT.

Quantifiable criteria	Feature source	JP2K	JPEG	WN	BLUR	FADING	ALL
PLCC	Feature of DWT	0.9651	0.9407	0.9488	0.9465	0.9306	0.8478
	Magnitude	0.9690	0.9629	0.9719	**0.9866**	0.9545	0.9300
	Relative phase	0.9725	0.9566	0.9589	0.9825	0.9596	0.9500
	Combination	**0.9768**	**0.9730**	**0.9804**	0.9836	**0.9689**	**0.9731**

3.7.3 Comparison with other metrics

To assess the detection performance of the mixed distortions, we compare the FMRP metric with six other metrics, including two FR metrics: SSIM [111], MS-SSIM [110] and four RR metrics: Wang Z. et al.'s RR method [112] (we call it DWT), ADI [54], QWT [101] and Lin Z. et al.'s RR method [55] (we call it RP). The implementation code of the SSIM, MS-SSIM and Wang Z. et al.'s methods are available at the Waterloo IVC Code Repository [107]. According to the algorithms in [54, 55, 101], we respectively have implemented the other metrics. To have a robust evaluation of the metric, two strategies are employed to design subsets.

Strategy 1: Two-thirds of data are randomly picked out for training, while the remaining data are for testing. The final result is the mean of 1000 replicates of the process.

Strategy 2: A 10-fold cross-validation style. The database is randomly partitioned into 10 equally sized subgroups. A single subgroup is selected as the testing data, while the remaining 9 subgroups are used as training data. This process is repeated in such a way that each of the 10 subgroups is selected just once as the validation (testing) data.

All the performances of these metrics are summarized in Table 3.8. We utilize Strategy 1 for set-partitioning in this step. To increase the readability of the results, we mark the first-best and the second-best metrics in bold. It is shown that the proposed metric FMRP performs quite well for the four databases. It provides a better prediction accuracy (with higher PLCC), better prediction monotonicity (with higher SROCC) and better prediction consistency (with lower OR) than most other metrics. Compared with other RR metrics, the FMRP metric achieves the best results on all the databases. Compared with the representative FR metric MS-SSIM, it also performs quite well, with a higher accuracy in LIVE database 2 and a very close accuracy in the other databases. By recalling that the proposed metric FMRP is an RR metric, it uses less information in image evaluation than the FR metrics. The experimental results verify that the proposed metric FMRP is a useful and reasonable RR metric.

Table 3.8 Performance comparison of the objective quality metrics.

		Objective quality metrics						
Database	Criteria	SSIM	MS-SSIM	DWT	ADI	QWT	RP	FMRP
LIVE	PLCC	0.9274	0.9424	0.8478	0.9334	0.9126	**0.9487**	**0.9731**
	SROCC	0.9047	**0.9387**	0.8349	0.9201	0.8845	0.9314	**0.9638**
	OR	0.0173	**0.0041**	0.3646	0.0217	0.0314	0.0110	**0.0035**
Toyama	PLCC	0.8851	**0.9349**	0.8231	0.9188	0.8035	0.8712	**0.9322**
	SROCC	0.8862	**0.9304**	0.8283	0.9164	0.8124	0.8760	**0.9313**
	OR	0.0223	**0.0000**	0.1071	0.0199	0.1526	0.0494	**0.0093**
CSIQ*	PLCC	0.8199	**0.9109**	0.7336	0.8591	0.7826	0.8371	**0.9114**
	SROCC	0.8304	**0.9183**	0,7484	0.8703	0.7683	0.8210	**0.9085**
	OR	0.1080	**0.0267**	0.1440	0.0719	0.1236	0.0748	**0.0307**
TID2013*	PLCC	0.7033	**0.8544**	0.6511	0.7565	0.7583	0.7394	**0.8504**
	SROCC	0.7178	**0.8574**	0.6097	0.7458	0.7512	0.7283	**0.8549**
	OR	0.1480	**0.0560**	0.2093	0.1243	0.1157	0.1261	**0.0686**

Further, we compare the FMRP with three other RR metrics to test their performances for different types of distortions. The results are illustrated in Table 3.9, where we list the major distortions appearing in communication systems. It is obvious that the FMRP metric performs better for most distortions, such as compression, transmission error, and additive noise distortions. It can be interpreted by the good perception and the robustness of the magnitude and the relative phase. For additive Gaussian noise and Gaussian blur distortions, the ADI metric performs best. As introduced before, it is an energy-based method in the wavelet transform domain and sensitive to changes in the magnitude.

Table 3.9 Performance comparison of RR metrics for different distortions.

Distortion type	Database	PLCC			
		DWT	ADI	RP	FMRP
JPEG compression	LIVE	0.9407	0.9068	0.9474	**0.9730**
	Toyama	0.8915	0.9110	0.8919	**0.9306**
	CSIQ*	0.8758	0.9079	0.8978	**0.9185**
	TID2013*	0.8541	0.8729	0.8638	**0.8934**
JPEG 2000 compression	LIVE	0.9651	0.9447	0.9552	**0.9768**
	Toyama	0.9469	0.9222	0.9438	**0.9475**
	CSIQ*	0.9326	0.9322	0.9308	**0.9330**
	TID2013*	**0.9335**	0.9201	0.9084	0.9254
Additive Gaussian noise	LIVE	0.9488	**0.9877**	0.9081	0.9804
	CSIQ*	0.8126	0.8857	0.8851	**0.8865**
	TID2013*	0.6789	**0.8851**	0.8446	0.8214
Gaussian blur	LIVE	0.9465	0.9719	0.9684	**0.9836**
	CSIQ*	0.8806	0.9086	**0.9218**	0.9090
	TID2013*	0.9158	**0.9736**	0.9552	0.9296
JPEG 2000 trans. error	LIVE	0.9306	0.8702	0.9439	**0.9731**
	TID2013*	0.7721	0.7054	0.7263	**0.8127**
JPEG trans. error	TID2013*	0.8793	0.8922	0.8277	**0.9060**
Additive pink noise	CSIQ*	0.7925	0.8454	0.8427	**0.8864**

3.7.4 Robustness and cross-validation

The proposed method is based on machine learning regression techniques, so its performance depends upon the training sets to some extent. To evaluate the robustness of the proposed FMRP metric, two strategies for designing the training and testing subsets (detailed in the previous subsection) are utilized.

The test is divided into two parts: cross-validation and cross-database validation. For the cross-validation, we utilize Strategy 1 and Strategy 2 for set-partitioning of the databases and then conduct the training and testing process for each database. The experimental results of cross-validation in databases are given in Table 3.10. The numerical values of the three evaluation criteria are very close. It shows that the proposed RR image quality metric is not sensitive to the original choice of training subset.

Moreover, to perform a cross-database validation, we integrate all the images of four databases and convert different objective scores for different databases to unify objective scores by linear scale-functions. With the integration, we have around 4200 images in

Table 3.10 Performance of cross-validation for FMRP.

Criteria	Subsets selection	LIVE	Toyama	CSIQ*	TID2013*
PLCC	Strategy 1	0.9731	0.9322	0.9114	0.8504
	Strategy 2	0.9751	0.9435	0.9067	0.8583
SROCC	Strategy 1	0.9638	0.9313	0.9085	0.8549
	Strategy 2	0.9643	0.9333	0.8966	0.8593
OR	Strategy 1	0.0035	0.0093	0.0307	0.0686
	Strategy 2	0.0025	0.0066	0.0301	0.0654

total and then carry out the training and testing with Strategy 1. To form a comparison of performance, two other RR-IQA methods based on machine learning, ADI [54] and RP [55], are compared with the proposed FMRP method. The results are presented in Table 3.11, where stdCC and stdSR are respectively the standard deviations of 1000 PLCCs and SROCCs (1000 iterations for the training and testing process). Since the database TID2013 including some exotic distortions such as image color quantization with dither distortion, additive noise in color component, which are a challenge for most image quality metrics, the accuracy of the cross-database validation is lower than that of other databases. However, the proposed RR image quality metric, FMRP, still performs better than the ADI and the RP metrics in cross-database validation evaluation and has a significant improvement in accuracy. This comfirms the combination of the magnitude and the relatvie phase features outperforms the single magnitude feature or the single relative phase feature in detecting mixed distortions (as described previously, the RP method is mainly based on relative phase feature and the ADI is an energy-based method).

Table 3.11 Performance of cross-database validation for FMRP and two other RR metrics.

Database	Metric	PLCC	SROCC	OR	stdCC	stdSR
merge of 4 database	FMRP	**0.8826**	**0.8806**	**0.0482**	**0.0114**	**0.0121**
	RP	0.8047	0.7964	0.0976	0.0210	0.0221
	ADI	0.7705	0.7581	0.1143	0.0198	0.0225

3.7.5 Complexity and two inaccurate examples

For complexity analysis, we enumerate memory cost and computation time of all the metrics used in the previous section, as shown in Tables 3.12 and 3.13, respectively. The extracted parameters are quantized to an 8-bit finite precision. As introduced in section 3.4, computation of each feature of the FMRP metric involves three parameters, so 36 features cost

$36 \times (8+8+8) = 864$ bits, i.e., 0.27% of the reference image, as shown in Table 3.12. It is a proper memory cost for the communication system. In addition, the memory cost of the FMRP metric is independent on the image size.

Table 3.12 The size of features of different metrics.

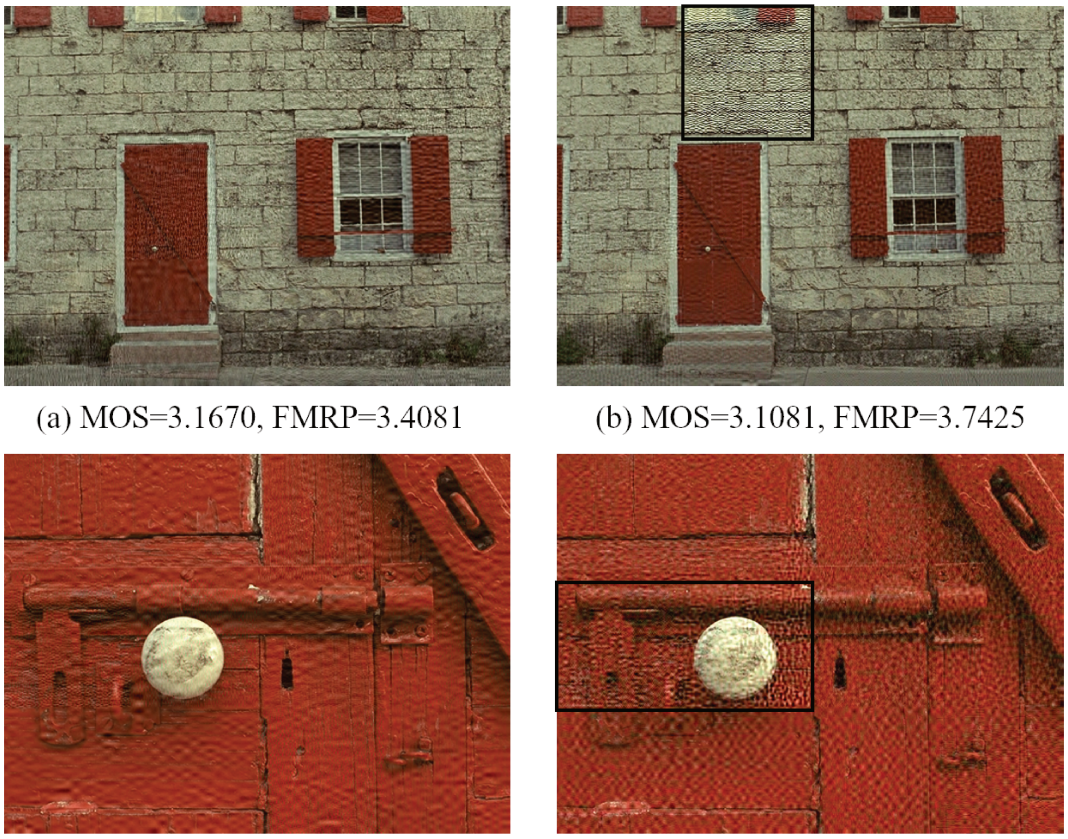
Metrics	SSIM	MS-SSIM	DWT	ADI	QWT	RP	FMRP
Bits to transmit (bit)	3145728	3145728	144	96	360	1248	864
Percentage (%)	100.00	100.00	0.05	0.03	0.11	0.40	0.27

The calculation of objective scores involves two steps: (1) Calculation of the similarity or specific distance between the distorted image and the reference one, (2) Transferring from the similarity or the specific distance to objective scores by a regression function. Since most of the metrics employ the same regression function and the time of this step is quite short, we compare the time of step (1) of all the metrics, as shown in Table 3.13. The test is performed on a PC with 2.7 GHz Intel Core i7 CPU and 8.00 GB RAM, and the development platform is Matlab R2016a. The computation time of the FMRP is 1.3618 seconds, which is slightly more than that of the DWT metric.

Table 3.13 Time consumption of different metrics.

Metrics	SSIM	MS-SSIM	DWT	ADI	QWT	RP	FMRP
Time (sec.)	0.8972	0.9123	1.2454	0.2972	2.1269	1.3435	1.3618

Two inaccurate examples of the FMRP metric are illustrated in Fig. 3.9, where the objective scores of the FMRP are not agreed with the assessors' MOS. Taking Fig. 3.9 (a) and (b) as the first example, the MOSes of theirs are in a decreasing order, but the FMRP scores are in an increasing order. The second example shown in Fig. 3.9 (c) and (d) reveals as the same information as Fig. 3.9 (a) and (b). The distortion resulting in the phenomenon often is the JPEG transmission error in database TID2013. For this type of distortions, the changes of the magnitude and the relative phase in the wavelet domain are quite small, which triggers bad perception to the distortion. Moreover, the distortion occurs in some partial blocks (we call it local distortion), as shown with the black rectangular box in Fig. 3.9. The local distortion is a challenge to most metrics, especially for the metrics with global statistical features. We will study this issue in the next chapter.



(a) MOS=3.1670, FMRP=3.4081 (b) MOS=3.1081, FMRP=3.7425

(c) MOS=3.3889, FMRP=2.8530 (d) MOS=2.8919, FMRP=3.0339

Fig. 3.9 Two inaccurate examples of the FMRP metric in TID2013 database.

3.8 Chapter summary

In this chapter, an RR-IQA metric based on the statistical model in the dual-tree complex wavelet domain is proposed. The features are extracted from the magnitude and the relative phase of the image DT-CWT coefficients. Moreover, a strategy based on IC is designed to find a statistical model to capture the magnitude or the relative phase information in the complex wavelet domain. The parameters of these models are used as RR features. Thus, the metric is named as feature of magnitude and relative phase (FMRP).

It is demonstrated that the magnitude feature is sensitive to JPEG compression, white noise and Gaussian blur distortions, while the relative phase feature is sensitive to JPEG 2000 compression and fast-fading distortions. These features offer additional information for the evaluation of the degradation of image quality. The degradation measure is obtained with the computation of the Kullback-Leibler Divergence between the sub-bands of the distorted image and the reference one. Finally, to clarify the relation between the features (the KLD measures) and subjective scores, a general regression neural network is introduced.

Compared with the representative FR-IQA metrics and the other RR-IQA metrics, the proposed reduced-reference image quality metric performs well in prediction accuracy, prediction monotonicity, prediction consistency and robustness among various distortion types, and demonstrates that the features are quality-aware and the metric is highly correlated with the human visual system. Moreover, the experimental results indicate our strategy is independent on the used database for the training process.

Therefore, the proposed FMRP metric is in line with the expectation: (1) detecting various distortions encountered in communication systems such as JPEG compression, JPEG2000 compression, JPEG transmission error, JPEG 2000 transmission error and white noise, (2) being competitive with the state-of-the-art FR-IQA methods, (3) summarizing the image information in a low data rate and a low computational complexity. The proposed RR image quality metric is very suitable to be applied in communication systems for evaluating the image quality. Next, we will utilize it to optimize the image decoding strategy in a wireless communication environment.

