Perceptual Quality-of-Experience of Stereoscopic 3D Images and Videos

by

Jiheng Wang

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Doctor of Philosophy
in
Electrical and Computer Engineering

Waterloo, Ontario, Canada, 2016

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Abstract

With the fast development of 3D acquisition, communication, processing and display technologies, automatic quality assessment of 3D images and videos has become ever important. Nevertheless, recent progress on 3D image quality assessment (IQA) and video quality assessment (VQA) remains limited. The purpose of this research is to investigate various aspects of human visual quality-of-experience (QoE) when viewing stereoscopic 3D images/videos and to develop objective quality assessment models that automatically predict visual QoE of 3D images/videos.

Firstly, we create a new subjective 3D-IQA database that has two features that are lacking in the literature, i.e., the inclusion of both 2D and 3D images, and the inclusion of mixed distortion types. We observe strong distortion type dependent bias when using the direct average of 2D image quality to predict 3D image quality. We propose a binocular rivalry inspired multi-scale model to predict the quality of stereoscopic images and the results show that the proposed model eliminates the prediction bias, leading to significantly improved quality predictions.

Second, we carry out two subjective studies on depth perception of stereoscopic 3D images. The first one follows a traditional framework where subjects are asked to rate depth quality directly on distorted stereopairs. The second one uses a novel approach, where the stimuli are synthesized independent of the background image content and the subjects are asked to identify depth changes and label the polarities of depth. Our analysis shows that the second approach is much more effective at singling out the contributions of stereo cues in depth perception. We initialize the notion of depth perception difficulty index (DPDI) and propose a novel computational model for DPDI prediction. The results show that the proposed model leads to highly promising DPDI prediction performance.

Thirdly, we carry out subjective 3D-VQA experiments on two databases that contain various asymmetrically compressed stereoscopic 3D videos. We then compare different mixed-distortions asymmetric stereoscopic video coding schemes with symmetric coding methods and verify their potential coding gains. We propose a model to account for the prediction bias from using direct averaging of 2D video quality to predict 3D video quality. The results show that the proposed model leads to significantly improved quality predictions and can help us predict the coding gain of mixed-distortions asymmetric video compression.

Fourthly, we investigate the problem of objective quality assessment of Multi-view-plus-depth (MVD) images, with a main focus on the pre- depth-image-based-rendering (pre-DIBR) case. We find that existing IQA methods are difficult to be employed as a guiding criterion in the optimization of MVD video coding and transmission systems when applied post-DIBR. We propose a novel pre-DIBR method based on information content weighting of both texture and depth images, which demonstrates competitive performance against state-of-the-art IQA models applied post-DIBR.

Acknowledgements

First of all, I wish to express my sincere appreciation and gratitude to my supervisor, Dr. Zhou Wang. Dr. Wang introduced me to my Ph.D. study in engineering five years ago and has always been very patient to guide me. His insightful observations and comments helped me establish the overall direction of my research as well as to move forward with my studies in depth. It is impossible for myself to finish any of my research work without his guidance. It has been my pleasure and a tremendous opportunity to work with Dr. Wang.

Thank you to the members of my examining committee, Dr. Dana Kulić, Dr. Liang-Liang Xie, Dr. Edward R. Vrscay and Dr. Konstantinos N. Plataniotis. I offer my sincere gratitude for your help, cooperation, and reviewing this thesis.

The University of Waterloo offers a rich and productive environment to explore new ideas. I am grateful to have had the chance to study amidst a greatly supportive community and be surrounded by wonderful colleagues. I would also like to thank all members at Image and Vision Computing Laboratory, Shahrukh Athar, Zhengfang Duanmu, Yuming Fang, Yang Gao, Ke Gu, Rania Hassen, Wentao Liu, Kede Ma, Xiongkuo Min, Rasoul Mohammadi Nasiri, Nima Nikvand, Ilona Kowalik-Urbaniak, Abdul Rehman, Shiqi Wang, Qingbo Wu, Kaiwen Ye, Hojatollah Yeganeh, Kai Zeng, Wei Zhang and Tiesong Zhao. I would also like to thank my friends, Zhiyue Huang and Zhenhao Li for their support throughout the course of my studies.

I would like to express my sincere gratitude to Dr. Dake He and Mr. Nguyen Nguyen at BlackBerry, who supervised me during my four months of internship. Their invaluable ideas and guidance introduced me to work in industry. I would also like to thank Dr. Guangzhe Fan, who supervised and supported me during my master study at University of Waterloo.

Last, but not least, I would like to thank my parents, Wei Wang, and Lingli Ji, for giving me life and privilege to pursue my dream.

Table of Contents

List of Figures xi

Acronyms xix

1 Introduction **1**

1.1 Motivation 1

1.2 Objectives 2

1.3 Contributions 2

1.4 Thesis Outline 3

2 Background **6**

2.1 2D Image Quality Assessment 6

2.1.1 Mean Squared Error and Peak Signal-to-Noise Ratio 8

2.1.2 Structural Similarity 8

2.1.3 Subjective 2D-IQA Studies and Objective Evaluations 11

2.2 3D Visual Quality Factors 15

2.2.1 Impact of Eye Dominance 15

2.2.2 Impact of Image Distortions 16

2.2.3 Impact of Depth Perception 16
2.2.4 Impact of Visual Discomfort 17
2.2.5 Overall 3D Quality-of-Experience 18
2.3 Subjective 3D Image/Video Quality Assessment 18
2.3.1 Subjective 3D-IQA Studies 18
2.3.2 Subjective 3D-VQA Studies 24
2.4 Objective 3D Image/Video Quality Assessment 27
2.4.1 Objective 3D-IQA Studies 27
2.4.2 Objective 3D-VQA Studies 29
2.5 Discussion on existing 3D-IQA and 3D-VQA studies 30

3 Quality prediction of asymmetrically distorted stereoscopic 3D images 33
3.1 Introduction 34
3.2 Subjective Database 34
3.2.1 Waterloo-IVC 3D Image Quality Database Phase I 34
3.2.2 Waterloo-IVC 3D Image Quality Database Phase II 40
3.2.3 Summary 48
3.3 Subjective Test 50
3.4 Impact of Eye Dominance 59
3.5 Analysis and Key Observations 60
3.5.1 Subjective data analysis 60
3.5.2 Key observation: Distortion type dependency 63
3.6 Objective Study I: 2D-to-3D Quality Prediction 71
3.6.1 Objective 2D Quality Assessment 71