Chapter 4

Optimization of JPWL decoding in realistic wireless transmission

The rapid development of communication technologies has enabled high-quality multimedia services and wide applications of high-quality devices. In turn, the popularity of high-quality multimedia services and devices are demanding the image or video communication system with better Quality of Service (QoS) and better Quality of Experience (QoE). In this chapter, we devote to applying the RR-IQA metric (FMRP proposed in Chapter 3) into a realistic wireless communication system to improve the QoE while ensuring the QoS. The features of this metric are employed to automatically optimize the decoding strategy by selecting an image with better quality in consistent with users. To this end, a new image database has been constructed to collect the preference of users when they are selecting different decoding images. Based on the new database, experimental comparison demonstrates the decoding strategy with the FMRP metric has a significant improvement than the typically classic decoding strategies.

The chapter is organized as follows: Section 4.1 introduces the motivation of taking the RR-IQA metric to optimize the decoding strategy and the database construction. The global scheme and the classifier design will be presented in Section 4.2. The database construction details including experimental methodology, simulation environment and raw data processing are described in Section 4.3. The performance of the decoding strategy with the FMRP metric is validated in Section 4.4. A new metric FMRP-LSD used to detect the local distortion is proposed in Section 4.5. The chapter summary is presented in Section 4.6.

4.1 Motivation

Over two decades, with massive development of wireless technologies, such as WIFI, 4G, cellular data and multimedia services, and wireless sensor network, wireless transmission has become more and more important. However, the wireless channel is very turbulent due to the multipath transmission mixed by direct path, reflection, diffraction and dispersion paths. Furthermore, the Doppler effect due to the mobility of the receiver also introduces the disturbance. Therefore, images are easily degraded due to channel noise and interference, when transmitted over a wireless channel, and they are very sensitive to noises, even with few errors.

To cope with the channel variability, many technologies, such as ECC [65, 92], OFDM [35, 113], MIMO [12, 28] technologies, have been developed in the wireless channel and been contained in JPEG 2000 standard and JPWL (JPEG2000 WireLess) standard, which is an extension of the JPEG 2000 baseline specification to enable the efficient transmission of JPEG 2000 code-stream over an error-prone network [32]. These technologies exploit the diversity of time, frequency and space to optimize the transmission. For example, a strategy combined the aforementioned technologies with the content-based hierarchical transmission technique has demonstrated a good improvement in QoS [2]. The strategy converts the JPWL image into hierarchical layers based on the image content and correspondingly divides the whole channel into independent and hierarchical subchannels, thereby improving the QoS by the UPA and the UEP technologies, as shown in Fig. 4.1.

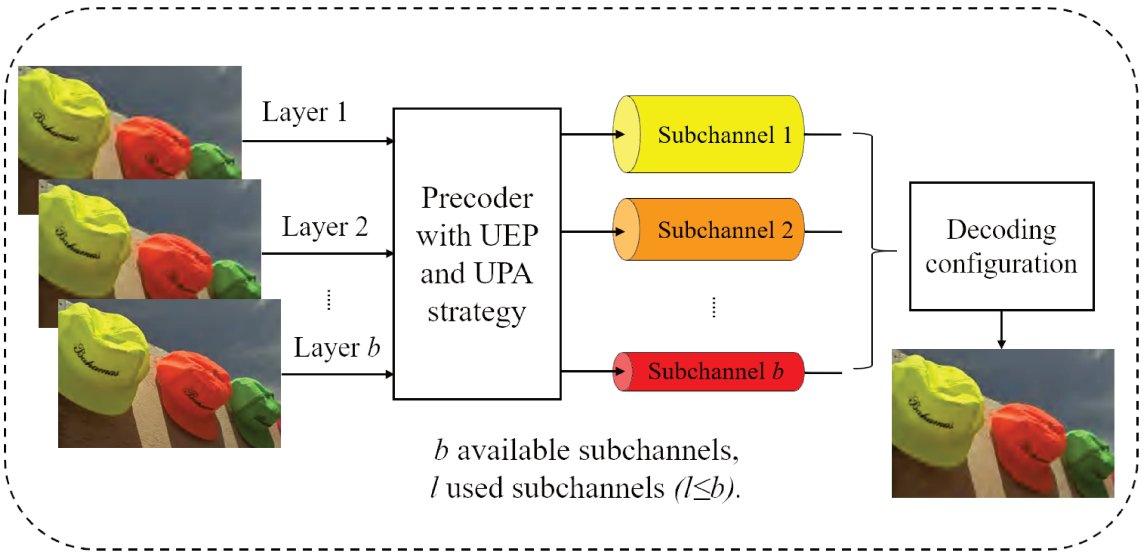
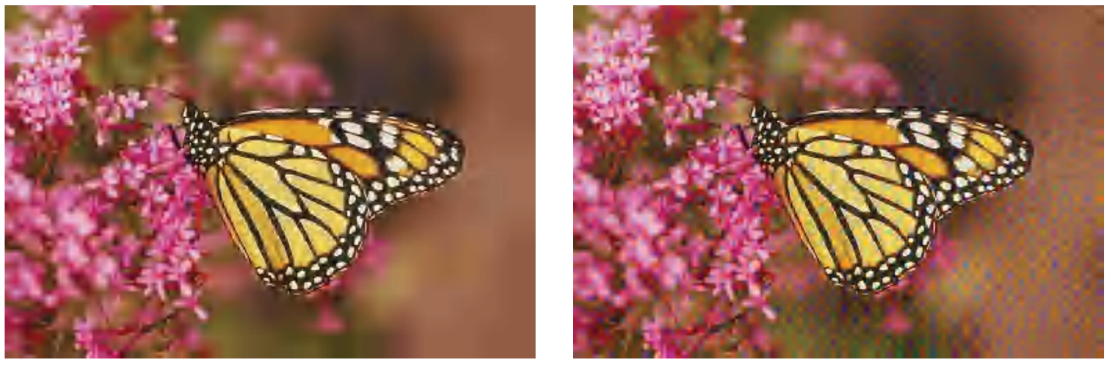


Fig. 4.1 Content-based hierarchical transmission system with UPA and UEP technologies.

According to the channel status, the UPA and the UEP strategies allocate scaled powers and protections to the independent subchannels and match the hierarchical subchannels with the hierarchical quality layers to improve the robustness up to a target data-bit-rate and a target Bit Error Rate (BER). At the receiver terminal, the decoding strategy selects the proper decoding configuration in line with the channel condition. For example, if the channel condition is good, the decoding configuration will be b layers, otherwise, the decoding configuration will be l layers ($l < b$). Thus, there are b available subchannels and l used subchannels ($l \leq b$), according to the specific channel condition.



(a) "Monarch" with better QoE

(b) "Monarch" with better QoS

(c) "House" with better QoE

(d) "House" with better QoS

Fig. 4.2 Image quality comparsion, where four images are obtained over a realistic wireless channel, (b) and (d) have better QoS, and (a) and (c) have better QoE.

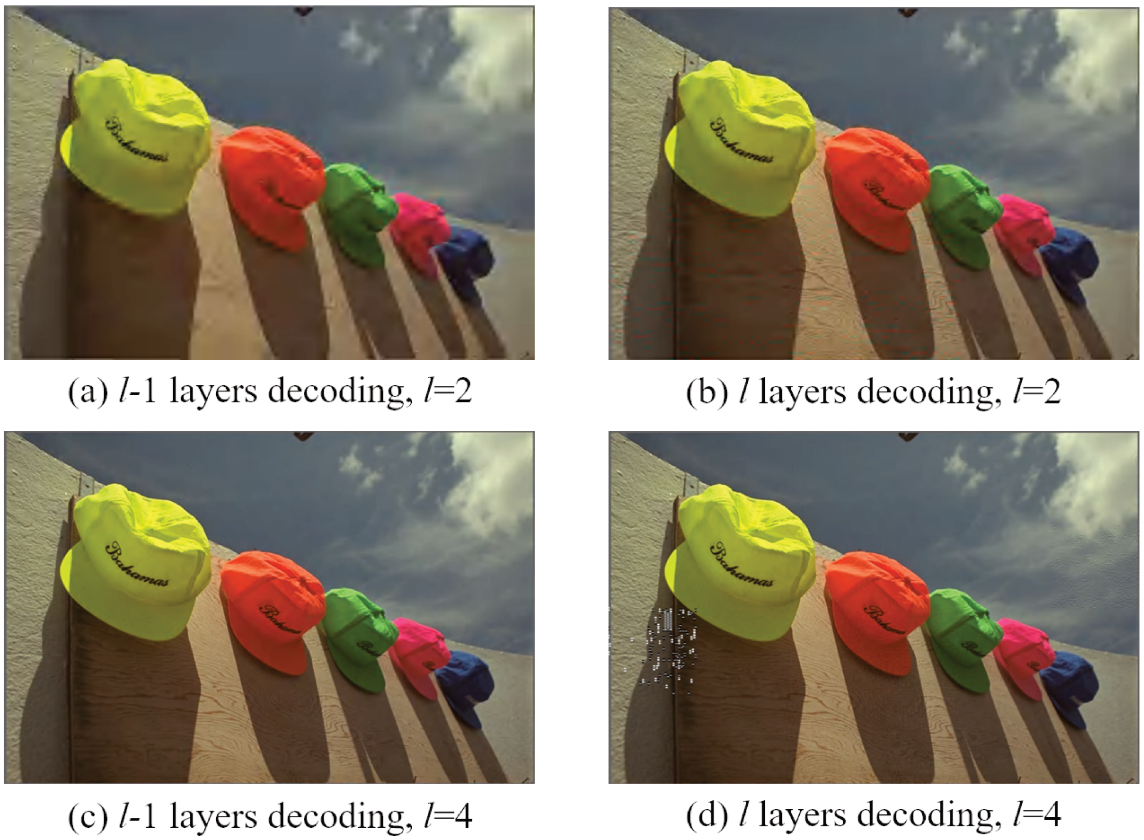
The classic decoding strategy utilizes l layers (l varies with channel condition) of decoding configuration to obtain a scalable image quality providing a high flexibility under an unstable channel condition. The codec system with the UPA strategy and the content-based hierarchical transmission function has demonstrated good performance in the optimization of QoS [1, 3]. However, it only considers specific communication parameters, such as Signal Noise Ratio (SNR), BER and channel capacity to optimize transmission and obtain a good

QoS, but does not consider the QoE of the terminal user. For the HVS, the high QoS, in a certain case, does not mean a high-quality experience, as shown in Fig. 4.2, where (b) and (d) have a better QoS but do not have a better QoE. The reason is because the resource allocation is made under the constraint of a target BER according to the channel state checked before the transmission. Nevertheless, the channel state may turn bad due to the mobility of the terminal and the change of channel state lead to the highest-indexed subchannel has not enough power and thus has not enough protection to reach the target BER, after allocation of the power and data rate on the lower-indexed subchannels. Moreover, at the receiver side, the robust JPWL decoder does not prevent the decoding of the highest-indexed layer since the code stream errors are limited and the content of the layer is helpful to the image quality improvement. Actually, for the terminal users, the layer with a few errors may lead to degradation or improvement for image quality, as shown in Fig 4.3 (a) and (c) respectively. Thus, in some cases, an l-layer decoding configuration may not be an optimal selection, but an $(l-1)$-layer decoding configuration may be.

The influences on images of two different decoding configurations are shown in Fig. 4.3, where (a) and (c) are decoded with $l-1$ layers, namely partial-decoding, (b) and (d) are decoded with l layers, namely full-decoding (the value of l is determined by the state of the channel). Generally, the image decoded with l layers has a better quality than the one decoded with $l-1$ layers since l is the number of quality layers and more quality layers signify higher quality, referring to (a) and (b). However, in some cases, the image decoded with $l-1$ layers has a better quality than that decoded with l layers, as illustrated in (c) and (d). Since the l-th layer is attacked by channel noises and parasitic oscillations are caused by false coefficients during the reconstruction.

Thus, a new decoding strategy to improve the QoE while ensuring the QoS is desirable. The solution is to exploit an image quality metric to evaluate the image quality and to determine the decoding configuration automatically. The most appropriate IQA method is an RR-IQA method since an RR-IQA method (referring to Chapter 3) provides a good balance between the prediction accuracy and the size of the required reference information. The new RR-IQA metric FMRP proposed in Chapter 3 will be embedded in the codec system to evaluate the image quality and select the decoding configuration automatically.

To optimize the decoding strategy according with the HVS, we need a database to collect people's preference when selecting two images decoded by different decoding configurations. During the last few years, some representative image databases like LIVE, TID2013, CSIQ and Toyama have been built to evaluate IQA metrics. These databases contain numbers of images and large physiological vision test, and provide a reliable benchmark for assessing IQA algorithms. However, these databases are not suitable to assess whether the FMRP



(a) l-1 layers decoding, l=2 (b) l layers decoding, l=2

(c) l-1 layers decoding, l=4 (d) l layers decoding, l=4

Fig. 4.3 The quality comparison between the $l-1$ layers and the l layers decoding images, where (a) decoded with $l-1$ layers and (b) decoded with l layers ($l=2$, because of the bad channel condition), and (b) has a better quality than (a). Figure (c) decoded with $l-1$ layers and (d) decoded with l layers ($l=4$, because of the good channel condition), but (d) has a worse quality than (c).

method can optimize the image quality by selecting the decoding configuration in wireless communication systems for multimedia contents. The reason is manifested in two aspects: (1) All of these databases, except LIVE and TID2013, did not contain JPEG 2000 transmission error distortion, which is an often-encountered distortion in communication systems utilizing JPEG 2000 coder. (2) All these databases were built under the simulated environment, namely, the degradative images were distorted not by the realistic channel noises but by the artificial noises. Thus, we have to build a new database based on a realistic error-prone environment. The purpose is to obtain the people's preference when selecting two images decoded by different decoding configurations.

4.2 Global scheme and classifier deployment

4.2.1 Global scheme

The RR-IQA metric FMRP proposed in Chapter 3 is very suitable to embed into the wireless communication system because of its sensitivity to various distortions and less binary cost for RR information. The RR features of FMRP metric are extracted from a three-scale six-orientation wavelet decomposition of the image, including 18 magnitude and 18 relative phase features. The KLD between sub-bands of the reference image and the distortion one is employed to measure the image quality degradation, and the objective score is derived from the KLDs by a GRNN.

The features of the FMRP metric are applied in a realistic wireless communication system to improve the output image quality by determining the optimal decoding configuration. The global scheme considering a hierarchical content and the users' visual perception is illustrated in Fig. 4.4 (as recognized by the Long-Term Evolution standard [86]), where the system comprises hierarchical JPWL coding section, precoder solutions, MIMO transmission channel, and a JPWL decoder. The JPWL coding section divides the image codestream into hierarchical layers with different contributions for image quality, according to the multi-scale characteristic of wavelet decomposition. Moreover, this system exploits a CL-MIMO technology related to the CSI at the transmitter terminal to divide the whole channel into independent and hierarchical SISO sub-channels [2]. The scheme allows the precoder algorithm to match the hierarchical power protection and the system parameters of each subchannel with the UPA and the UEP strategies. The novelty of the proposition is utilizing the new RR-IQA metric to optimize the decoding configuration, and achieving the purpose of improving users' QoE while ensuring the system's QoS.

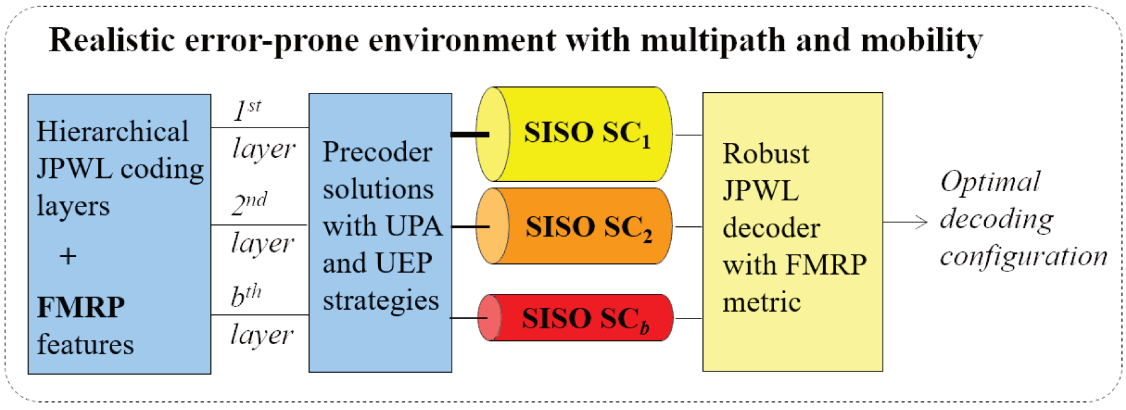


Fig. 4.4 The global scheme of realistic MIMO wireless channel.