3.6.2 2D-to-3D Quality Prediction Model 74
3.6.3 Validation 83
3.7 Objective Study II: Blind 2D-to-3D Quality Prediction 93
3.7.1 Introduction 93
3.7.2 Blind 2D-to-3D Quality Prediction 100
3.7.3 Validation 103
3.8 Summary 113

4 Quantifying Perceptual Depth Quality in Distorted Stereoscopic Images 114

4.1 Introduction 115
4.2 Subjective Study I 117
4.2.1 Image Database 117
4.2.2 Subjective Test 118
4.2.3 Observations and Discussions 120
4.3 Subjective Study II 124
4.3.1 Image Database 124
4.3.2 Subjective Test 129
4.3.3 Depth Perception Difficulty Index (DPDI) 131
4.3.4 Analysis and Key Observations 132
4.3.5 Impact of Eye Dominance 133
4.4 Objective Study: Prediction Of Depth Perception Difficulty Index 135
4.4.1 DPDI Prediction Model 135
4.4.2 Validation 142
4.5 Discussions 145
4.6 Summary 147

5 Asymmetrically Compressed Stereoscopic 3D Videos: Quality Assessment and Rate-Distortion Performance Evaluation 148

5.1 Introduction 149

5.2 Subjective Study 151

5.2.1 WATERLOO-IVC 3D Video Quality Databases Phase I and Phase II 151

5.2.2 Subjective Test 156

5.2.3 Impact of Eye Dominance 158

5.3 Analysis and Findings 159

5.3.1 Relationship between 2D and 3D Video Quality 159

5.3.2 Quality of Asymmetric Stereoscopic Video with Postprocessing . . . 164

5.3.3 Rate-Distortion Performance of Mixed Distortions Asymmetric Stereoscopic Video 172

5.4 A Model for 2D-to-3D Quality Prediction 174

5.4.1 2D Video Quality Prediction 174

5.4.2 2D-to-3D Quality Prediction 175

5.4.3 Validation 177

5.5 Summary 178

6 Quality Assessment of Multi-View-Plus-Depth Images 179

6.1 Introduction 180

6.2 Post-DIBR Quality Assessment 181

6.3 Pre-DIBR Quality Assessment 188

6.3.1 Distortion type dependency 188

6.3.2 Pre-DIBR 3D-IQA algorithm 192

6.4 Summary 196

7 Conclusion and Future Work 197

7.1 Conclusion 197

7.2 Future Work 199

7.2.1 Research on Depth Quality, Visual Comfort and Overall 3D Quality-of-Experience 200

7.2.2 Quality Assessment of Asymmetrically Distorted Videos 201

7.2.3 Quality Assessment of Multi-View-Plus-Depth Images 201

References 203

List of Figures

3.1 The 6 pristine images in Waterloo-IVC 3D Image Database Phase I. Only the right-views are shown here. (a) Art. (b) Books. (c) Dolls. (d) Moebius. (e) Laundry. (f) Reindeer. 35

3.2 The 10 pristine images in Waterloo-IVC 3D Image Database Phase II. Only the right-views are shown here. (a) Barrier. (b) CraftLoom. (c) Dancer. (d) Hall. (e) Laboratory. (f) OldTownCar. (g) Persons. (h) Soccer. (i) Tree. (j) Umbrella. 47

3.3 ASUS 27” VG278H 3D LED monitor with NVIDIA 3D Vision$^{\mathrm{TM}}$2 active shutter glasses. 51

3.4 Experimental setup: the subject is facing the 3D LED monitor with active shutter glasses using a holder to control the viewing distance. 52

3.5 The procedure of the subjective test in Waterloo-IVC 3D Image Database Phase I and Phase II. 55

3.6 A test image in VC test: the right-view is JPEG distorted while the left-view is noise distorted. 55

3.7 Rating strategies for 2DIQ, 3DIQ, DQ, VC and 3DQoE. 57

3.8 The customized GUI application for recording scores 59

3.9 MOS scores of all stimuli in Waterloo-IVC 3D Image Database Phase I. (a) 3DIQ. (b) DQ. (c) VC. (d) 3DQoE. 61

3.10 3DIQ-MOS scores of all stimuli in Waterloo-IVC 3D Image Database Phase II. 62

3.11 3DIQ-MOS versus predictions from 2DIQ-MOS of 2D left- and right-views. (a) All Images, Average 2DIQ-MOS on Waterloo-IVC Phase I. (b) All Images, Weighted 2DIQ-MOS on Waterloo-IVC Phase I. (c) All Images, Average 2DIQ-MOS on Waterloo-IVC Phase II. (d) All Images, Weighted 2DIQ-MOS on Waterloo-IVC Phase II. 66

3.12 3DIQ-MOS versus predictions from 2DIQ-MOS of 2D left- and right-views. (a) Noisy Images, Average 2DIQ-MOS on Waterloo-IVC Phase I. (b) Noisy Images, Weighted 2DIQ-MOS on Waterloo-IVC Phase I. (c) Noisy Images, Average 2DIQ-MOS on Waterloo-IVC Phase II. (d) Noisy Images, Weighted 2DIQ-MOS on Waterloo-IVC Phase II. 67

3.13 3DIQ-MOS versus predictions from 2DIQ-MOS of 2D left- and right-views. (a) Blurred Images, Average 2DIQ-MOS on Waterloo-IVC Phase I. (b) Blurred Images, Weighted 2DIQ-MOS on Waterloo-IVC Phase I. (c) Blurred Images, Average 2DIQ-MOS on Waterloo-IVC Phase II. (d) Blurred Images, Weighted 2DIQ-MOS on Waterloo-IVC Phase II. 68

3.14 3DIQ-MOS versus predictions from 2DIQ-MOS of 2D left- and right-views. (a) JPEG Images, Average 2DIQ-MOS on Waterloo-IVC Phase I. (b) JPEG Images, Weighted 2DIQ-MOS on Waterloo-IVC Phase I. (c) JPEG Images, Average 2DIQ-MOS on Waterloo-IVC Phase II. (d) JPEG Images, Weighted 2DIQ-MOS on Waterloo-IVC Phase II. 69

3.15 3DIQ-MOS versus predictions from 2DIQ-MOS of 2D left- and right-views. (a) Mixed Distortion, Average 2DIQ-MOS on Waterloo-IVC Phase I. (b) Mixed Distortion, Weighted 2DIQ-MOS on Waterloo-IVC Phase I. (c) Mixed Distortion, Average 2DIQ-MOS on Waterloo-IVC Phase II. (d) Mixed Distortion, Weighted 2DIQ-MOS on Waterloo-IVC Phase II. 70

3.16 3DIQ-MOS versus average 2DIQ-MOS - 3DIQ-MOS for Group 3D.2 on Waterloo-IVC 3D Image Database Phase I and Phase II. 72

3.17 Binocular fusion versus Binocular Rivalry. 77

3.18 Diagram of the proposed 2D-to-3D quality prediction model. 78

3.19 Local Energy Map E. (a) Reference. (b) Noise. (c) Blur. (d) JPEG. 79

3.20 Local Energy Ratio Map R. (a) Reference. (b) Noise. (c) Blur. (d) JPEG. 80

3.21 Binocular Rivalry Example 1. (a) Left-view: Reference. (b) Right-view: Blur 81

3.22 Binocular Rivalry Example 2. (a) Left-view: Reference. (b) Right-view: JPEG. 82

3.23 3DIQ-MOS versus predictions from IDW-SSIM of 2D left- and right-views. (a) All Images, Average IDW-SSIM on Waterloo-IVC Phase I. (b) All Images, Weighted IDW-SSIM on Waterloo-IVC Phase I. (c) All Images, Average IDW-SSIM on Waterloo-IVC Phase II. (d) All Images, Weighted IDW-SSIM on Waterloo-IVC Phase II. 87

3.24 3DIQ-MOS versus predictions from IDW-SSIM of 2D left- and right-views. (a) Noisy Images, Average IDW-SSIM on Waterloo-IVC Phase I. (b) Noisy Images, Weighted IDW-SSIM on Waterloo-IVC Phase I. (c) Noisy Images, Average IDW-SSIM on Waterloo-IVC Phase II. (d) Noisy Images, Weighted IDW-SSIM on Waterloo-IVC Phase II. 88

3.25 3DIQ-MOS versus predictions from IDW-SSIM of 2D left- and right-views. (a) Blurred Images, Average IDW-SSIM on Waterloo-IVC Phase I. (b) Blurred Images, Weighted IDW-SSIM on Waterloo-IVC Phase I. (c) Blurred Images, Average IDW-SSIM on Waterloo-IVC Phase II. (d) Blurred Images, Weighted IDW-SSIM on Waterloo-IVC Phase II. 89

3.26 3DIQ-MOS versus predictions from IDW-SSIM of 2D left- and right-views. (a) JPEG Images, Average IDW-SSIM on Waterloo-IVC Phase I. (b) JPEG Images, Weighted IDW-SSIM on Waterloo-IVC Phase I. (c) JPEG Images, Average IDW-SSIM on Waterloo-IVC Phase II. (d) JPEG Images, Weighted IDW-SSIM on Waterloo-IVC Phase II. 90

3.27 3DIQ-MOS versus predictions from IDW-SSIM of 2D left- and right-views. (a) Mixed Distortion, Average IDW-SSIM on Waterloo-IVC Phase I. (b) Mixed Distortion, Weighted IDW-SSIM on Waterloo-IVC Phase I. (c) Mixed Distortion, Average IDW-SSIM on Waterloo-IVC Phase II. (d) Mixed Distortion, Weighted IDW-SSIM on Waterloo-IVC Phase II. 91

3.28 Diagram of the proposed blind 2D-to-3D quality prediction model. 101

3.29 3DIQ-MOS versus predictions from 2DIQ-MOS of 2D left- and right-views. (a) All Images, predictions by direct averaging the 2DIQ-MOS scores of both views on Waterloo-IVC Phase I. (b) All Images, predictions by the proposed model on Waterloo-IVC Phase I. (c) All Images, predictions by direct averaging 2DIQ-MOS scores of both views on Waterloo-IVC Phase II. (d) All Images, predictions by the proposed model on Waterloo-IVC Phase II. 105

3.30 3DIQ-MOS versus predictions from 2DIQ-MOS of 2D left- and right-views. (a) Noisy Images, predictions by direct averaging the 2DIQ-MOS scores of both views on Waterloo-IVC Phase I. (b) Noisy Images, predictions by the proposed model on Waterloo-IVC Phase I. (c) Noisy Images, predictions by direct averaging 2DIQ-MOS scores of both views on Waterloo-IVC Phase II. (d) Noisy Images, predictions by the proposed model on Waterloo-IVC Phase II. 106

3.31 3DIQ-MOS versus predictions from 2DIQ-MOS of 2D left- and right-views. (a) Blurred Images, predictions by direct averaging the 2DIQ-MOS scores of both views on Waterloo-IVC Phase I. (b) Blurred Images, predictions by the proposed model on Waterloo-IVC Phase I. (c) Blurred Images, predictions by direct averaging 2DIQ-MOS scores of both views on Waterloo-IVC Phase II. (d) Blurred Images, predictions by the proposed model on Waterloo-IVC Phase II. 107

3.32 3DIQ-MOS versus predictions from 2DIQ-MOS of 2D left- and right-views. (a) JPEG Images, predictions by direct averaging the 2DIQ-MOS scores of both views on Waterloo-IVC Phase I. (b) JPEG Images, predictions by the proposed model on Waterloo-IVC Phase I. (c) JPEG Images, predictions by direct averaging 2DIQ-MOS scores of both views on Waterloo-IVC Phase II. (d) JPEG Images, predictions by the proposed model on Waterloo-IVC Phase II. 108

3.33 3DIQ-MOS versus predictions from 2DIQ-MOS of 2D left- and right-views. (a) Mixed Distortion, predictions by direct averaging the 2DIQ-MOS scores of both views on Waterloo-IVC Phase I. (b) Mixed Distortion, predictions by the proposed model on Waterloo-IVC Phase I. (c) Mixed Distortion, predictions by direct averaging 2DIQ-MOS scores of both views on Waterloo-IVC Phase II. (d) Mixed Distortion, predictions by the proposed model on Waterloo-IVC Phase II. 109

4.1 Procedure of the depth comparison test: Art Stereopairs. 119

4.2 Means and standard deviations of depth preference scores in depth comparison test. 119

4.3 Comparison of two subjects' 3DIQ and DQ scores on the Art stereopairs. . 122

4.4 Relationships between (3DIQ and 2DIQ) and (3DIQ and 3D DQ) in Subjective Study I. 123

4.5 The texture images used in Subjective Study II. (a) Bark. (b) Brick. (c) Flower. (d) Food. (e) Grass. (f) Water. 125

4.6 Procedure of generating a symmetrically blurred stereoscopic image in Subjective Study II. 130

4.7 The ground truth disparity maps for different depth levels from the pristine Bark stereopair used in Subjective Study II. Brighter pixels represent higher disparity values. (a) Level 1. (b) Level 2. (c) Level 3. (d) Level 4. (e) Level 5. (f) Level 6. 136