For the realistic error-prone environment, we utilize a 3D-ray tracer validated by experimental comparison [2, 20]. To improve the robustness of the system, the JPWL coder decomposes the image content into b hierarchical layers and extracts the FMRP features (the crucial parameters of models) of the whole image. Corresponding to the b hierarchical layers, the system generates b code-streams sorted in a descending magnitude order, and the precoder solutions decouple an MIMO channel into hierarchical, parallel and independent sub-channels of different magnitudes and sorted in a descending SNR order. Therefore, the image can be transmitted by the hierarchical layers (1st layer by 1st sub-channel, 2nd layer by 2nd sub-channel, etc.) and a quality-variable version of the transmitted image will be obtained even under channel perturbation. The UPA and UEP strategies are applied to guarantee partial or total reception of the image depending on the channel state [81]. During the JPWL coding process, a small amount of data, extracted from the original image and called Reduced-reference (RR) is embedded in the code-stream by the image quality assessment system FMRP. At the decoding side, the system is jointly used with a robust JPWL decoder [1] to provide the best decoding configuration to the user by exploiting the embedded information. To ensure the embedded information is not distorted in the transmission, the Error Protect Block (EPB) [33] is used in the JPWL decoder.

4.2.2 Classifier deployment

To decide the decoding configuration automatically, a classifier is necessary, which can select a better-quality image in the same manner with users. The input of the classifier can be the objective score or the crucial features of image quality metric. For the classic quality metrics, the objective score is obtained by a nonlinear mapping through a logistic function. Since the logistic function is a single-input and single-output function, the multiple features of the metric usually are integrated into a quality index in a fixed manner. For machine-learning-based metrics, the mapping relationship is determined by machine learning tools such as the neural network and support vector regression. The multiple features of the metric are transformed to an objective score, directly, by machine learning tools. Moreover, according to different databases, the mapping relation can be adjusted to suit different applications to achieve more accurate predictions. To test the second strategy that allows one to tune according to different applications, the multiple features of the metric are used as the input of the classifier.

The classifier trained by the database is utilized to automatically select the optimal decoding configuration in line with the HVS, as illustrated in Fig. 4.5. The output of the JPWL decoder is two images decoded with l and $l-1$ layers and the RR features of FMRP metric for the original image. Then, the features of FMRP metric for two images are extracted

and compared with the RR of the orginial image. Finally, according to the training of the new database, the classifier selects the image with a better quality as the optimal decoding output and determines the optimal decoding configuration (l or $l-1$ layers).

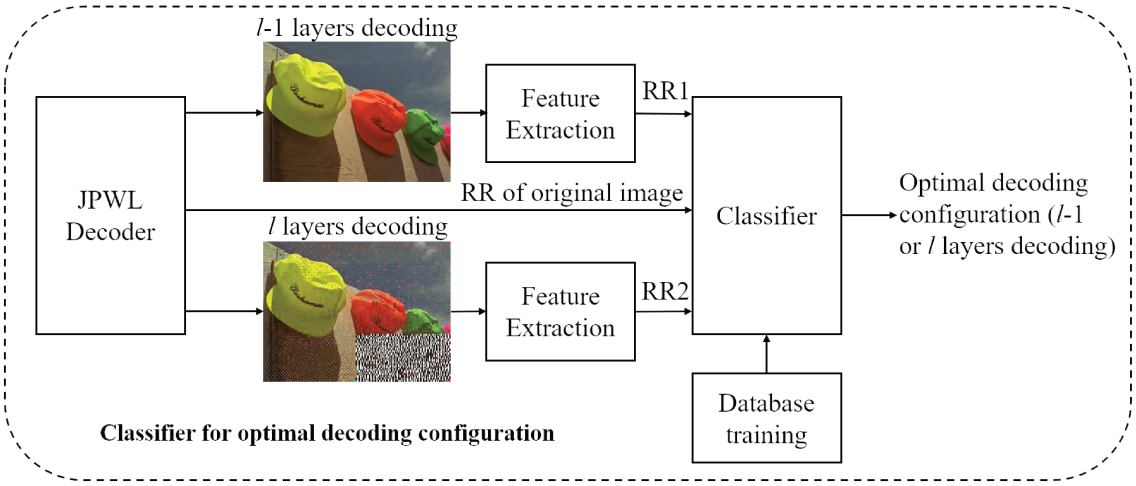


Fig. 4.5 Optimization of decoding configuration with the classifier.

4.3 Database construction

The details of the database establishment are introduced in this section, including reference and distorted images, testing methodology and environment, display configuration and equipment, and assessors.

4.3.1 Reference and distorted images

4.3.1.1 Reference and distorted images

The reference images named "House", "Caps", "Bikes" and "Monarch" are selected from the LIVE database. These 4 images with representative features have different scenes, details, colors and contrasts, as shown in Fig. 4.6. The image "Caps" has a bright color and a simple background. In opposite, the image "House" has more details but dim colors. A clear contrast between a blurred background and a detailed foreground is provided by the image "Monarch". The image "Bikes" has a bright contrast and rich contours.

All the distorted images are derived from the reference images through wireless communication system over a realistic wireless channel. To cope with the channel instability, the wireless communication system coded the image into 1 to 4 quality hierarchies and transmitted them with the UPA and the UEP techniques. At the receiving terminal, the system will



Fig. 4.6 The reference images used in the database.

select the appropriate number of layers ($l = 1, ..., 4$) as the decoding configuration. However, the selection is only related to the parameters of the transmission system such as BER and SNR, without involving the perception of HVS. To compare the quality of the images decoded with l and $l-1$ layers, we collected two types of images at different receiving points. The realistic suburban environment is the campus of University of Poitiers in France. In the simulation, buildings are represented in red (see Fig. 4.7 (a)). The MIMO (4×4, due to the rich characteristic of this wireless channel) transmitter is fixed and the MIMO receiver moves at a speed of 5m/s over a distance of 180 meters. We alternate successively bad (Non Line Of Sight (NLOS) in the area 1), moderately good (NLOS in areas 2 and 4) and good (Line Of Sight (LOS) in the area 3) areas. The evolution of gain of MIMO channel is presented in the Fig. 4.7 (b). The gains at every positions of the MIMO receiver are obtained by means of a simulator of 3D rays [20]. Thus, different levels of distortions will occur in the images at different receiving locations. These distorted images will be submitted to subjective test with the aim of identifying the quality of both images (total and partial decoding).

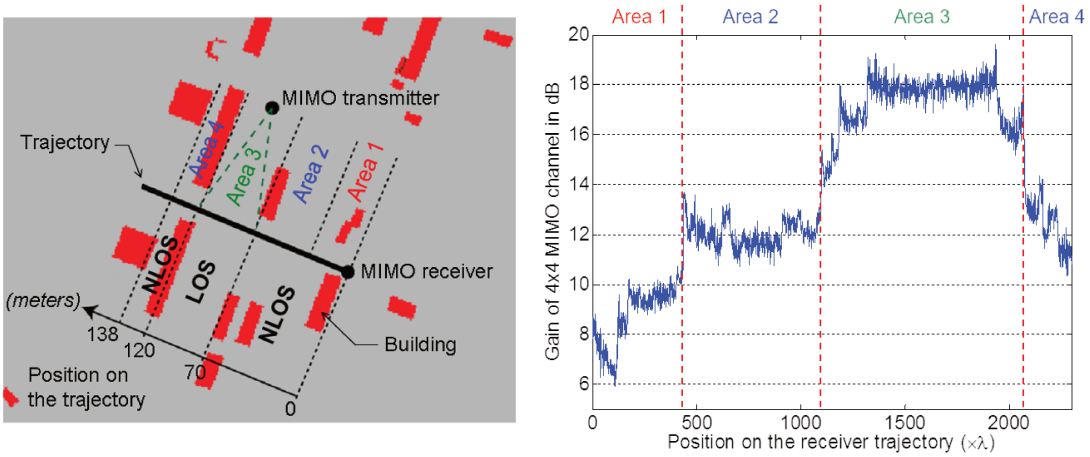


Fig. 4.7 The topology of the experimental environment and the gain variation of the system: (left) Topology of the transmission scene and (right) gain evolution of the MIMO channel.

4.3.1.2 Distortion types

The two images suffer JPEG 2000 compression distortion and hybrid distortion of JPEG 2000 compression and JPEG 2000 wireless transmission errors, respectively. For the $(l-1)$-layer decoding image, the distortion is caused by the compression procedure in the transmitter. There are very few transmission errors because of using the EPB of JPWL decoder. For the l-layer decoding image, the distortion is due to the compression and transmission errors since the l-th layer code-stream is more easily attacked by channel noise (the higher the index, the more sensitive to the perturbations). Thus, if the l-th layer code-stream has a good quality, i.e., without large noise attack, the image decoded with l layers has a better quality than the one decoded with $l-1$ layers, and vice versa.

4.3.2 Testing methodology and environment

4.3.2.1 Display configuration and testing environment

The experimental interface is a GUI interface in Matlab as illustrated in Fig. 4.8. Two types of images are shown in a random order, and the assessors classify the images by discrete values -2, 0 and 2 representing respectively that the left one has a better quality, they have the same quality, and the right one has a better quality.

Following the Recommendation ITU-R BT.500-13, we deployed the experiment and collected the preference of the assessors. We chose 210 pairwise impaired images covering the range of all distortions for subjective test. To avoid visual fatigue and restrict the test time not more than 30 minutes, we partitioned all the images to 7 sets. Each set contains 30

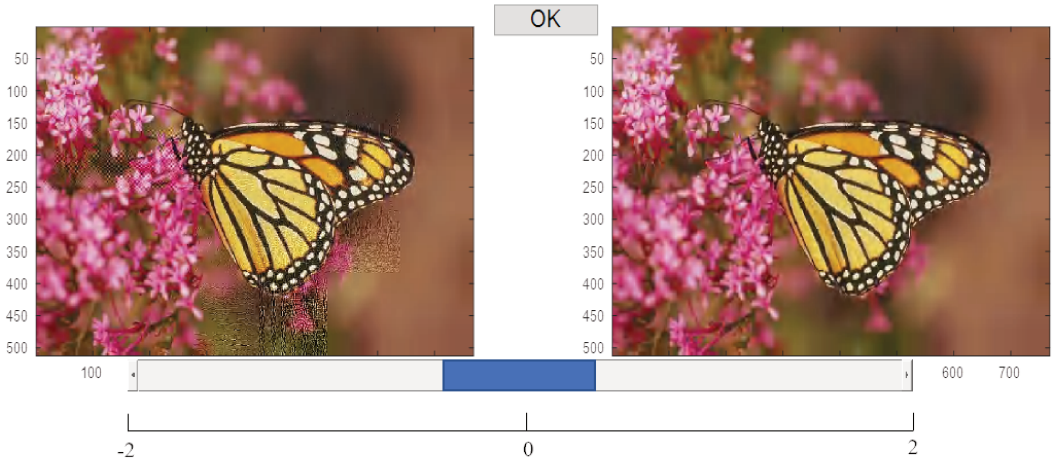


Fig. 4.8 The interface used in the experiment, $-2, 0, 2$ represent 3 classifications.

pairwise images in a random order and each assessor completes one or two sets test. The number of assessors for each set is ensured to be more than 16, as shown in Table 4.1. The workstations were placed in an office environment with normal in-door illumination levels. The display monitor was 14-inch notebook monitors displaying at a resolution of 1366×768 pixels.

Table 4.1 Subjective evaluation sets

Sets	Number of pairwise images	Number of assessors
1	30	16
2	30	15
3	30	18
4	30	16
5	30	22
6	30	22
7	30	15
Total	210	Average: 17.71

4.3.2.2 Assessors, Training and Testing

Around 100 assessors recruited respectively from 3 Universities: University of Poitiers (Ph.D. and Engineer students), Nanchang University (postgraduate students) and Jiujiang University (undergraduate students and teachers) participated in the study, over a course of half a year. Most of them have no experience with image quality assessment and image impairments. The demographic data of subjective assessors is shown in Table 4.2. The assessors were tested

for vision problems, and people with poor eyesight (corrected eyesight) were forbidden to take part in the experiment. Each assessor was individually briefed about the goal of the test and given a demonstration of the experimental procedure.

Table 4.2 The demographic data of subjective assessors

Items	Nationality			Age			Gender	
	French	Chinese	Indian	19-29	30-45	46-55	Male	Female
Percentage(%)	15.38	81.54	3.07	46	50	4	66.67	33.33

4.3.3 Post-processing and properties

4.3.3.1 Post-processing of raw data

To reduce the accidental errors, we utilized a simple algorithm to conduct outlier detection and assessor rejection, according to the suggestion given by VQEG [16]. The points satisfying the Eq. (4.1) is defined as the outlier points.

$$|c - c_{\text{mean}}| > 2\sigma, \tag{4.1}$$

where c is the classification value of the image given by an assessor, c_{mean} and σ are respectively the average classification value and the standard deviation given by all assessors for an image. For assessor rejection, we considered a similar solution used in LIVE database [17]. For any set, all quality evaluations of an assessor were rejected if more than 16.67% of his evaluations are outliers. Overall, a total of 5 assessors were rejected, and about 4.67% of the difference values were rejected as being outliers (we took all data points of rejected assessors as outliers).

After the outlier detection and assessor rejection, a majority rule is exploited to select the classification value, which most frequently occurs, as the final classification value. The classification statistics of the database is shown in Table 4.3, where for the pairwise images, Label -2, Label 0 and Label 2 respectively denote that partial-decoding image with a better quality, two images have the same quality and full-decoding image has a better quality. It is obvious that the samples with the label of 2 are very few. Since the test images were selected randomly and most partial-decoding images have the same or a better quality than the full-decoding ones.

To analyze the relation of classification label and the transmission channel states, Fig. 4.9 illustrates that the distribution of classification labels varies with the channel states. In

Table 4.3 Classified statistics of the database

Items	Total samples	Label −2	Label 0	Label 2
Number	210	160	46	4

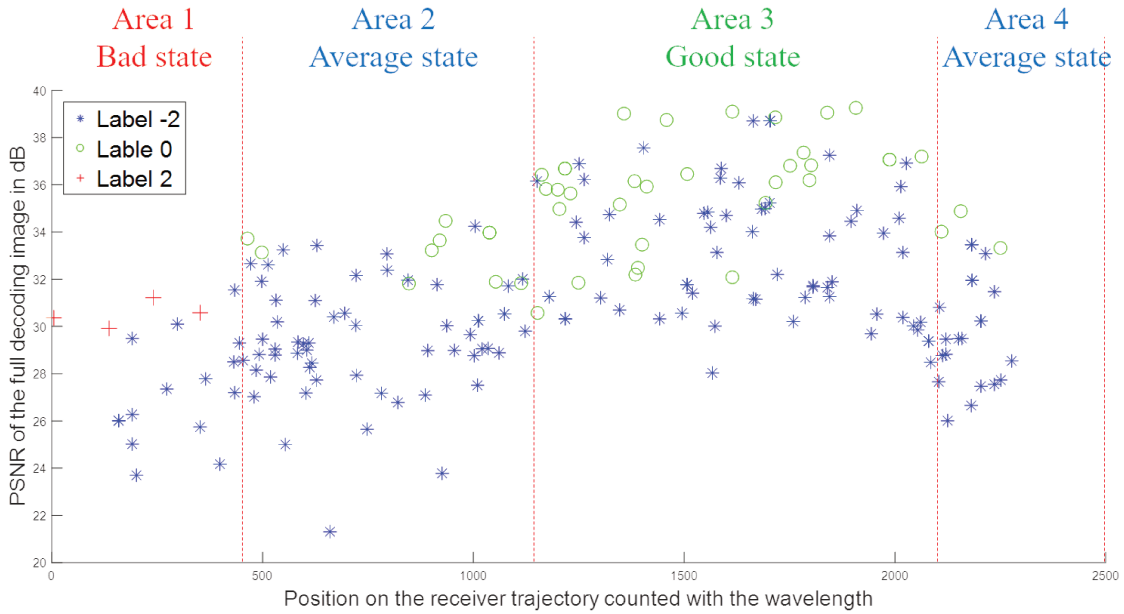


Fig. 4.9 The optimization of decoding configuration with a classifier.

area 1 (bad state), the difference between two types of images is very large and it is easy to identify which one has a better quality, where for the HVS, 4 labels equal to 2 and the rest one equal to −2. In the areas 2 and 4 (average state), most labels are equal to −2 and few labels are equal to 0. In area 3 (good state), a small part of labels are equal to 0 and the rest are equal to 2. The phenomena above indicate that the image quality transmitted in different channel states will vary significantly. Although the robust JPWL decoder can provide good QoS based on the system parameters, the QoE based on the HVS needs to be improved with IQA metric.