4.8 The estimated disparity maps for different depth levels from the pristine Bark stereopair used in Subjective Study II. Brighter pixels represent higher disparity values. (a) Level 1. (b) Level 2. (c) Level 3. (d) Level 4. (e) Level 5. (f) Level 6. 137

5.1 Sample frames from the pristine videos used in the subjective study in Waterloo-IVC 3D Video Database Phase I. Only the right-views are shown here. (a) Balloons. (b) Book. (c) Kendo. (d) Lovebird. 152

5.2 Sample frames from the pristine videos used in the subjective study in Waterloo-IVC 3D Video Database Phase II. Only the right-views are shown here. (a) Barrier. (b) Craft. (c) Laboratory. (d) Soccer. (e) Tree. (f) Dancer. 153

5.3 The procedure of the subjective test in Waterloo-IVC 3D Video Database Phase I and Phase II. 158

5.4 3DVQ-MOS versus predictions from 2DVQ-MOS of 2D left- and right-views on Waterloo-IVC 3D Video Database Phase II. (a) Average of 2DVQ-MOS. (b) Weighted average of 2DVQ-MOS by the proposed method. 164

5.5 3DVQ-MOS versus predictions from 2D-IQA/VQA estimations of 2D left- and right-views. (a) Average of PSNR on Waterloo-IVC Phase I. (b) Weighted average of PSNR by the proposed method on Waterloo-IVC Phase I. (c) Average of PSNR on Waterloo-IVC Phase II. (d) Weighted average of PSNR by the proposed method on Waterloo-IVC Phase II. 165

5.6 3DVQ-MOS versus predictions from 2D-IQA/VQA estimations of 2D left- and right-views. (a) Average of SSIM on Waterloo-IVC Phase I. (b) Weighted average of SSIM by the proposed method on Waterloo-IVC Phase I. (c) Average of SSIM on Waterloo-IVC Phase II. (d) Weighted average of SSIM by the proposed method on Waterloo-IVC Phase II. 166

5.7 3DVQ-MOS versus predictions from 2D-IQA/VQA estimations of 2D left- and right-views. (a) Average of MS-SSIM on Waterloo-IVC Phase I. (b) Weighted average of MS-SSIM by the proposed method on Waterloo-IVC Phase I. (c) Average of MS-SSIM on Waterloo-IVC Phase II. (d) Weighted average of MS-SSIM by the proposed method on Waterloo-IVC Phase II. . 167

5.8 3DVQ-MOS versus predictions from 2D-IQA/VQA estimations of 2D left- and right-views. (a) Average of IW-SSIM on Waterloo-IVC Phase I. (b) Weighted average of IW-SSIM by the proposed method on Waterloo-IVC Phase I. (c) Average of IW-SSIM on Waterloo-IVC Phase II. (d) Weighted average of IW-SSIM by the proposed method on Waterloo-IVC Phase II. . 168

5.9 3DVQ-MOS versus predictions from 2D-IQA/VQA estimations of 2D left- and right-views. (a) Average of VQM on Waterloo-IVC Phase I. (b) Weighted average of VQM by the proposed method on Waterloo-IVC Phase I. (c) Average of VQM on Waterloo-IVC Phase II. (d) Weighted average of VQM by the proposed method on Waterloo-IVC Phase II. 169

5.10 Diagram of the proposed 2D-to-3D quality prediction model. 176

6.1 Two types of quality assessment of stereoscopic 3D images rendered from MVD representations. ($T_{o,1}$, $D_{o,1}$), ($T_{o,2}$, $D_{o,2}$), ($T_{o,3}$, $D_{o,3}$) and ($T_{d,1}$, $D_{d,1}$), ($T_{d,2}$, $D_{d,2}$), ($T_{d,3}$, $D_{d,3}$) are the (texture, depth) images for 3 views in the original and distorted MVD representations, respectively. ($R_{o,l}$, $R_{o,r}$) and ($R_{d,l}$, $R_{d,r}$) are the (left, right) views of DIBR-synthesized original and distorted stereoscopic 3D images, respectively. Type 1: post-DIBR 3D-IQA; Type 2: pre-DIBR 3D-IQA. 182

6.2 All texture images in MCL-3D database. (a) Balloon. (b) Kendo. (c) Lovebird. (d) Dancer. (e) GtFly. (f) MicroWorld. (g) PoznanHall. (h) PoznanStreet. (i) Shark. 183

6.3 All depth maps in MCL-3D database. (a) Balloon. (b) Kendo. (c) Lovebird. (d) Dancer. (e) GtFly. (f) MicroWorld. (g) PoznanHall. (h) PoznanStreet. (i) Shark. 184

6.4 3D-MOS versus predictions from PSNR of 2D left- and right-views. (a) By different rendering combinations. (b) By different distortion types. 186

6.5 3D-MOS versus predictions from SSIM of 2D left- and right-views. (a) By different rendering combinations. (b) By different distortion types. 186

6.6 3D-MOS versus predictions from MS-SSIM of 2D left- and right-views. (a) By different rendering combinations. (b) By different distortion types. . . . 187

6.7 3D-MOS versus predictions from IW-SSIM of 2D left- and right-views. (a) By different rendering combinations. (b) By different distortion types. . . . 187

6.8 3D-MOS versus predictions from IDW-SSIM of 2D left- and right-views. (a) By different rendering combinations. (b) By different distortion types. . . . 188

6.9 3D-MOS scores for different distortion types and different rendering combinations. (a) Noise. (b) Blur. (c) JPEG. (d) JP2K. (e) Downsample. (f) Transmission Loss. 190

6.10 SSIM maps of the rendered Balloon images for all distortion types at level 4. (a) Rendered, Texture-Depth-Distortion. (b) Rendered, Texture-Distortion-Only. (c) Rendered, Depth-Distortion-Only. 191

6.11 Diagram of the proposed 2D-to-3D quality prediction model. 192

6.12 Information maps of the pristine Balloon texture and depth images. (a) Texture Information Map. (b) Depth Information Map. 193

6.13 SSIM maps of the source Balloon images for all distortion types at level 4. (a) Source, Texture. (b) Source, Depth. 195

6.14 3D-MOS versus predictions from the proposed Pre-DIBR method. (a) By different rendering combinations. (b) By different distortion types. 196

Acronyms

2DIQ 2D image quality.

2DVQ 2D video quality.

3DIQ 3D image quality.

3DQoE 3D Quality-of-Expeirence.

3DVQ 3D video quality.

ACR absolute category rating.

ACR-HR absolute category rating with hidden reference.

BIQA blind image quality assessment.

BIQI Blind Image Quality Index.

BLIINDS-II BLind Image Integrity Notator using DCT-Statistics II.

BRISQUE Blind/Referenceless Image Spatial QUality Evaluator.

CCD charge-coupled device.

CORNIA Codebook Representation for No-Reference Image Assessment.

DIIVINE Distortion Identification-based Image Verity and INtegrity Evalutation.

DPDI Depth Perception Difficulty Index.

DQ Depth quality.

DSCQS double stimulus continuous quality scale.

FR Full-reference.

HVS Human Visual System.

IDW-SSIM Information Content and Distortion Weighted Structural Similarity Index.

IQA image quality assessment.

IW-SSIM Information Content Weighted Structural Similarity Index.

KRCC Kendall's Rank Correlation Coefficient.

LPSI Local Pattern Statistics Index.

MAE Mean Absolute Error.

MOS Mean Opinion Scores.

MS-SSIM Multi-scale Structural Similarity Index.

MSE Mean Squared Error.

MVD Multi-view-plus-depth.

NIQE Naturalness Image Quality Evaluator.

NR No-reference.

OA opinion-aware.

OF opinion-free.

PLCC	Pearson Linear Correlation Coefficient.
PSNR	Peak Signal-to-Noise Ratio.
QAC	Quality-Aware Clustering.
QoE	quality-of-experience.
RMSE	Root Mean Squared Error.
RRIQA	Reduced Reference Image Quality Assessment.
SRCC	Spearman's Rank Correlation Coefficient.
SSCQS	single stimulus continuous quality scale.
SSIM	Structural Similarity Index.
TCLT	Distortion Type Classification and Label Transfer.
UQI	Universal Quality Index.
VC	Visual comfort.
VIF	Visual Information Fidelity.
VQA	video quality assessment.
VQEG	Video Quality Experts Group.
VSNR	Visual Signal-to-noise Ratio.

Chapter 1

Introduction

1.1 Motivation

Over the past decade, we have observed an exponential increase in the demand for 3D image and video services. High-quality 3D movies can now be seen in thousands of new generation 3D theaters all around the world. Meanwhile, 3D TV has become technologically mature and won an increasing market share in the consumption market since 2011, where non-cinematic 3D contents could be obtained from various sources such as Blu-ray 3D and 3D broadcasting [1]. Looking forward, it is expected that mobile phones will be the largest 3D display application on a unit shipment basis in 2018, when 71 million units will have 3D capability [55].

Nevertheless, 3D contents are not universally accepted. Indeed, they are also widely criticized. These criticisms mainly focus on the following aspects: 1) Brightness loss; 2) Fakedness (No 3D effect as it claims); 3)Unnaturalness; 4) Discomfortness (eyestrain, nausea and headaches) and 4) Inconvenience (must wear thick 3D glasses for some 3D displays). It is worth noting that nearly all these criticisms are regarding certain perceptual quality aspects of stereoscopic images/videos.

Automatically assessing the quality of 3D visual experience is a challenging problem [134, 8, 84], especially due to the sophistication and interaction between multiple

3D visual cues including image quality, depth quality and visual comfort [119, 26]. As a result, recent progress on 3D image and video quality assessment remains limited. This lack of successful objective quality assessment methods for 3D visual experience has limited the development of 3D imaging applications and services. Therefore, with the fast development of 3D acquisition, communication, processing and display technologies, in contrast to the slow progress on the understanding of 3D visual quality, automatic quality assessment of 3D images and videos has become ever more important. Consequently, the development of objective visual quality assessment models of 3D images and videos is highly desirable.

1.2 Objectives

The objectives of this research is to investigate various aspects of human visual quality-of-experience (QoE) when viewing stereoscopic 3D images/videos and to develop objective quality assessment models that automatically predict visual QoE of 3D images/videos.

1.3 Contributions

The main contributions of this thesis are summarized as follows:

- A new subjective 3D image quality assessment (IQA) database is developed that has two unique features − the inclusion of both 2D and 3D images that allows for direct observation of the relationship between 2D and 3D quality; and the inclusion of mixed distortion types that allows for studying human behaviors at handling complicated distortion scenarios. A strong distortion type dependent bias is observed when using the direct average of 2D image quality of both views to predict 3D image quality.
- A binocular rivalry inspired multi-scale 2D-to-3D model is proposed to predict the quality of stereoscopic image from that of its single-view images that leads to significantly improved quality prediction performance without explicitly identifying image distortion types.

- A novel subjective 3D depth quality experiment is carried out that is effective at singling out the contributions of stereo cues in depth perception, resulting in several interesting findings regarding distortion type dependency, image content dependency, and the impact of symmetric and asymmetric distortions on the perception of depth.
- A notion of Depth Perception Difficulty Index (DPDI) is proposed based on the percentage of correct and incorrect subject judgements to quantify depth perception induced by stereo cues. A novel computational model for DPDI prediction that leads to highly promising DPDI prediction performance.
- Two new subjective 3D video quality assessment (VQA) databases are developed that have two unique features − the inclusion of both 2D and 3D videos, and the inclusion of asymmetrically compressed stereoscopic 3D videos obtained from mixed-resolution coding, asymmetric transform-domain quantization coding, their combinations, and multiple choices of postprocessing techniques.
- A comparison of different mixed-distortions asymmetric stereoscopic video coding schemes with symmetric coding methods is conducted to verify their potential coding gains. A model to predict the quality of stereoscopic video from that of its single-view videos is proposed, leading to significantly improved quality prediction performance, which can help us predict the coding gain of mixed-distortions asymmetric video compression.
- A novel pre- depth-image-based-rendering (pre-DIBR) 2D-to-3D method based on information content weighting of both texture and depth images is proposed, which demonstrates competitive performance against state-of-the-art IQA models applied post-DIBR.