4.3.3.2 Properties of the new database

The presented database is based on JPWL standard over a realistic wireless channel, so it is named the JPWL-RWC database. The JPWL-RWC database contains 480 distortion images and large physiological vision test of around 100 people from three countries (France, China and India), which can provide a benchmark for optimization of decoding configuration and also provide a useful tool to understand JPEG 2000 compression distortion and JPEG

2000 wireless transmission error distortion. The new database reveals the properties of the distortions encountered in the wireless communication system, and the preference of the HVS when selecting two types of distortion images, as summarized as follows:

1. The main distortions encountered in the JPWL transmission system are JPEG 2000 Compression Distortion (JP2K-CD) and JPEG 2000 Transmission Error Distortion (JP2K-TED) as well as their mixture.

2. The JP2K-CD is manifested with blur and ringing, as shown in Fig. 4.10. The blur in an image is the loss of spatial details, which is related to the filtering or data compression. The ringing is caused by the quantization or truncation of the high-frequency transform coefficients resulting from DCT- or wavelet-based coding [62]. This will lead to oscillations or ripples around edges and contours in the image, which is also called the Gibbs phenomenon [103, 62]. The distortion can be well detected by the wavelet-based IQA metric for example, WBRRM [109] and FMRP proposed in Chapter 3.

3. The JP2K-TED is a local and random distortion since the distortion occurs in local areas and the locations of distortions are unpredictable. The distortion is a challenge to most IQA metric because of its locality and randomicity.

4. The local and random distortion resulting from the noise oscillation usually cause a large variation of the luminance in the location (the luminance oscillates greatly with the frequency of noise), as illustrated in Fig. 4.11.

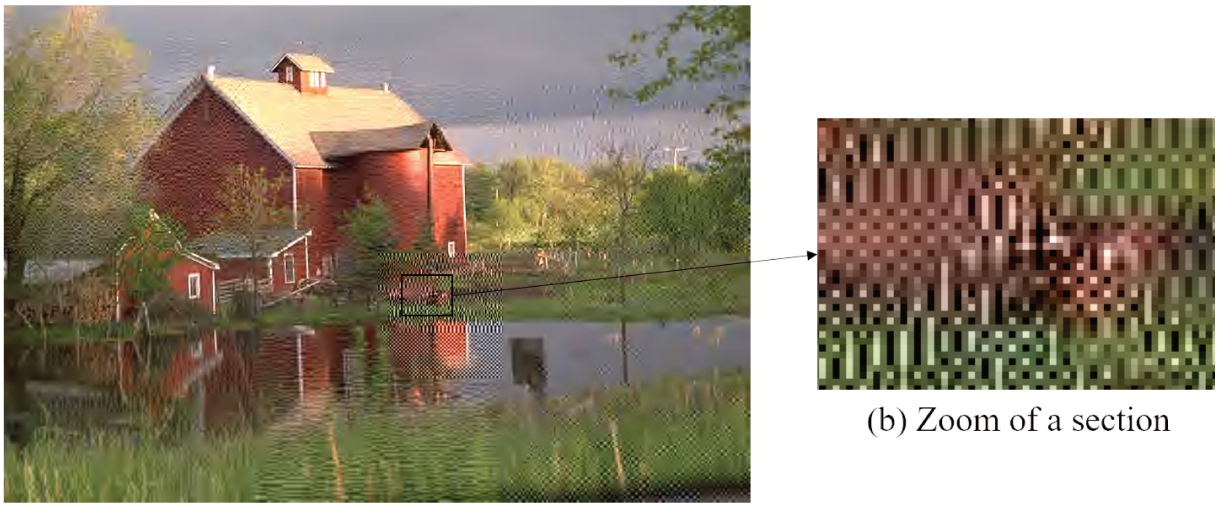
5. The HVS has a preference when comparing the image quality of two types of distortions. Most subjects prefer the uniform distributed distortion namely JP2K-CD, and are sensitive to the local and high luminance distortions, i.e., JP2K-TED.



(b) Zoom of a section

(a) Image with blur and ringing

Fig. 4.10 Image with blur and ringing distortion.



(b) Zoom of a section

(a) Image with local and random distortion

Fig. 4.11 Image with local and random distortion.

4.4 Assessment of representative metrics with database

4.4.1 Classifier design

With the database, we can evaluate the effectiveness of representative IQA metrics on distinguishing between two types of images. The representative IQA metrics comprise four FR-IQA metrics and four RR-IQA metrics, some of which are based on machine learning method. As introduced in Subsection 4.2.2, the input of classifier can be objective score or crucial features of the IQA metric. For the machine-learning-based metrics, we assume the crucial features as the input of the classifier, referring to Fig. 4.5. For the classic IQA metrics, we propose that the objective score is employed as the input of the classifier, as shown in Fig. 4.12. There is indeed another simple way to classify with the IQA scores directly. In this way, the scores of two types of images are compared and the decoding layer of the image with the high score is selected as the optimal decoding configuration. However, it is a very rough classification method without supervision and we observed that in the range of low-quality score, the people prefer the l-layer decoding image and in the range of high-quality score, the people prefer the $(l-1)$-layer decoding image. Thus, the classifier and the supervised training are necessary to take into account this effect. After the training of new database, the classifier selects the better quality image as the optimal decoding output and determines the decoding configuration (l or $l-1$ layers). Two classifiers: K-Nearest Neighbors (K-NN) and a multiclass Support Vector Machine (SVM) with the Error-Correcting Output Codes (ECOC) [29] were employed to select the better-quality image based on the input of the metric features.

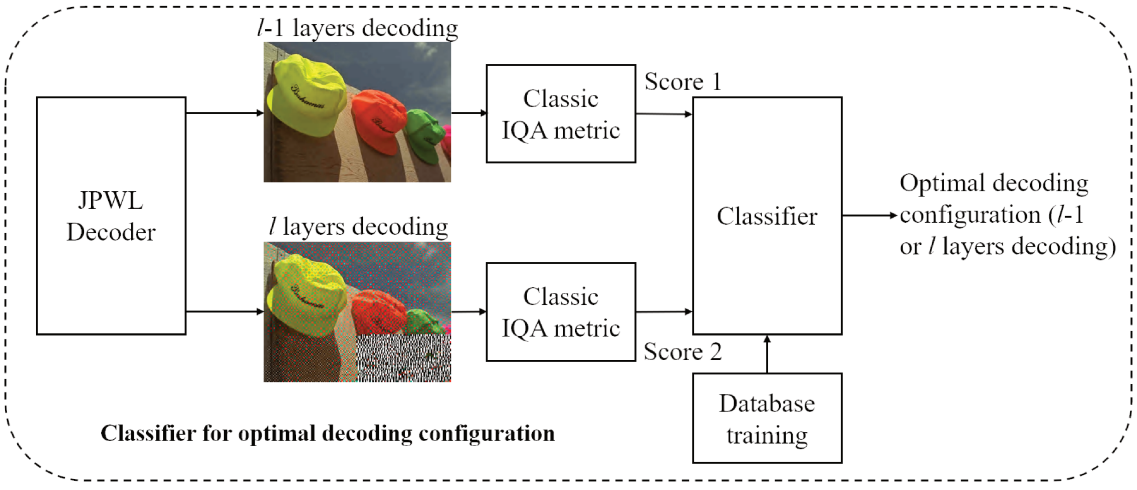


Fig. 4.12 Classifier design for classic IQA metric.

The K-NN is a simple yet popular approach for classification. For a given new example, it classifies the example into the class of the nearest training example to the observation. It is easy to implement and has 2 parameters to set. Here we set the K to 5 and choose the Euclidian distance as the basis of classification. The K-NN is a lazy algorithm without training and sensitive to noises. It is easy to implement and less time-consuming. However, this method is a non-parametric method that requires all training data to be stored for a new classification. It is inconvenient for practical application. Therefore, the classification results of the K-NN method are only used as a comparison with that of SVM.

The SVM is a binary classifier which can be extended to a multiclass classifier by using multiple classifiers. The standard approach is to learn k individual binary classifiers $c_1, ..., c_k$, one for each class (k denotes the number of classes). Here we use another advanced approach which taking an ECOC as the representation of k classes and utilize n individual binary classifiers ($n = k(k-1)/2 = 3$, $k = 3$, for error correcting). With ECOC, the SVM algorithm is robust with respect to changes in the size of the training samples, the assignment of distributed representations to particular classes and the application of overfitting avoidance techniques such as decision-tree pruning [7, 29]. Considering that the feature vector has a large dimension (36) and samples of the database are few (210 in total), we select the Radial Basis Function (RBF) as the kernel function of SVM.

For the design of machine learning set, we employ a 5-fold cross-validation style. The database is randomly partitioned into 5 equally sized subgroups. A single sub-group is selected as the test data, and the remaining 4 subgroups are used as training data, then switching the testing and the training data. Each of the 5 subgroups is selected just once as the validation (test) data.

4.4.2 Effectiveness validation

To form a comprehensive evaluation, 7 other representative image quality metrics were used, including 4 FR metrics: PSNR, SSIM [1], MS_SSIM [2] and FSIM [3], and three RR metrics: Wang Z. et al.'s RR method [14] (we call it WBRRM), ADI [15] and Lin Z. et al.'s RR method [16] (we call it RP). The implementation codes of the SSIM, MS_SSIM, Wang Z. et al.'s RR and FSIM methods are provided by the authors of [17, 18]. For the other metrics, we implemented them according to the algorithms in [15, 16, 19]. For the machine-learning-based metrics such as ADI, RP and FMRP, the crucial features of the metric serve as the input of the classifier. For the other classic metrics, the objective scores of the metric is employed as the input of the classifier.

The performance indicators are the ratio of right classified samples and overall samples, and the standard deviation. All the performances of the metrics are summarized in Table

4.4, where "Acc" denotes accuracy measured in percentage, namely the mean ratio of the right classified samples and overall samples (5-fold cross-validation), and "Std" denotes the standard deviation of 5 datasets. To increase the readability of the results, we mark the first-best and the second-best metrics in bold. One can see that the metric FMRP performs quite well for both classifiers. It provides a better prediction accuracy than any other metrics except MS_SSIM. Compared with the state-of-the-art FR metric MS_SSIM, it also performs quite well, with a higher accuracy in the SVM classifier and the comparable accuracy in the K-NN classifier. While the proposed metric FMRP is an RR metric, it uses much few data in image evaluation and is more appropriate when applying in the real-time communication system than the FR metrics.

Table 4.4 Performance of the objective quality metrics based on the proposed database

Metric		PSNR	SSIM	MS-SSIM	FSIM	WBRRM	ADI	RP	FMRP
KNN	Acc(%)	78.10	80.48	**82.71**	80.95	76.67	79.52	75.71	**81.43**
	Std	4.58	6.16	4.76	2.38	5.15	7.82	2.61	6.39
SVM	Acc(%)	75.71	77.62	**83.10**	83.05	76.19	80.95	74.76	**83.24**
	Std	0.82	1.51	0.58	0.26	0.00	1.15	1.74	1.03

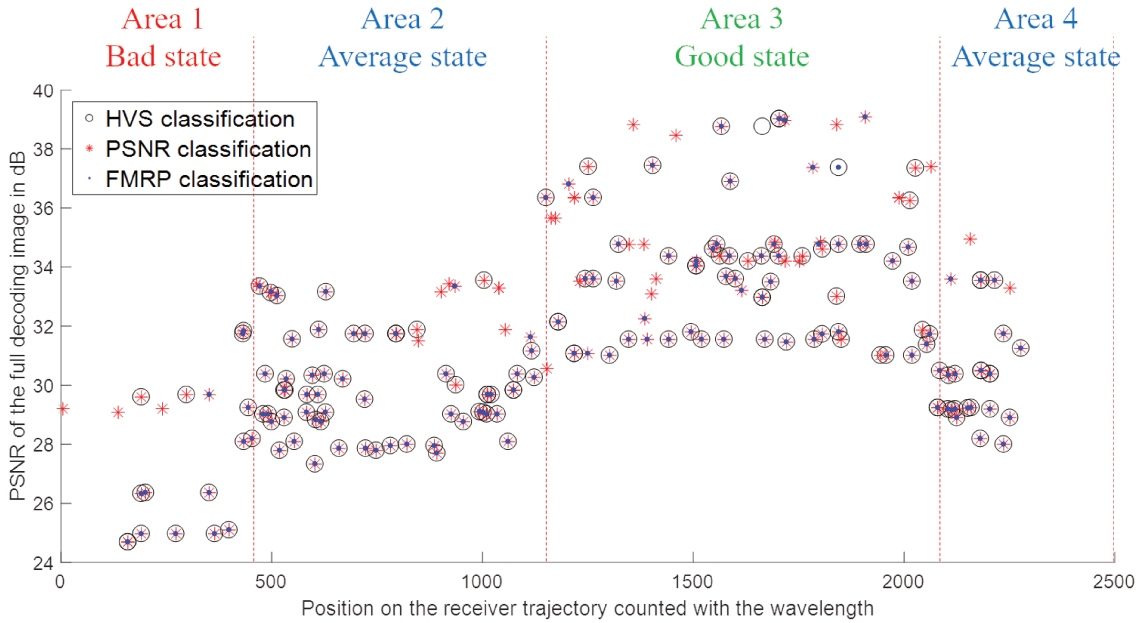


Fig. 4.13 Comparison of classification results of HVS, PSNR and FMRP, where all the points correspond to the classification label -2 and the number of isolated points indicates the degree of difference.

To show the classification performance more intuitively, classification results based on HVS, PSNR and FMRP are shown in Fig. 4.13, where all the points correspond to classification label -2. Noted that the number of points of HVS, PSNR and FMRP classifications are different from each other and they are 160, 204 and 158, respectively. Moreover, the isolated points show the difference and the coincide points show the consistency. Fig. 4.13 intuitively shows that the isolated points of the PSNR classification are much more than that of the FMRP classification. Thus, the classification result of the FMRP metric is closer to the classification result of the HVS, as compared with that of the PSNR metric.

After the comprehensive evaluation, one can see that the FMRP metric has a good performance to classify the images in consistent with the HVS. The performance in optimizing decoding configuration needs to be evaluated now. Since there are three classification labels and two decoding configurations, the three labels should be converted to two categories. Label -2 should be assigned to $(l-1)$-layer configuration and label 2 should be assigned to l-layer configuration. The point labelled with 0 means the two images have the same quality and denotes the optimal configuration is either $(l-1)$–layer configuration or l-layer configuration. If the two images are in the same quality and it is sure that l-layer configuration has more details than $(l-1)$-layer configuration, we assign the label 0 to l-layer configuration.