1.4 Thesis Outline

The outline of this thesis is organized as follows:

Chapter 2 discusses the related work on the topics addressed in the thesis. We starts with a brief introduction about the characteristics of traditional 2D visual data followed by a summary of existing 2D-IQA models. We then perform an overview of well-known 3D visual quality issues followed by a summary of existing 3D subjective and objective IQA/VQA studies.

In Chapter 3, we create a new subjective 3D-IQA database that has two unique features − the inclusion of both 2D and 3D images, and the inclusion of mixed distortion types. Second, we observe strong distortion type dependent bias when using the direct average of 2D image quality of both views to predict 3D image quality. Third, we propose a series of binocular rivalry inspired multi-scale Full-reference (FR) and No-reference (NR) models to predict the quality of stereoscopic images from that of its single-view 2D images. Our results show that the proposed model, without explicitly identifying image distortion types, successfully eliminates the prediction bias, leading to significantly improved quality prediction of stereoscopic 3D images.

In Chapter 4, we carry out two subjective studies on depth perception of stereoscopic 3D images. The first one follows a traditional framework where subjects are asked to rate depth quality directly on distorted stereopairs. The second one uses a novel approach, where the stimuli are synthesized independent of the background image content and the subjects are asked to identify depth changes and label the polarities of depth. Our analysis shows that the second approach is much more effective at singling out the contributions of stereo cues in depth perception. We propose the notion of DPDI and a novel computational model for DPDI prediction. Our results show that the proposed model, without explicitly identifying image distortion types, leads to highly promising DPDI prediction performance.

In Chapter 5, we carry out subjective 3D-VQA experiments on two databases that contain various asymmetrically compressed stereoscopic 3D videos obtained from mixed-resolution coding, asymmetric transform-domain quantization coding, their combinations, and multiple choices of postprocessing techniques. We compare different mixed-distortions asymmetric stereoscopic video coding schemes with symmetric coding methods and verify their potential coding gains. We observed a strong systematic bias when using direct averaging of 2D video quality of both views to predict 3D video quality. We propose a model to account for the prediction bias, leading to significantly improved quality predictions of stereoscopic videos. We show that the proposed model can help us predict the coding gain of mixed-distortions asymmetric video compression.

In Chapter 6, we investigate the problem of objective quality assessment of Multi-view-plus-depth (MVD) images, with a main focus on the pre-DIBR case. We find that although existing IQA methods can be applied post-DIBR to provide reasonable quality prediction of MVD images, they are difficult to be employed as a guiding criterion in the optimization of MVD video coding and transmission systems. We propose a novel pre-DIBR method based on information content weighting of both texture and depth images, which demonstrates competitive performance against state-of-the-art IQA models applied post-DIBR.

Finally, Chapter 7 concludes the thesis and discusses different avenues for future research.

Chapter 2

Background

This chapter starts with a brief introduction about the characteristics of traditional 2D visual data followed by a summary of existing 2D-IQA models. We then overview well-known 3D visual quality issues followed by a summary of existing 3D subjective and objective IQA/VQA studies. The review focuses on the most relevant results in the literature.

2.1 2D Image Quality Assessment

Natural photographical images refer to the visual data obtained by a camera − these include pictures of physical scenes, man-made objects and natural environments. The "amount" of incoming photons entering a camera, through an open aperture, is recorded on an array of charge-coupled device (CCD) receptors. The analog values measured in the form of difference of voltage are converted to digital form using an analog-to-digital converter. The digital data is then transformed into an array of pixels (picture elements). The current state of research in the area of natural image statistics has not yet been able to find a comphrehensive natural image model. The main observations include:

- second-order pixel correlations [39];
- importance of phases [18];

- heavy-tail non-Gaussian marginals in wavelet domain [27, 30];
- near elliptical shape of joint densities in wavelet domain [129];
- decay of dependency in wavelet domain [80];

A common assumption in computational vision is that the Human Visual System (HVS) is optimized for processing the spatial information in natural visual images [98]. In the classical approach for image quality assessment, researchers use a bottom-up approach that attempts to emulate the computational system of the HVS in order to reach a realistic model of image quality perception [31, 77]. A top-down philosophy, towards image quality assessment, makes hypotheses about the overall functionality of the HVS. The main purpose of such an approach is to use a simpler solution by treating HVS as a black-box and concentrating only on its input-output relationship [125, 153]. The desirable properties in an image and video quality assessment method include:

- high correlation with subjective opinions;
- low computational complexity;
- accurate local quality prediction that can help determine local quality variations;
- good mathematical properties, for example, a valid distance metric, convexity, differentiability, symmetry, etc;

which are no doubt desirable properties for 3D image and video quality measures also.

Reviews of existing 2D-IQA methods are given in [150]. Among existing methods, the best known and most widely used models are Mean Squared Error (MSE)/Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index (SSIM) [153]. The remainder of this section introduces MSE/PSNR and SSIM and also provides their comparison with subjective scores.

2.1.1 Mean Squared Error and Peak Signal-to-Noise Ratio

The goal of an image/video fidelity measure is to provide a quantitative comparison between two images/videos, where one of the image/video is considered pristine or treated as a reference that has perfect quality. The most widely used method to measure image/video fidelity is PSNR, a monotonic function of MSE. The MSE between two images $\mathbf{y}$ and $\hat{\mathbf{y}}$ is

$$\mathrm{MSE}(\mathbf{y}, \hat{\mathbf{y}}) = \frac{1}{L_1 L_2} \sum_{i_1=1}^{L_1} \sum_{i_2=1}^{L_2} (\mathbf{y}(i_1, i_2) - \hat{\mathbf{y}}(i_1, i_2))^2, \tag{2.1}$$

where L_1 and L_2 are the height and the width of the images, respectively. As MSE computation is based on the error signal, $\mathbf{e} = \mathbf{y} - \hat{\mathbf{y}}$, between the reference image, $\mathbf{y}$, and its distorted version, $\hat{\mathbf{y}}$, it can be regarded as a measure of image quality degradation. In image and video processing literature, MSE is often converted to PSNR using the expression:

$$\mathrm{PSNR}(\mathbf{y}, \hat{\mathbf{y}}) = 10 \log_{10} \frac{R^2}{\mathrm{MSE}(\mathbf{y}, \hat{\mathbf{y}})} \tag{2.2}$$

where R is the dynamic range of image pixel intensities e.g. for an 8-bit/pixel gray-scale image, $R = 2^8 - 1 = 255$. The only advantage of PSNR over MSE, as a perceptual quality measure, is its capability to provide normalized meaning for images with different dynamic ranges.