Table 4.5 shows the performances of four decoding strategies including classic strategy (l-layer decoding configuration), strategy with PSNR metric, QIP metric [66] and FMRP metric. The performances are measured with a ratio between the samples consistent with the HVS classification and the total samples. The QIP metric is an RR-IQA metric based on the points of interests and the objects' saliency of color images [66], which utilizes the change of numbers of interest points in different salient areas to measure the image quality. Based on the QIP metric, J. Abot et al. proposed an optimization scheme of decoding configuration over a realist MIMO channel [1]. The scheme has been proved to have a good improvement for image quality but the performance validation is measured by the PSNR metric, not the real perception of users, and the validation of QIP effectiveness is only based on the JPEG compression and JPEG 2000 compression distortions. Therefore, to verify the effectiveness of the decoding strategy suitable for the HVS, we proposed to construct the JPWL-RWC database and present the FMRP metric. With the new JPWL-RWC database, the comprehensive validation based on the real perception of users is carried out. Compared with the classic strategy, all the three strategies based on IQA metric can improve the experience of image quality and this indicates the new decoding strategies with IQA metric is necessary. Further, from the comparison, one can see that the decoding strategy with FMRP metric performs best for all channel states. The improvement of all strategies for good channel state

is lower than that for other two states. Thus, the qualities of two images are good and very close, and this necessarily cause difficulty for IQA to detect the difference.

Table 4.5 Performances of decoding strategies (ratio of correct classification samples based on the HVS)

	Decoding strategy			
Channel state	Classic(%)	With PSNR(%)	With QIP(%)	With FMRP(%)
Bad state	18.21	78.95	78.95	**84.21**
Average state	15.54	84.95	86.02	**90.32**
Good state	32.65	67.35	75.51	**77.51**

4.4.3 Inaccurate samples analysis

Although the classifiers with FMRP metric can provide the right classification for most of the samples, there are about 6 or 7 misclassified samples for each subset (42 samples) of training-testing procedure (5-fold cross-validation style introduced before). These misclassified samples are associated with local random distortions. The samples are very difficult to right-classify by the objective IQA metric since the same distortion in different areas may cause different perceptions of the HVS. As shown in Fig. 4.14, there are very slight distortions in (b) and (d) but the distortion areas are different, where the distortions of (b) occur in the apparent area and the distortions of (d) occur in the non-significant area. Although the distortions of images are very similar, the classification labels of the HVS are completely different, where the classification label of the first pair of images is -2, and for the second pair of images, the classification label is 0. Anyway, for these samples, the distortion is very slight and if in the image sequence or video, it is difficult for the HVS to detect the distortions. In the next section, we propose to detect the local random distortions effectively and design a new RR-IQA metric with other features.



(a) Image with better quality (b) Image with slight distortion in notable area



(c) Image with better quality (d) Image with slight distortion in dim area

Fig. 4.14 Comparison of two misclassified samples, where for the HVS, (a) has a better quality than (b), and (c) and (d) have the same quality; for the FMRP metric, (a) and (b) have the same quality, and (c) has a better quality than (d).

4.5 A new metric FMRP-LSD

In this section, we try to exploit new features sensitive to local random distortion and integrate them into FMRP metric to improve the performance of detecting local distortions.

4.5.1 FMRP feature and LSD feature

As introduced in the previous section, the JP2K-TED distortion is a local random distortion and a challenge for most of IQA metrics. Moreover, the FMRP metric is a global statistical metric based on KLD differences between the DT-CWT coefficients of reference and distortion images. Although the FMRP metric performs best in the classification of images with different decoding layers, the accuracy needs to be improved further. For the local- and random-distortion images, especially for the images with obvious distortions only in local small areas, the FMRP metric does not perform well. As illustrated in Fig 4.15, (a) and (b) are images respectively decoded with 3 layers and 4 layers. Obviously, (b) has a worse quality than (a) since an obvious distortion occurrs in a local small area. As mentioned before, the FMRP metric measures the distortion with a global statistical feature, i.e., KLD between

PDFs of reference image and distortion image. As shown in Fig. 4.15 (c), (d), (e) and (f), the KLDs of magnitude and relative phase of two images are very close. The numerical comparison of KLDs is demonstrated in Table 4.6, where KLDs of (a) are larger than that of (b). Since the image (a) loses a quality layer compared with the image (b), most of the pixels in the former have a larger variation than that in the latter. For the global statistical measure FMRP, the KLD value of (a) is larger than that of (b), although some pixels in the partial area of (b) have a larger variation than that of (a). Thus, the FMRP feature cannot detect the local distortions well and a new feature to measure local distortions should be developed.

Table 4.6 KLDs of six sub-bands of images (a) and (b)

	Magnitude		Relative phase	
Sub-bands	image (a)	image (b)	image (a)	image (b)
1	0.0514	0.0110	0.0178	0,0236
2	0.0607	0.0195	0.0350	0.0297
3	0.1081	0.0301	0.0111	0.0178
4	0.0525	0.0102	0.0256	0.0283
5	0.0823	0.0195	0.0466	0.0389
6	0.1916	0.0338	0.0179	0.0254

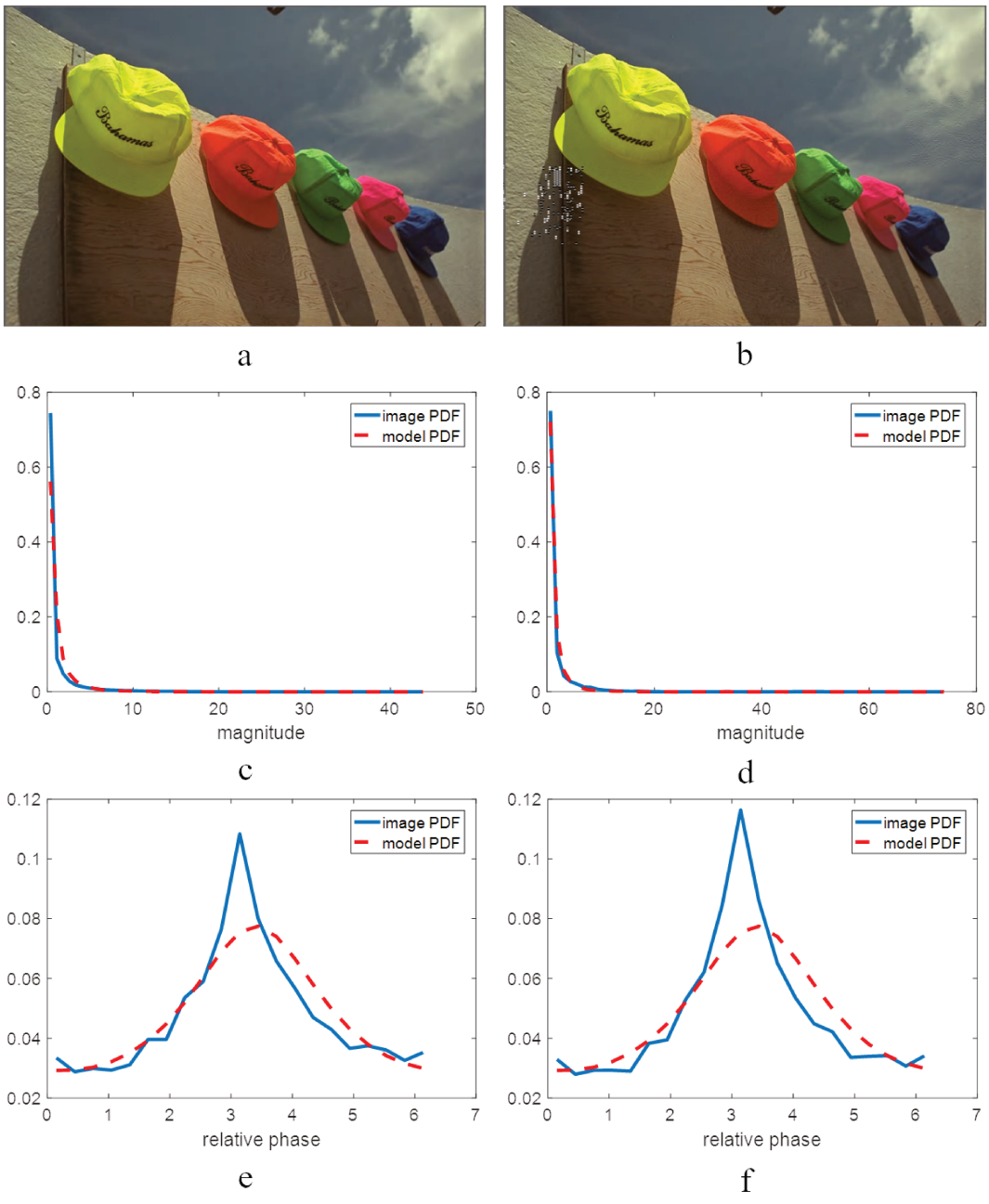


Fig. 4.15 Comparison of two images with different decoding configurations: (a) and (b) are respectively decoded with 3 and 4 layers configuration, (c) and (d) are the comparison of magnitude's PDFs between the model and the image, (e) and (f) are the comparison of relative phase's PDFs between the model and the image.

As mentioned earlier, it is necessary to introduce the local measure. Since the local features in the CWT domain will cost a large transmission binary bits, it is difficult to deploy them in communication systems. Then, after a study of literature, we found that the Local Standard Deviation (LSD) map in the spatial domain is sensitive to the local distortion. The LSD map is a matrix consisting of the standard deviation of image sub-blocks, which has been used in many image applications such as contrast enhancement [19], noise estimation [36] and image quality evaluation [37, 118]. In [37], Gore et al. applied a local standard deviation (LSD) map to image quality estimation for JPEG and JPEG 2000 compression, but the method needs a full access to the original image, which usually is impossible for the communication system. Zhang et al. proposed an RR-IQA method which reduces reference information by calculating the LSD of each block of the original image of a 12×12 size [118]. However, the method still involves a 4.88% binary cost of the original image and the RR size will increase with the image size. Therefore, we try to find a briefer way to apply the LSD map into image quality estimation. In this section, the local distortions will be coped with a LSD map in the spatial domain since the LSD map in the wavelet domain will cost a lot for transmission (18 sub-bands with real and imaginary parts). The local and random distortions mean a large change of luminance in the local areas caused by the parasitic oscillations, as illustrated in Fig. 4.11 in Section 4.3. Thus, the local standard deviation in a neighboring area can indicate the distortion well. Considering the FMRP feature is sensitive to the global distortions and LSD feature is sensitive to the local distortions, we combine these two features to improve the accuracy in both regression and classification applications in the thesis.

The scheme of the new metric is shown in Fig. 4.16. As it combines the FMRP feature with the LSD feature, we name it the FMRP-LSD metric. The FMRP features are extracted from the DT-CWT decomposition of the reference and the distortion images by the same method in Chapter 3. The LSD map consists of the local standard deviation of blocks, as shown in Fig. 4.17. The local feature, i.e., LSD feature, is computed from the linear correlation coefficients between the LSD map of the reference and the distortion images. With the FMRP features and the LSD ones, the FMRP-LSD metric can well detect the JP2K-CD and JP2K-TED distortions frequently encountered in the wireless channel. According to different application scenarios, regression or classification algorithm will be utilized to obtain the mapping relation or the classification basis and then objective scores or classification labels will be obtained.

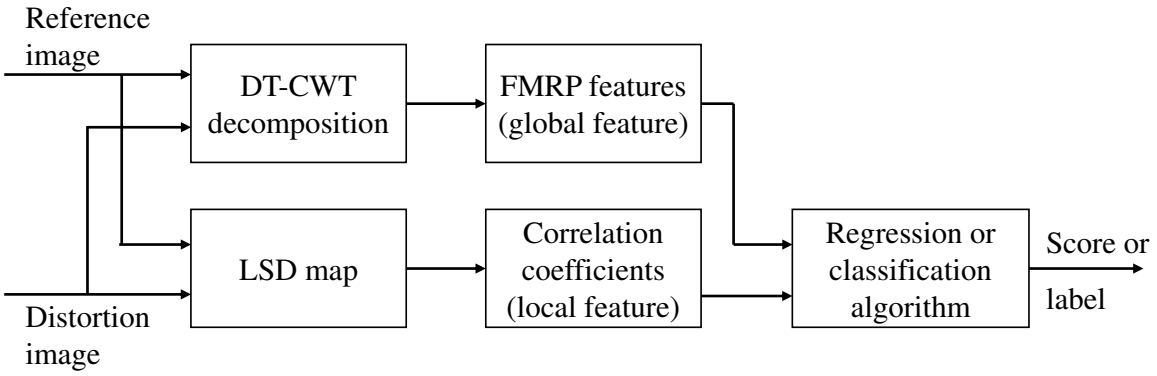


Fig. 4.16 Block diagram of new metric.

According to the size of most local distortions and the binary cost for transmission, we partition the image into 12 blocks and calculate the local standard deviation of luminance value for each block, as shown in Fig. 5.3. Different from the method adopted in [37, 118], we proposed to measure the local distortion with the Pearson linear correlation coefficient (PLCC) of the LSD of all blocks since the FMRP features already have good ability to detect the overall distortion in our metric. The PLCC of the LSD map of the reference and the distortion images can well detect the local distortions within the blocks and the structure change of the overall image. Formally, the PLCC of blocks is defined as

$$\mathrm{CC} = \frac{\sum_{i=1}^{N}(A_i-\mu_A)(B_i-\mu_B)}{\sqrt{\sum_{i=1}^{N}(A_i-\mu_A)^2}\sqrt{\sum_{i=1}^{N}(B_i-\mu_B)^2}}, \tag{4.2}$$

where N is the number of blocks, set as 12, A_i, B_i are block standard deviation vector of reference image and distortion one, respectively, and μ_A, μ_B are the mean values of A_i, B_i.

As a feature juxtaposed the FMRP feature (KLDs between the wavelet sub-bands of the reference and the distortion images), CC should be normalized in the same way as the KLD. Thus, the difference of CC is utilized to measure the degradation of the image quality, rather than using CC directly. Formally, the difference of CC is defined as

$$\mathrm{DCC} = \log_{10}[k(1-\mathrm{CC})+1] \tag{4.3}$$

where k is a constant to adjust the scope of the feature.

The new feature DCC can detect the local and the random distortions very well. Fig. 4.18 illustrates some pairs of images with different KLD, DCC and S values, where KLD is the mean of all sub-bands (after the normalization of Eq. (3.12)), DCC is the new feature for detecting local distortions, the larger the values of KLD and DCC, the worse the image



a. Image blocking method

Std 1	Std 2	Std 3	Std 4
Std 5	Std 6	Std 7	Std 8
Std 9	Std 10	Std 11	Std 12

b. LSD map

Fig. 4.17 Image blocking method and LSD map, where "Std 1" denotes the standard deviation of luminance value for block 1.

quality, S is a discrete score with two values: 0 and 1. If $S = 0$, the assessor thinks the image is of a better quality. Obviously, DCC can significantly reflect the local distortions, and its variation fits the subjective value S. Thus, the DCC feature is exploited as a supplement to KLD feature to cope with local distortion. The new metric is named FMRP-LSD since it includes the FMRP feature and the new LSD feature. Its performance on optimizing the decoding strategy will be validated in the next section.



Fig. 4.18 Comparison of three features for classification, where KLD is the mean of all sub-bands, DCC is the new feature for detecting local distortion, and S is the preference score of HVS.

4.5.2 Transmission binary cost

As an RR-IQA method used in wireless communication system, the binary cost of transmission is also an important performance indicator. In Chapter 3, we roughly assign 8-bit for quantization of each parameter and the total quantization cost for the FMRP features is $36 \times 3 \times 8 = 864$ bits. In this subsection, we will determine the quantization bits for both FMRP and FMRP-LSD metrics by a numerical method.

4.5.2.1 Relation between feature reduction and accuracy

In Chapter 3, we proposed an RR-IQA method named FMRP, which comprises 36 features extracted from the magnitude and the relative phase of the DT-CWT. The image is decomposed by a 3-scale 6-direction DT-CWT and thus there are 18 wavelet sub-bands with magnitude and relative phase features after the decomposition. We have demonstrated that the 3-scale decomposition obtains the best trade-off between the accuracy and the size of the transmitted information, referring to Subsection 3.7.3. In this subsection, we mainly analyze the influence of directions on accuracy.