2.1.2 Structural Similarity

The SSIM index is a method for measuring the similarity between two images that has shown superior performance over MSE in predicting visual quality of 2D images [151]. SSIM and its derivations have been applied to a broad range of applications, ranging from image restoration and compression, to visual communication and pattern recognition [151].

The basic spatial domain SSIM algorithm is based upon separated comparisons of local luminance, contrast and structure between an original and a distorted images. Given two local image patches $\mathbf{x} = \{x_i | i = 1, 2, \cdots, L\}$ and $\mathbf{y} = \{y_i | i = 1, \cdots, L\}$ extracted from the original and distorted images, respectively, the luminance, contrast and structural similarities between them are evaluated as

$$l(\mathbf{x}, \mathbf{y}) = \frac{2\mu_x\mu_y + C_1}{\mu_x^2 + \mu_y^2 + C_1} \tag{2.3}$$

$$c(\mathbf{x}, \mathbf{y}) = \frac{2\sigma_x\sigma_y + C_2}{\sigma_x^2 + \sigma_y^2 + C_2} \tag{2.4}$$

$$s(\mathbf{x}, \mathbf{y}) = \frac{\sigma_{xy} + C_3}{\sigma_x\sigma_y + C_3} \tag{2.5}$$

respectively. Here, μ_x, σ_x and σ_{xy} represent the mean, standard deviation and cross-correlation evaluations, respectively. $C_1 = (K_1L)^2$, $C_2 = (K_2L)^2$, $C_3 = C_2/2$ are small constants that have been found to be useful in characterizing the saturation effects of the visual system at low luminance and contrast regions and stabilizing the performance of the measure when the denominators are close to zero. The local SSIM index is defined as the product of the three components, which gives

$$\text{SSIM}_{\text{local}} = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)} \tag{2.6}$$

When this local measurement is applied to an entire image using a sliding window approach, an SSIM quality map is created. The overall SSIM value of the whole image is simply the average of the SSIM map.

It has been found that the performance of the direct single-scale SSIM algorithm depends upon the scale it is applied to [158]. In [158], a Multi-scale Structural Similarity Index (MS-SSIM) approach was proposed that incorporates SSIM evaluations at different scales. Psychovisual experiments were carried out to find the relative weights between scales. Interestingly, the measured weight function peaks at middle-resolution scales and drops at both low- and high-resolution scales, consistent with the contrast sensitivity function extensively studied in the vision literature [141]. Let $\mathbf{x}_{j,i}$ and $\mathbf{y}_{j,i}$ be the ith local image patches (extracted from the ith evaluation window) at the jth scale, and let N_j be the number of evaluation windows in the scale, then the jth scale SSIM evaluation is computed as

$$\mathrm{SSIM}_j = \frac{1}{N_j}\sum_i c(\mathbf{x}_{j,i}, \mathbf{y}_{j,i}) s(\mathbf{x}_{j,i}, \mathbf{y}_{j,i}) \tag{2.7}$$

for $j = 1, \cdots, M-1$, and

$$\mathrm{SSIM}_j = \frac{1}{N_j}\sum_i l(\mathbf{x}_{j,i}, \mathbf{y}_{j,i}) c(\mathbf{x}_{j,i}, \mathbf{y}_{j,i}) s(\mathbf{x}_{j,i}, \mathbf{y}_{j,i}) \tag{2.8}$$

for $j = M$. The overall MS-SSIM measure is defined as

$$\text{MS-SSIM}_j = \prod_{j=1}^{M} (\mathrm{SSIM}_j)^{\beta_j} \tag{2.9}$$

where the β_j values were obtained through psychophysical measurement [158].

By combining information content weighting with MS-SSIM, an Information Content Weighted Structural Similarity Index (IW-SSIM) was obtained [154]. Let $w_{j,i}$ be the information content weight [154] computed at the ith spatial location in the jth scale, the jth scale IW-SSIM measure is defined as

$$\text{IW-SSIM}_j = \frac{\sum_i w_{j,i} c(\mathbf{x}_{j,i}, \mathbf{y}_{j,i}) s(\mathbf{x}_{j,i}, \mathbf{y}_{j,i})}{\sum w_{j,i}} \tag{2.10}$$

for $j = 1, \cdots, M-1$, and

$$\text{IW-SSIM}_j = \frac{1}{N_j}\sum_i l(\mathbf{x}_{j,i}, \mathbf{y}_{j,i}) c(\mathbf{x}_{j,i}, \mathbf{y}_{j,i}) s(\mathbf{x}_{j,i}, \mathbf{y}_{j,i}) \tag{2.11}$$

for $j = M$. The final overall IW-SSIM measure is then computed as

$$\text{IW-SSIM}_j = \prod_{j=1}^{M} (\text{IW-SSIM}_j)^{\beta_j} \tag{2.12}$$

using the same set of scale weights β_j's as in MS-SSIM.

SSIM, MS-SSIM and IW-SSIM algorithms do not involve any training process or any new parameters for tuning. All parameters are defined in [153, 158]. These include $K_1 = 0.01$ and $K_2 = 0.03$ from [153]; $M = 5$ from [158]; and the fine-to-coarse scale weights $\{\beta_1, \beta_2, \beta_3, \beta_4, \beta_5\} = \{0.0448, 0.2856, 0.3001, 0.2363, 0.1333\}$ from [158].

2.1.3 Subjective 2D-IQA Studies and Objective Evaluations

The ultimate goal of a 2D-IQA algorithm is to predict subjective quality scores of 2D images. To the best of our knowledge, there are currently seven publicly available subject-rated image databases that are widely recognized in the IQA research community. These databases include:

- The Cornell-A57 database [22] was created at Cornell University. It contains 54 distorted images with six types of distortions including a) quantization of the LH subbands of a 5-level discrete wavelet transform, where the subbands were quantized via uniform scalar quantization with step sizes chosen such that the root mean-squared contrast of the distortions was equal; b) additive Gaussian white noise; c) baseline JPEG compression; d) JPEG2000 compression without visual frequency weighting; e) JPEG2000 compression with the dynamic contrast-based quantization algorithm, which applies greater quantization to the fine spatial scales relative to the coarse scales in an attempt to preserve global precedence; and f) blurring by using a Gaussian filter.

- The Categorical Image Quality (CSIQ) Database [71] was developed at Oklahoma State University. 30 original images were used to create a total of 866 distorted images using six types of distortions at four to five distortion levels. The distortion types include JPEG compression, JPEG2000 compression, global contrast decrements, additive pink Gaussian noise, and Gaussian blurring.