First, the experiments are based on the LIVE database 2 to analyze the relation between the accuracy and the feature dimension of the FMRP metric. The feature dimension is reduced gradually from 36 to 6 to observe its influence on accuracy. Since the feature dimension is determined by the wavelet direction, the variation in the number of directions is also listed, as shown in Table 4.7, where the direction 1,2,3,4,5,6 respectively denote the wavelet sub-bands $-75°$, $-15°$, $-45°$, $75°$, $15°$, $45°$. The performance indicators, PLCC, SROCC and OR, described in Chapter 3 are employed to evaluate the accuracy. From Table 5.2, one can see that the accuracy decreases slowly as the dimension decreases but even with only one direction left, there is still relatively high accuracy. It is because the distortions in LIVE database 2 are isotropic, that is to say, most distortions are global distortions. Therefore, the reduction in the number of directions does not result in a significant drop in accuracy. In this case, the dimension of the FMRP metric can be reduced. To make a balance between the accuracy and the dimension of features, the best strategy is to keep 3 directions namely 18 features, for example, directions 1,2,3, or 4,5,6 or 1,3,5 in Table 5.2. However, if only the accuracy is considered, all directions (namely 36 features) should be retained.

Table 4.7 Variation of accuracy with feature dimensions

Directions	Feature number	PLCC	SROCC	OR
1,2,3,4,5,6	36	0.9751	0.9643	0.0025
2,3,4,5,6	30	0.9720	0.9631	0.0000
3,4,5,6	24	0.9689	0.9608	0.0000
1,2,3	18	0.9657	0.9557	0.0051
4,5,6	18	0.9654	0.9548	0.0051
1,3,5	18	0.9665	0.9525	0.0051
1,2	12	0.9512	0.9442	0.0102
1,4	12	0.9432	0.9299	0.0102
1,5	12	0.9510	0.9467	0.0102
1	6	0.9360	0.9178	0.0102
2	6	0.9364	0.9206	0.0102
3	6	0.9346	0.9128	0.0102
4	6	0.9328	0.9246	0.0102
5	6	0.9356	0.9212	0.0102
6	6	0.9319	0.9044	0.0153

Table 4.8 Classification accuracy varies with feature dimensions

Directions	Feature number	Classification accuracy(%)
1,2,3,4,5,6	36	83.24
2,3,4,5,6	30	82.38
3,4,5,6	24	80.48
1,2,3	18	79.33
4,5,6	18	80.38
1,3,5	18	81.90
1,2	12	81.88
1,4	12	78.38
1,5	12	81.62
1	6	80.48
2	6	78.57
3	6	76.57
4	6	80.29
5	6	81.52
6	6	75.24

Second, based on the JPWL-RWC database, the effect of dimension reduction of the FMRP metric on classification accuracy will be investigated. The similar procedure used in the previous experiment is employed to reduce the FMRP metric's dimensions or directions. The performance indicator is the classification accuracy namely the rate between the right classification labels and total labels. The experimental results are shown in Table 4.8. Unlike the results in Table 5.2, the accuracy decreases quickly as the dimension decreases, and the classification accuracies of all directions are different from each other. For example, the accuracy of direction 1 is 80.48% but that of direction 6 is only 75.24%. The local and random distortion of JP2K-TED in JPWL-RWC database cause this. The JP2K-TED distortions are anisotropic and thus the wavelet sub-bands in different directions have different accuracies. To detect the local and random distortions in different directions more comprehensively, all the 36 features of FMRP metric are kept. Next, the quantization bits cost will be discussed.

4.5.2.2 Quantization bits of parameters

The FMRP-LSD metric comprises FMRP and LSD features. The FMRP feature is the relevant parameters of the magnitude and the relative phase model, and the LSD feature includes a local standard deviation map with 12 blocks. As summarized in Subsection 3.4.3, the IGD model of the magnitude has two parameters, i.e., the sharp parameter λ and the mean μ_v. The WCD model of the relative phase has the sharp parameter ρ and the mean μ_w. In addition, including the modeling errors of the two models, thereby the total parameters are six. To determine the quantization bits for each parameter, the quantization error should be considered. The quantization error is decided by the quantization interval, that is, the smaller the quantization interval, the less the error. However, the extra binary transmission will increase and thus the balance between the quantization bits and the quantization error should be taken into account.

The balance is determined as follows:

1. Taking all reference images of the LIVE database 2 as research objects and listing maximal, minimal and median values of all parameters, as shown in Table 4.9.

2. The median value of KLD error is selected as the threshold to determine whether the parameter variation will cause many errors or not. Since the median value of KLD errors of the magnitude and the relative phase are respectively 0.0086 and 0.0016, the error within these values is not supposed to make a large error for the modeling.

3. We change one parameter in turn while fix other parameters, observing whether the relevant parameter variation will affect the model error apparently or not. If

Table 4.9 Determination of quantization bits

Feature	Parameter	Range			Quantization step			
		Max	Median	Min	10 bits	8 bits	6 bits	4 bits
Magnitude	λ	38.5058	2.2413	0.2124	0.0376	0.1502	0.6078	2.5529
	μ_v	40.6627	5.2629	0.5440	0.0392	0.1573	0.6368	2.6746
	KLD error	0.1586	0.0086	0.0000	0.0002	0.0006	0.0025	0.0100
Relative phase	ρ	0.4715	0.2855	0.0683	0.0004	0.0016	0.0064	0.0269
	μ_w	6.6221	3.1708	2.2448	0.0043	0.0172	0.0695	0.2918
	KLD error	0.0391	0.0016	0.0000	0.0001	0.0002	0.0006	0.0026

the variation is larger than the threshold, the quantization bits should be enlarged, otherwise, the quantization bits should be settled.

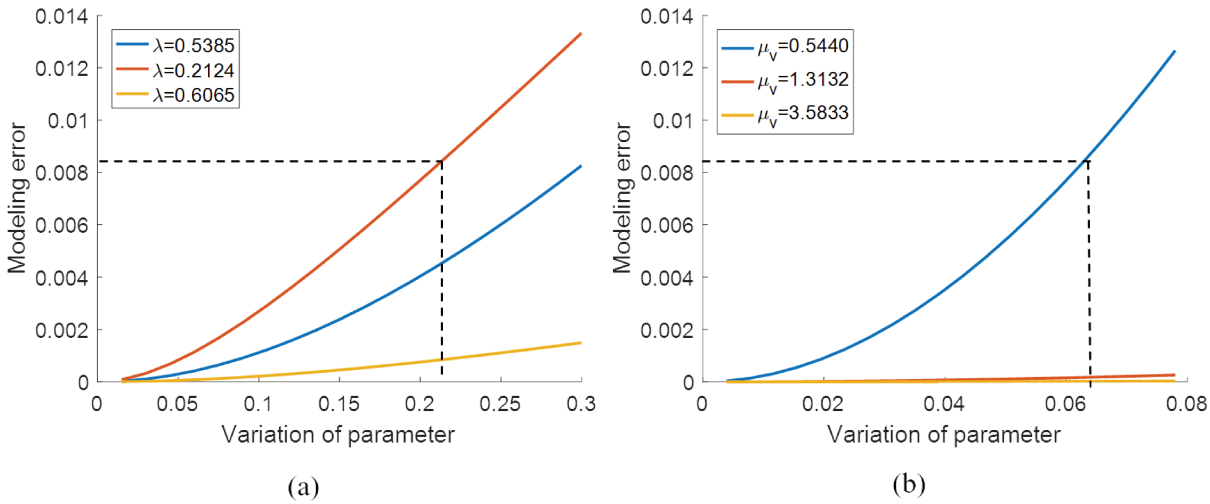


Fig. 4.19 The influences of parameter variation on the model error: (a) Variation of parameter λ and (b) Variation of parameter μ_v.

Taking the magnitude parameters as an example, we choose the minimal value to analyze the influences of parameter variation on the model error, since the minimal value is most sensitive to the quantization interval. The influences of parameter variation on the model error are illustrated in Fig. 4.19, where (a) indicates that the modeling error is within the threshold 0.0086 if the variation of parameter λ is lower than 0.21, and (b) indicates that the modeling error is lower than the threshold 0.0086 if the parameter μ_v vary within 0.065. Therefore, the quantization bits for parameters λ and μ_v are settled as 8 bits and 10 bits, respectively. The quantization bits of the relative phase parameters are determined in a similar way.

Table 4.10 The quantization bits of each parameter

Iterm	Magnitude			Relative phase		
	λ	μ_v	KLD error	ρ	μ_w	KLD error
Quantization bits	8	10	6	8	9	6

The specific quantization bits for all parameters are compiled in Table 4.10, where the six parameters are quantized with 8, 10, 6, 8, 9 and 6 bits, respectively. For the FMRP features, the total quantization bits are $18\times(8+10+6+8+9+6) = 846$ bits, if all the directions are retained. Since one pixel intensity of an image is usually represented with an 8-bit integer, hence, each block standard deviation of the LSD map can be determined as 8 bits. Thus, for the LSD feature, the transmission of the LSD map will cost $12 \times 8 = 96$ bits in total. The new FMRP-LSD metric will cost $846 + 96 = 942$ bits in parameter transmission, which is 0.3% of the cost of image transmission. It is slightly larger than that of the cost of the FMRP metric (0.27%) and far less than another RR-IQA metric SB44 (4.88%) [118].

4.5.3 Improvement for decoding strategy

In this section, we will verify the effectiveness of the new metric for optimizing the decoding strategy. The FMRP-LSD metric is deployed in the JPWL decoder in the same way as FMRP metric, as shown in Fig. 4.5 in Chapter 4. The difference is the FMRP-LSD metric costs more for RR transmission since the LSD map of the reference image is transmitted along with the FMRP features. The LSD map will take 12 bytes of additional cost (details will be discussed in the next section). The FMRP features and LSD feature of the reference image and the distortion one serve as the classification basis and the input of the classifier. Then, after the training of JPWL-RWC database, the classifier selects the better-quality image as the decoding output to optimize the decoding configuration.

To validate the performance, the decoding strategies with FMRP, DCC and FMRP-LSD metrics are compared in Table 4.11, where the experiment utilizes a 5-fold cross-validation and the numbers "1" to "5" denote five subsets of 210 samples, respectively. The experimental results demonstrate that the decoding strategy with FMRP-LSD metric could greatly improve the ratio of the correctly classified samples. Moreover, two wrongly classified samples of decoding strategy with FMRP are shown in Fig. 4.20, where (b) and (d) have obvious local distortions in some areas, which are not detected by the FMRP metric, but they are detected by the FMRP-LSD metric. For the HVS, the classification lables of the two pairs of images are both -2, which means (a) and (c) have a better quality than (b) and (d), respectively. However, for the FMRP metric, (b) and (d) have a better quality, which is opposite to the

experience of the HVS. For the FMRP-LSD metric, (a) and (c) are considered having better quality, which is consistent with the HVS.

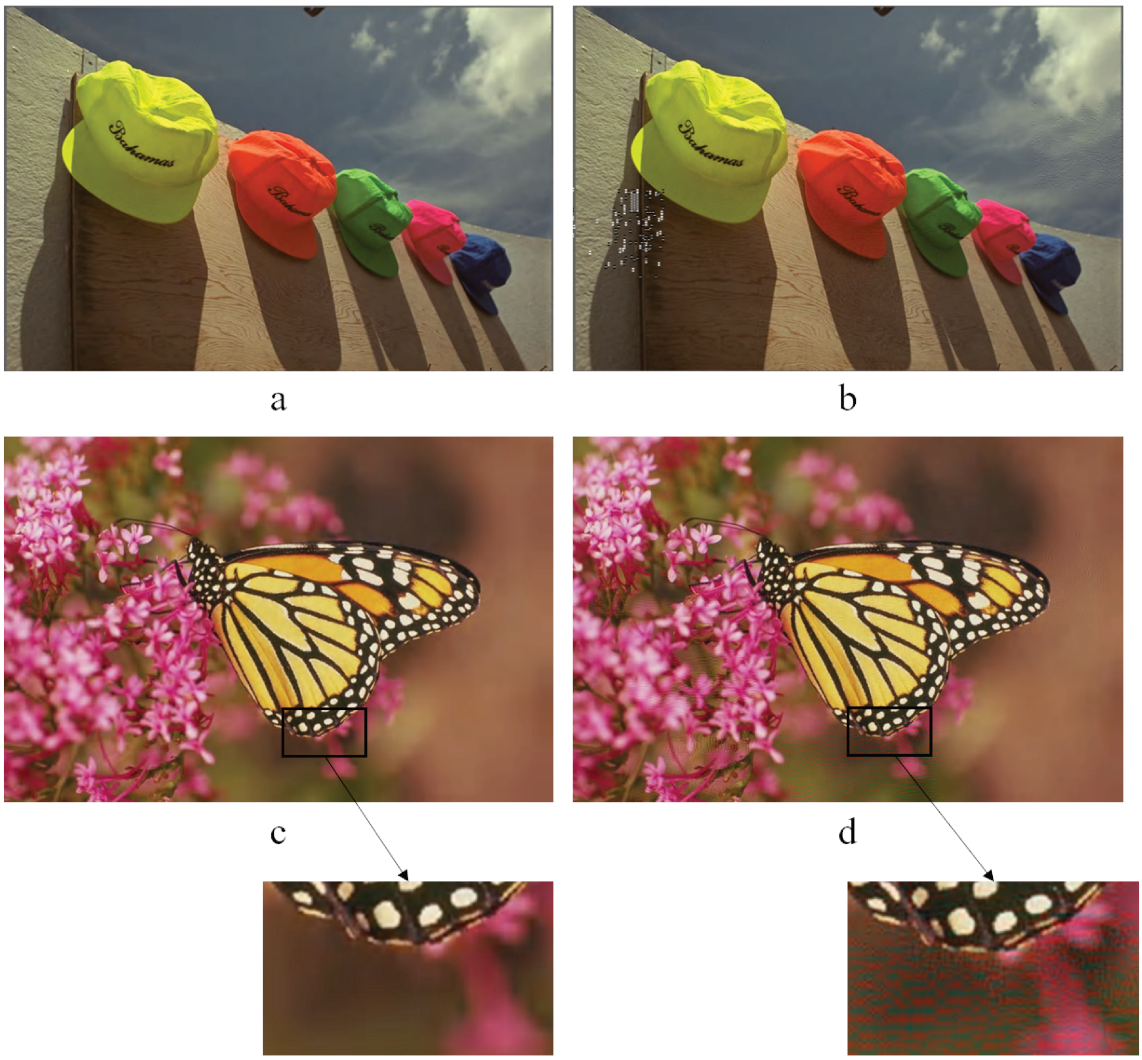


Fig. 4.20 Two wrongly classified samples for FMRP metric, which are correctly classified by FMRP-LSD metric.

Table 4.11 Comparison of decoding strategies with FMRP, DCC and FMRP-LSD

	Ratio of correctly classified samples				
Decoding strategy	1	2	3	4	5
with FMRP	82.86	83.81	82.86	83.81	82.86
with DCC	76.19	76.19	76.19	76.19	76.19
with FMRP-LSD	**88.10**	**89.05**	**87.62**	**87.62**	**88.57**

4.5.4 Validation on comprehensive database

Furtherly, the new metric is validated based on the comprehensive IQA evaluation database TID2013*, which keeps 18 types of distortions of TID2013 and does not contain six exotic distortions that will not occur in the communication system (details illustrated in Subsection 3.7.1). For one single distortion type, there are 125 distortion images and thus the 5-fold cross-validation is utilized. For all 18 distortion types, there are 2250 distortion images and thus the 10-fold cross-validation is used. To compare the performances, PLCC, SROCC and OR are employed to evaluate the prediction accuracy, the prediction monotonicity and the prediction consistency, as shown in Table 5.3. Obviously, the FMRP-LSD metric performs better for most distortion types such as JPEG compression and JPEG 2000 compression, JPEG and JPEG 2000 transmission error, and Blur and White noise distortions. Moreover, the metric also has an obvious improvement for the mixture of all distortions in the TID2013* database. This demonstrates that the LSD feature is effective to detect the image degradation and can improve the accuracy of the FMRP metric. The only one type of distortions without higher accuracy than FMRP metric is JPEG-TE, which is associated with global blur and ringing at the edges. These distortions indeed cannot be detected well by the local LSD feature, however they can be well detected by the FMRP feature. Thus, the FMRP metric performs well in detecting this kind of distortions. In summary, the FMRP-LSD metric could improvement for the objective score prediction.

Table 4.12 Comparison between FMRP and FMRP-LSD metrics on TID2013* database

Metric	Criteria	All	JPEG	JP2K	JPEG-TE	JP2K-TE	Blur	White noise
FMRP	PLCC	0.8504	0.8934	0.9254	**0.9060**	0.8127	0.9296	0.8214
	SROCC	0.8549	0.8478	0.8873	**0.8130**	0.8065	0.9354	0.8036
	OR	0.0686	0.0413	0.0140	0.0327	0.0927	0.0073	0.0900
FMRP-LSD	PLCC	**0.8812**	**0.9238**	**0.9268**	0.9030	**0.8340**	**0.9493**	**0.8314**
	SROCC	**0.8820**	**0.8657**	**0.8902**	0.8124	**0.8256**	**0.9491**	**0.8119**
	OR	**0.0504**	**0.0203**	**0.0087**	**0.0227**	**0.0805**	**0.0027**	**0.0717**

4.6 Chapter summary

In this chapter, we proposed an optimization strategy for decoding configuration over a realistic wireless transmission. The strategy utilizes the RR-IQA metric FMRP to select a better-quality image and determine a decoding configuration. It improves the users' QoE while ensuring the system' QoS.

To validate the decoding strategy, we constructed a new database named JPWL-RWC which comprises 480 images with JP2K-CD and JP2K-TED distortions and contains a large physiological vision test of around 100 people from three countries (France, China and India). The database reveals properties of JP2K-CD and JP2K-TED distortions and thus is helpful to understand and analyze these two distortions. The final purpose of the database construction is to provide a benchmark for optimization of decoding configuration.

According to the properties of JP2K-CD and JP2K-TED distortions, a new metric FMRP-LSD has been proposed. The metric detects the global distortions with the FMRP feature and detects the local distortions with the LSD feature.

The experimental results demonstrate that the decoding strategies with FMRP and FMRP-LSD metrics can greatly improve the QoE, as compared with the classic decoding strategy and the decoding strategy with other quality metric, in various kinds of channel states.

Chapter 5

Conclusion and perspective

5.1 Summary of thesis

The image quality assessment based on the visual features is a popular research topic in image processing field. This thesis proposed two new reduced reference image quality assessment metrics and applied them into optimization of the decoding configuration over a realistic wireless communication channel. The main contributions are as follows:

Firstly, we have proposed a new reduced-reference image quality metric, FMRP, which is based on statistical models in the complex domain. The metric exploits the multi-scale and multi-direction features of the DT-CWT decomposition to detect the image degradation. The features are extracted from statistical models of the magnitude and the relative phase and the parameters serve as the reduced-reference information. The Kuallback-Leibler divergence between models of the reference image and the distortion image is utilized to quantify the image quality and then the objective score is obtained by a generalized regression neural network approach. Usually, the magnitude information is sensitive to the energy change while the phase information is sensitive to the structure change. The existing metrics in the wavelet domain either utilize the magnitude of the DWT real-value wavelet or utilize the phase information of the CWT. This leads to a limited accuracy and lack of robustness. The FMRP metric utilizes advantages from both the magnitude and the phase, and thus has a higher accuracy and a better robustness. The effectiveness has been demonstrated by experiments based on the large public databases.

Secondly, with the new FMRP metric, a new decoding strategy has been proposed for a realistic wireless transmission channel, which can improve the quality of experience while ensuring the quality of service. For this, a new database including 420 distortion images and large physiological vision tests of around 100 people from three countries, has been constructed to collect the visual preference of different decoding configurations.

Then, a classifier based on support vector machine or K-nearest neighboring is exploited to automatically select the decoding configuration. The classical transmission strategies guarantee the quality of service by adjusting the system parameters but do not consider the quality of experience of users. The new strategy improves the quality of experience by applying the quality metric FMRP to measure the image quality. Comparative experiments verify the effectiveness of the decoding strategy in the improvement of image experience. Moreover, the new database reveals the characteristics of the distortions encountered in the wireless communication system, which helps to develop new methods to detect distortions.

Thirdly, according to the specific characteristics of the distortions in realistic wireless channel, a new quality metric FMRP-LSD has been proposed, which is based on the global statistical feature in the wavelet domain and the local statistical feature in the spatial domain. The global feature in wavelet domain is utilized to detect the global distortion such as blur and ringing while the local feature is exploited to detect the local random distortion caused by the noise oscillation. The experimental results verify the effectiveness of the FMRP-LSD metric in the applications of image quality assessment and optimization of decoding strategy.

5.2 Future research

In the thesis, the image quality evaluation technology and the optimization of decoding strategy in wireless communication systems have been studied and some methods have been proposed. However, with the continuous development of multimedia technology, there are still many new applications and problems in the field of image quality evaluation. According to the experience in the research process, we believe that the following research directions or technologies deserve further study.

1. Image quality metric for 3D application scenarios. The FMRP and the FMRP-LSD metrics proposed in this thesis can be applied to evaluate the image quality of both left and right views. Then, if we can find a well-designed weighting coefficient to balance the effects of the degradations of left and right views on total image quality, the 3D image quality metric will be obtained.

2. Image quality evaluation based on a multi-feature fusion. Usually, the image distortions are caused by different variations such as variations in luminance, contrast, color, structure and texture. The metric with multi-feature of different types and different measurement distances can evaluate the image quality in a more comprehensive way. The FMRP metric proposed in Chapter 3 evaluates the image with multi-scale and multi-direction features but they are all statistical features based on the PDF model. It

performs well for global distortions but badly for local distortions. The FMRP-LSD metric proposed in Chapter 5 combines the global feature with local feature and thus has a better performance than FMRP metric. In the future, we may develop a more robust and comprehensive image quality metric by utilizing more features that measure the quality in different ways.

3. No-reference image quality metric. The thesis does not involve the research of no-reference image quality metric but it has a prominent advantage due to its evaluating image without access to the reference image. Although it is difficult to design a no-reference image quality metric suitable for multiple distortions, the no-reference metric has drawn much attention of researchers, recently. With the development of human-visual-system research and the in-depth understanding of various distortion characteristics, it is possible to design no-reference image quality metrics for specific distortions are possible.

4. Image quality metric based on deep neural network. We did not adopt the deep neural network-based IQA metric since the JPWL-RWC database only includes 210 samples, which is too small for deep neural network training. In the future, we can extend the database and introduce the deep neural network-based IQA metric to optimize the decoding strategy.

5. Application of the metrics for other communication systems. We can also extend the FMRP and the FMRP-LSD metrics to other communication systems with different compression and transmission standards.

References

[1] Julien Abot, Michael Nauge, Clency Perrine, Chaker Larabi, Cyril Bergeron, Yannis Pousset, and Christian Olivier. A robust content-based JPWL transmission over a realistic MIMO channel under perceptual constraints. In *International Conference Image Processing (ICIP), 2011 18th IEEE International Conference*, pages 3241–3244. IEEE, 2011.

[2] Julien Abot, Christian Olivier, Clency Perrine, and Yannis Pousset. A link adaptation scheme optimized for wireless JPEG 2000 transmission over realistic MIMO systems. *Signal Processing: Image Communication*, 27(10):1066–1078, 2012.

[3] Julien Abot, Clency Perrine, Yannis Pousset, and Christian Olivier. An unequal power allocation designed for MIMO systems using content characteristics. *International Journal of Research and Reviews in Applied Sciences*, 16(1):16, 2013.

[4] Max Agueh, Jean-François Diouris, Magaye Diop, François-Olivier Devaux, Christophe De Vleeschouwer, and Benoit Macq. Optimal jpwl forward error correction rate allocation for robust jpeg 2000 images and video streaming over mobile ad hoc networks. *EURASIP Journal on Advances in Signal Processing*, 2008:167, 2008.

[5] Hirotugu Akaike. A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19(6):716–723, 1974.

[6] Olivier Alata, Christian Olivier, and Yannis Pousset. Law recognitions by information criteria for the statistical modeling of small scale fading of the radio mobile channel. *Signal Processing*, 93(5):1064–1078, 2013.

[7] Mohammad Ali Bagheri, Gholam Ali Montazer, and Sergio Escalera. Error correcting output codes for multiclass classification: application to two image vision problems. In *Artificial Intelligence and Signal Processing (AISP), 2012 16th CSI International Symposium on*, pages 508–513. IEEE, 2012.

[8] Peter Almers, Ernst Bonek, A Burr, Nicolai Czink, Mérouane Debbah, Vittorio Degli-Esposti, Helmut Hofstetter, Pekka Kyösti, Dave Laurenson, Gerald Matz, et al. Survey of channel and radio propagation models for wireless mimo systems. *EURASIP Journal on Wireless Communications and Networking*, 2007(1):019070, 2007.

[9] Ryan Anderson, Nick Kingsbury, and Julien Fauqueur. Determining multiscale image feature angles from complex wavelet phases. In *International Conference Image Analysis and Recognition*, pages 490–498. Springer, 2005.

[10] Alejandro Aragon-Zavala. *Antennas and propagation for wireless communication systems*. John Wiley & Sons, 2008.

[11] Luigi Atzori. Transmission of jpeg2000 images over wireless channels with unequal power distribution. *IEEE Transactions on Consumer Electronics*, 49(4):883–888, 2003.

[12] Ertuğrul Başar. Multiple-input multiple-output ofdm with index modulation. *IEEE Signal Processing Letters*, 22(12):2259–2263, 2015.

[13] Lucien Birge. Statistical estimation with model selection. *Indagationes Mathematicae*, 17(4):497–537, 2006.

[14] Sebastian Bosse, Dominique Maniry, Klaus-Robert Muller, Thomas Wiegand, and Wojciech Samek. Deep Neural Networks for No-Reference and Full-Reference Image Quality Assessment. *IEEE Transactions on Image Processing*, 27:206–219, 2018.

[15] Alan C. Bovik, Marianna Clark, and Wilson S. Geisler. Multichannel texture analysis using localized spatial filters. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, (1):55–73, 1990.

[16] Serdar Cakir and A Enis Cetin. Image quality assessment using two-dimensional complex mel-cepstrum. *Journal of Electronic Imaging*, 25(6):061604, 2016.

[17] Mathieu Carnec, Patrick Le Callet, and Dominique Barba. An image quality assessment method based on perception of structural information. In *Proceedings of International Conference on Image Processing (ICIP)*, volume 3, pages III–185. IEEE, 2003.

[18] Mathieu Carnec, Patrick Le Callet, and Dominique Barba. Visual features for image quality assessment with reduced reference. In *IEEE International Conference on Image Processing (ICIP)*, volume 1, pages I–421. IEEE, 2005.

[19] Dah-Chung Chang and Wen-Rong Wu. Image contrast enhancement based on a histogram transformation of local standard deviation. *IEEE Transactions on Medical Imaging*, 17(4):518–531, 1998.

[20] Y Chartois, Y Pousset, and R Vauzelle. A siso and mimo radio channel characterization with a 3d ray tracing propagation model in urban environment. *European Conference on Propogation and Systems (ECPS) 2005 Proceedings*, 2005.

[21] Raj Chhikara. *The Inverse Gaussian Distribution: Theory: Methodology, and Applications*, volume 95. CRC Press, 1988.

[22] Peter S Chow, John M Cioffi, and John AC Bingham. A practical discrete multitone transceiver loading algorithm for data transmission over spectrally shaped channels. *IEEE Transactions on Communications*, 43(234):773–775, 1995.

[23] Marianna Clark and Alan C Bovik. Experiments in segmenting texton patterns using localized spatial filters. *Pattern Recognition*, 22(6):707–717, 1989.

[24] Marianna Clark, Alan C Bovik, and Wilson S Geisler. Texture segmentation using gabor modulation/demodulation. *Pattern Recognition Letters*, 6(4):261–267, 1987.

[25] Guilhem Coq, Xiang Li, Olivier Alata, Yannis Pousset, and Christian Olivier. Law recognition via histogram-based estimation. In *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP'09)*, pages 3425–3428, 2009.

[26] Ingrid Daubechies. *Ten lectures on wavelets*, volume 61. SIAM, 1992.

[27] Jeremy S De Bonet and Paul Viola. Texture recognition using a non-parametric multi-scale statistical model. In *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 641–647. IEEE, 1998.

[28] Marco Di Renzo, Harald Haas, and Peter M Grant. Spatial modulation for multiple-antenna wireless systems: A survey. *IEEE Communications Magazine*, 49(12), 2011.

[29] Thomas G Dietterich and Ghulum Bakiri. Solving multiclass learning problems via error-correcting output codes. *Journal of Artificial Intelligence Research*, 2:263–286, 1994.

[30] Yong Ding, Yang Zhao, and Xinyu Zhao. Image quality assessment based on multi-feature extraction and synthesis with support vector regression. *Signal Processing: Image Communication*, 54:81–92, 2017.

[31] Minh N Do and Martin Vetterli. Wavelet-based texture retrieval using generalized Gaussian density and Kullback-Leibler distance. *IEEE Transactions on Image Processing*, 11(LCAV-ARTICLE-2002-001):146–158, 2002.

[32] Frederic Dufaux and Didier Nicholson. Jpwl: Jpeg 2000 for wireless applications. In *Applications of Digital Image Processing XXVII*, volume 5558, pages 309–319. International Society for Optics and Photonics, 2004.

[33] Sorina Dumitrescu, Xiaolin Wu, and Zhe Wang. Globally optimal uneven error-protected packetization of scalable code streams. *IEEE Transactions on Multimedia*, 6(2):230–239, 2004.

[34] Abdelaziz El Matouat and Marc Hallin. Order selection, stochastic complexity and Kullback-Leibler information. In *Athens Conference on Applied Probability and Time Series Analysis*, pages 291–299. Springer, 1996.

[35] Rui Fan, Ya Jun Yu, and Yong Liang Guan. Generalization of orthogonal frequency division multiplexing with index modulation. *IEEE Transactions on Wireless Communications*, 14(10):5350–5359, 2015.

[36] Lian-ru Gao, Bing Zhang, Xia Zhang, and Qian Shen. Study on the method for estimating the noise in remote sensing images based on local standard deviations. *Journal of Remote Sensing-Beijing*, 11(2):201, 2007.

[37] Akshay Gore and Savita Gupta. Full reference image quality metrics for jpeg compressed images. *AEU-International Journal of Electronics and Communications*, 69 (2):604–608, 2015.

[38] Shalini Gupta, Mehul P Sampat, Mia K Markey, Alan C Bovik, and Zhou Wang. Facial range image matching using the complexwavelet structural similarity metric. In *Applications of Computer Vision, 2007. WACV'07. IEEE Workshop on*, pages 4–4. IEEE, 2007.

[39] IEEE-802-11-Working-Group et al. IEEE standard for information technology–telecommunications and information exchange between systems–local and metropolitan area networks–specific requirements–part 11: Wireless lan medium access control (mac) and physical layer (phy) specifications amendment 6: Wireless access in vehicular environments. *IEEE Std*, 802(11), 2010.

[40] IEEE-802-16-Working-Group et al. IEEE 802.16 e–2005 ieee standard for local and metropolitan area networks part 16: Air interface for fixed and mobile broadband wireless access systems amendment for physical and medium access control layers for combined fixed and mobile operation in licensed bands. *Retrieved from standards. ieee. org/getieee802/download/802.16 e-2005. pdf*, 2005.

[41] Magdy F Iskander and Zhengqing Yun. Propagation prediction models for wireless communication systems. *IEEE Transactions on Microwave Theory and Techniques*, 50(3):662–673, 2002.

[42] ISO/IEC 154444-11:2007. Information technology–jpeg 2000 image coding system: Wireless. Standard, International Organization for Standardization, Geneva, Switzerland, 2007.

[43] ISO/IEC 154444-1:2016. Information technology–jpeg 2000 image coding system: Core coding system. Standard, International Organization for Standardization, Geneva, Switzerland, 2016.

[44] ITU-T. Objective perceptual assessment of video quality: Full reference television. *ITU-T Telecommunication Standardization Bureau (2004)*, 2004.

[45] Frédéric Jouzel, Christian Olivier, and Abdelaziz El Matouat. Information criteria based edge detection. In *9th European Signal Processing Conference (EUSIPCO)*, pages 1–4. IEEE, 1998.

[46] Nick Kingsbury. Image processing with complex wavelets. *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 357(1760):2543–2560, 1999.

[47] Nick Kingsbury. Complex wavelets for shift invariant analysis and filtering of signals. *Applied and Computational Harmonic Analysis*, 10(3):234–253, 2001.

[48] Eleftherios Kofidis, Dimitrios Katselis, Athanasios Rontogiannis, and Sergios Theodoridis. Preamble-based channel estimation in ofdm/oqam systems: A review. *Signal Processing*, 93(7):2038–2054, 2013.

[49] Jeffrey C Lagarias, James A Reeds, Margaret H Wright, and Paul E Wright. Convergence properties of the nelder–mead simplex method in low dimensions. *Journal on Optimization*, 9(1):112–147, 1998.

[50] Eric Cooper Larson and Damon Michael Chandler. Most apparent distortion: full-reference image quality assessment and the role of strategy. *Journal of Electronic Imaging*, 19(1):011006, 2010.

[51] Michael C Lawton and JP McGeehan. The application of a deterministic ray launching algorithm for the prediction of radio channel characteristics in small-cell environments. *IEEE Transactions on Vehicular Technology*, 43(4):955–969, 1994.

[52] Chaofeng Li, Alan Conrad Bovik, and Xiaojun Wu. Blind image quality assessment using a general regression neural network. *IEEE Transactions on Neural Networks*, 22(5):793–799, 2011.

[53] Qiang Li and Zhou Wang. Reduced-reference image quality assessment using divisive normalization-based image representation. *IEEE Journal of Selected Topics in Signal Processing*, 3(2):202–211, 2009.

[54] Zhichao Lin, Jinxu Tao, and Zhufeng Zheng. Reduced-reference image quality assessment based on average directional information. In *IEEE 11th International Conference on Signal Processing (ICSP)*, volume 2, pages 787–791. IEEE, 2012.

[55] Zhichao Lin, Zhufeng Zheng, Ronghua Guo, and Liangfeng Pei. Reduced-reference image quality assessment based on phase information in complex wavelet domain. In *IEEE 12th International Conference on Signal Processing (ICSP)*, pages 966–971. IEEE, 2014.

[56] Delei Liu, Yong Xu, Yuhui Quan, and Patrick Le Callet. Reduced reference image quality assessment using regularity of phase congruency. *Signal Processing: Image Communication*, 29(8):844–855, 2014.

[57] Lin Ma, Songnan Li, Fan Zhang, and King Ngi Ngan. Reduced-reference image quality assessment using reorganized dct-based image representation. *IEEE Trans. Multimedia*, 13(4):824–829, 2011.

[58] Florence Jessie MacWilliams and Neil James Alexander Sloane. *The theory of error-correcting codes*. Elsevier, 1977.

[59] Julian Magarey and Nick Kingsbury. Motion estimation using a complex-valued wavelet transform. *IEEE Transactions on Signal Processing*, 46(4):1069–1084, 1998.

[60] Cristina Mairal and Max Agueh. Scalable and robust jpeg 2000 images and video transmission system for multiple wireless receivers. In *2010 IEEE Latin-American Conference on Communications (LATINCOM)*, pages 1–6. IEEE, 2010.

[61] Kanti V Mardia and Peter E Jupp. volume 494. John Wiley & Sons, 2009.

[62] Pina Marziliano, Frederic Dufaux, Stefan Winkler, and Touradj Ebrahimi. Perceptual blur and ringing metrics: application to jpeg2000. *Signal Processing: Image Communication*, 19(2):163–172, 2004.

[63] Anish Mittal, Anush Krishna Moorthy, and Alan Conrad Bovik. No-reference image quality assessment in the spatial domain. *IEEE Transactions on Image Processing*, 21 (12):4695–4708, 2012.

[64] Anush Krishna Moorthy and Alan Conrad Bovik. Blind image quality assessment: From natural scene statistics to perceptual quality. *IEEE Transactions on Image Processing*, 20(12):3350–3364, 2011.

[65] Robert H Morelos-Zaragoza. *The art of error correcting coding*. John Wiley & Sons, 2006.

[66] Michael Nauge, M-C Larabi, and C Fernandez. A reduced-reference metric based on the interest points in color images. In *28th Picture Coding Symposium*, pages 610–613. IEEE, 2010.

[67] Truong Nguyen and Herve Chauris. Uniform discrete curvelet transform for seismic processing. In *70th EAGE Conference and Technical Exhibition, Eur. Ass. of Geoscientists and Engineers*, page P066, 2008.

[68] Truong T Nguyen and Soontorn Oraintara. The shiftable complex directional pyramid-Part I: Theoretical aspects. *IEEE Transactions on Signal Processing*, 56(10):4651–4660, 2008.

[69] Alan V Oppenheim and Jae S Lim. The importance of phase in signals. *Proceedings of the IEEE*, 69(5):529–541, 1981.

[70] Patrice Pajusco. Propagation channel models for mobile communication. *Comptes Rendus Physique*, 7(7):703–714, 2006.

[71] Nikolay Ponomarenko and Oleg Ieremeiev. Color image database TID2013: Peculiarities and preliminary results. In *European Workshop on Visual Information Processing (EUVIP)*, pages 106–111. IEEE, 2013.

[72] Nikolay Ponomarenko, Vladimir Lukin, Alexander Zelensky, Karen Egiazarian, Marco Carli, and Federica Battisti. Tid2008-a database for evaluation of full-reference visual quality assessment metrics. *Advances of Modern Radioelectronics*, 10(4):30–45, 2009.

[73] M Ali Qureshi and Mohamed Deriche. A fast no reference image quality assessment using laws texture moments. In *IEEE Global Conference on Signal and Information Processing (GCSIP)*, pages 979–983. IEEE, 2014.

[74] SM Mahbubur Rahman, M Omair Ahmad, and MNS Swamy. Statistics of 2-D DT-CWT coefficients for a Gaussian distributed signal. *IEEE Transactions on Circuits and Systems I: Regular Papers*, 55(7):2013–2025, 2008.

[75] Yothin Rakvongthai and Soontorn Oraintara. Statistical image modeling with the magnitude probability density function of complex wavelet coefficients. In *IEEE International Symposium on Circuits and Systems, 2009. ISCAS 2009.*, pages 1879–1882. IEEE, 2009.

[76] ITURBT Recommendation. 500-11,"methodology for the subjective assessment of the quality of television pictures," recommendation itu-r bt. 500-11. *ITU Telecom. Standardization Sector of ITU*, 7, 2002.

[77] Abdul Rehman and Zhou Wang. Reduced-reference image quality assessment by structural similarity estimation. *IEEE Transactions on Image Processing*, 21(8): 3378–3389, 2012.

[78] Ann Marie Rohaly, Philip J Corriveau, John M Libert, Arthur A Webster, Vittorio Baroncini, John Beerends, Jean-Louis Blin, Laura Contin, Takahiro Hamada, David Harrison, et al. Video quality experts group: Current results and future directions. In *Visual Communications and Image Processing 2000*, volume 4067, pages 742–754. International Society for Optics and Photonics, 2000.

[79] Justin K Romberg, Hyeokho Choi, and Richard G Baraniuk. Multiscale edge grammars for complex wavelet transforms. In *International Conference on Image Processing Proceedings, 2001.*, volume 1, pages 614–617. IEEE, 2001.

[80] Michele A Saad, Alan C Bovik, and Christophe Charrier. Blind image quality assessment: A natural scene statistics approach in the dct domain. *IEEE Transactions on Image Processing*, 21(8):3339–3352, 2012.

[81] Muhammad Farooq Sabir, Alan Conrad Bovik, and Robert W Heath. Unequal power allocation for jpeg transmission over mimo systems. *IEEE Transactions on image processing*, 19(2):410–421, 2010.

[82] Mehul P Sampat, Zhou Wang, Shalini Gupta, Alan Conrad Bovik, and Mia K Markey. Complex wavelet structural similarity: A new image similarity index. *IEEE transactions on image processing*, 18(11):2385–2401, 2009.

[83] Gideon Schwarz et al. Estimating the dimension of a model. *The Annals of Statistics*, 6(2):461–464, 1978.

[84] Ivan W Selesnick and Ke Yong Li. Video denoising using 2D and 3D dual-tree complex wavelet transforms. In *Wavelets: Applications in Signal and Image Processing X*, volume 5207, pages 607–619. International Society for Optics and Photonics, 2003.

[85] Ivan W Selesnick, Richard G Baraniuk, and Nick C Kingsbury. The dual-tree complex wavelet transform. *IEEE Signal Processing Magazine*, 22(6):123–151, 2005.

[86] Stefania Sesia, Matthew Baker, and Issam Toufik. *LTE-the UMTS long term evolution: from theory to practice*. John Wiley & Sons, 2011.

[87] Cián W Shaffrey, Nick G Kingsbury, and Ian H Jermyn. Unsupervised image segmentation via Markov Trees and complex wavelets. In *Proceedings of IEEE International Conference on Image processing*, pages 801–804. IEEE, 2002.

[88] Joseph Shapira and Shmuel Y Miller. The mobile radio propagation channel. *CDMA Radio with Repeaters*, pages 45–99, 2007.

[89] Hamid R Sheikh and Alan C Bovik. Image information and visual quality. In *Proceedings of International Conference on Acoustics, Speech, and Signal Processing (ICASSP'04)*, volume 3, pages iii–709. IEEE, 2004.

[90] Hamid R Sheikh, Zhou Wang, Lawrence Cormack, and Alan C Bovik. Live image quality assessment database release 2 (2005), 2005.

[91] Hamid R Sheikh, Muhammad F Sabir, and Alan C Bovik. A statistical evaluation of recent full reference image quality assessment algorithms. *IEEE Transactions on Image Processing*, 15(11):3440–3451, 2006.

[92] Natalia Silberstein, Ankit Singh Rawat, and Sriram Vishwanath. Error-correcting regenerating and locally repairable codes via rank-metric codes. *IEEE Transactions on Information Theory*, 61(11):5765–5778, 2015.

[93] Marvin K Simon and Mohamed-Slim Alouini. *Digital communication over fading channels*, volume 95. John Wiley & Sons, 2005.

[94] Jingge Song, Carlos E Romero, Zheng Yao, and Boshu He. A globally enhanced general regression neural network for on-line multiple emissions prediction of utility boiler. *Knowledge-Based Systems*, 118:4–14, 2017.

[95] Donald F Specht. A general regression neural network. *IEEE Transactions on Neural Networks*, 2(6):568–576, 1991.

[96] J-L Starck and Fionn Murtagh. Astronomical image and signal processing: looking at noise, information and scale. *IEEE Signal Processing Magazine*, 18(2):30–40, 2001.

[97] Gordon L Stuber, John R Barry, Steve W Mclaughlin, Ye Li, Mary Ann Ingram, and Thomas G Pratt. Broadband mimo-ofdm wireless communications. *Proceedings of the IEEE*, 92(2):271–294, 2004.

[98] Marco Tagliasacchi, Giuseppe Valenzise, Matteo Naccari, and Stefano Tubaro. A reduced-reference structural similarity approximation for videos corrupted by channel errors. *Multimedia Tools and Applications*, 48(3):471–492, 2010.

[99] SY Tan and HS Tan. A microcellular communications propagation model based on the uniform theory of diffraction and multiple image theory. *IEEE Transactions on Antennas and Propagation*, 44(10):1317–1326, 1996.

[100] S Tourancheau, F Autrusseau, and P Sazzad. Impact of subjective dataset on the performance of image quality metrics. In *IEEE International Conference on Image*. Citeseer, 2008.

[101] Albekaye Traoré, Philippe Carré, and Christian Olivier. Quaternionic wavelet coefficients modeling for a reduced-reference metric. *Signal Processing: Image Communication*, 36:127–139, 2015.

[102] David Tse and Pramod Viswanath. *Fundamentals of wireless communication*. Cambridge university press, 2005.

[103] Alexey V Umnov, Andrey S Krylov, and Andrey V Nasonov. Ringing artifact suppression using sparse representation. In *International Conference on Advanced Concepts for Intelligent Vision Systems*, pages 35–45. Springer, 2015.

[104] Martin Vetterli. Wavelets, approximation, and compression. *IEEE Signal Processing Magazine*, 18(5):59–73, 2001.

[105] An Vo and Soontorn Oraintara. A study of relative phase in complex wavelet domain: Property, statistics and applications in texture image retrieval and segmentation. *Signal Processing: Image Communication*, 25(1):28–46, 2010.

[106] An PN Vo, Truong T Nguyen, and Soontorn Oraintara. Image denoising using shiftable directional pyramid and scale mixtures of complex gaussians. In *IEEE International Symposium on Circuits and Systems (ISCAS)*, pages 4000–4003. IEEE, 2007.

[107] Zhou Wang. *Waterloo IVC Code Repository*. http://ivc.uwaterloo.ca/ivc-code.php.

[108] Zhou Wang and Eero P Simoncelli. Reduced-reference image quality assessment using a wavelet-domain natural image statistic model. In *Human Vision and Electronic Imaging X*, volume 5666, pages 149–160. International Society for Optics and Photonics, 2005.

[109] Zhou Wang and Eero P Simoncelli. Translation insensitive image similarity in complex wavelet domain. In *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP'05)*, volume 2, pages ii–573. IEEE, 2005.

[110] Zhou Wang, Eero Simoncelli, Alan Bovik, et al. Multi-scale structural similarity for image quality assessment. In *ASILOMAR Conference on Signals, Systems and Computers*, volume 2, pages 1398–1402. IEEE; 1998, 2003.

[111] Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4):600–612, 2004.

[112] Zhou Wang, Guixing Wu, Hamid R Sheikh, Eero P Simoncelli, En-Hui Yang, and Alan C Bovik. Quality-aware images. *IEEE Transactions on Image Processing*, 15(6): 1680–1689, 2006.

[113] Stephen B Weinstein. The history of orthogonal frequency-division multiplexing [history of communications]. *IEEE Communications Magazine*, 47(11), 2009.

[114] Meiling Xie. *Indoor radio propagation modeling for system performance prediction*. PhD thesis, INSA de Lyon, 2013.

[115] Zixiang Xiong, Kannan Ramchandran, Michael T Orchard, and Ya-Qin Zhang. A comparative study of dct-and wavelet-based image coding. *IEEE Transactions on circuits and systems for video technology*, 9(5):692–695, 1999.

[116] Lin Zhang, Lei Zhang, Xuanqin Mou, David Zhang, et al. FSIM: a feature similarity index for image quality assessment. *IEEE Transactions on Image Processing*, 20(8): 2378–2386, 2011.

[117] Qi Tu Zhang. Maximal-ratio combining over nakagami fading channels with an arbitrary branch covariance matrix. *IEEE Transactions on Vehicular Technology*, 48 (4):1141–1150, 1999.

[118] Yi Zhang, Thien D Phan, and Damon M Chandler. Reduced-reference image quality assessment based on distortion families of local perceived sharpness. *Signal Processing: Image Communication*, 55:130–145, 2017.

[119] Debin Zhao, Wen Gao, and YK Chan. Morphological representation of dct coefficients for image compression. *IEEE Transactions on Circuits and Systems for Video Technology*, 12(9):819–823, 2002.