- The IVC database [93, 92] was developed at Ecole Polytechnique de l'Universit'e de Nantes. It includes 185 distorted images generated from 10 pristine images. There are four types of distortions that are a) JPEG compression; b) JPEG2000 compression; c) Local adaptive resolution (LAR) coding; and d) Blurring.

- The LIVE database [124] was developed at The University of Texas at Austin. It contains seven data sets of 982 subject-rated images, including 779 distorted images created from 29 pristine images with five types of distortions at different distortion levels. The distortion types include a) JPEG2000 compression (2 sets); b) JPEG compression (2 sets); c) White noise contamination (1 set); d) Gaussian blur (1 set); and e) fast fading channel distortion of JPEG2000 compressed bitstream (1 set). The subjective test was carried out with each data set individually. A cross-comparison set that mixes images from all distortion types is then used to help align the subject scores across data sets. The subjective scores of all images are then adjusted accordingly. The alignment process is rather crude. However, the aligned subjective scores (all data) are still very useful references, which are particularly important for testing general-purpose IQA algorithms, for which cross-distortion comparisons are highly desirable.

- The Tampere Image Database 2008 (TID2008) [103, 104] was developed with a joint international effort between Finland, Italy, and Ukraine. It includes 1700 distorted images generated from 25 pristine images with 17 distortion types at four distortion levels. The types of distortions include: a) Additive Gaussian noise; b) Additive noise in color components is more intensive than additive noise in the luminance component; c) Spatially correlated noise; d) Masked noise; e) High frequency noise; f) Impulse noise; g) Quantization noise; h) Gaussian blur; i) Image denoising; j) JPEG compression; k) JPEG2000 compression; l) JPEG transmission errors; m) JPEG2000 transmission errors; n) Non eccentricity pattern noise; o) Local block-wise distortions of different intensity; p) Mean shift (intensity shift); and q) Contrast change.

- The Tampere Image Database 2013 (TID2013) [106, 105] was developed with a joint international effort between Finland, France, Italy, Ukraine and USA. It includes 3000 distorted images generated from 25 reference images with 24 distortion types at five distortion levels. The types of distortions include: a) Additive Gaussian noise; b) Additive noise in color components is more intensive than additive noise in the luminance component; c) Spatially correlated noise; d) Masked noise; e) High frequency noise; f) Impulse noise; g) Quantization noise; h) Gaussian blur; i) Image denoising;

j) JPEG compression; k) JPEG2000 compression; l) JPEG transmission errors; m) JPEG2000 transmission errors; n) Non eccentricity pattern noise; o) Local block-wise distortions of different intensity; p) Mean shift (intensity shift); q) Contrast change; r) Change of color saturation; s) Multiplicative Gaussian noise; t) Comfort noise; u) Lossy compression of noisy images; v) Image color quantization with dither; w) Chromatic aberrations and x) Sparse sampling and reconstruction.

- The Toyama-MICT database [54] was created at University of Toyama. It contains 196 images, including 168 distorted images generated by JPEG and JPEG2000 compression.

Five evaluation metrics are widely used to compare the performance of IQA measures. Some of the metrics were included in previous tests carried out by the Video Quality Experts Group (VQEG) [46]. Other metrics are adopted from previous publications [103, 126]. These evaluation metrics are Pearson Linear Correlation Coefficient (PLCC), Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), Spearman's Rank Correlation Coefficient (SRCC), and Kendall's Rank Correlation Coefficient (KRCC).

- PLCC after a nonlinear mapping between the subjective and objective scores. For the ith image in an image database of size N, given its subjective score o_i mean opinion scores (MOS) or difference of MOS between reference and distorted images) and its raw objective score r_i, we first apply a nonlinear function to r_i given by [126]

$$q(r) = a_1\{\frac{1}{2} - \frac{1}{1+\exp[a_2(r-a_3)]}\} + a_4 r + a_5\,, \tag{2.13}$$

  where a_1 to a_5 are model parameters found numerically using a nonlinear regression process in MATLAB®optimization toolbox to maximize the correlations between subjective and objective scores. The PLCC value can then be computed as

$$\text{PLCC} = \frac{\sum_i (q_i - \bar{q}) \times (o_i - \bar{o})}{\sqrt{\sum_i (q_i - \bar{q})^2 \times (o_i - \bar{o})^2}}\,. \tag{2.14}$$

- MAE is calculated using the converted objective scores after the nonlinear mapping described previously

$$\mathrm{MAE} = \frac{1}{N}\sum |q_i - o_i| \,. \tag{2.15}$$

- RMSE is computed similarly as

$$\mathrm{RMSE} = \sqrt{\frac{1}{N}\sum |q_i - o_i|^2} \,. \tag{2.16}$$

- SRCC is defined as:

$$\mathrm{SRCC} = 1 - \frac{6\sum_{i=1}^{N} d_i^2}{N(N^2-1)} \,, \tag{2.17}$$

  where d_i is the difference between the ith image's ranks in subjective and objective evaluations. SRCC is a nonparametric rank-based correlation metric, independent of any monotonic nonlinear mapping between subjective and objective scores.

- KRCC is another nonparametric rank correlation metric given by

$$\mathrm{KRCC} = \frac{N_c - N_d}{\frac{1}{2}N(N-1)} \,, \tag{2.18}$$

where N_c and N_d are the numbers of concordant and discordant pairs in the data set, respectively.

Among the aforementioned metrics, PLCC, MAE and RMSE are adopted to evaluate prediction accuracy [46], and SRCC and KRCC are employed to assess prediction monotonicity [46]. A better objective IQA measure should have higher PLCC, SRCC, and KRCC while lower MAE and RMSE values.

The performance comparison of PSNR, SSIM, MS-SSIM and IW-SSIM on the above-mentioned seven 2D-IQA databases is conducted and reported in Table 2.1. We can observe that SSIM and its variants perform significantly better than PSNR, in predicting subjective scores of all the databases. As a result, we can conclude that SSIM is a better 2D perceptual quality measure as compared to MSE and a new 3D perceptual quality measure may be built upon SSIM and its variants.

Table 2.1: Performance comparison of 2D-IQA models on 2D-IQA databases

	Cornell A57 Database					CSIQ Database				
2D-IQA	PLCC	SRCC	KRCC	RMSE	MAE	PLCC	SRCC	KRCC	RMSE	MAE
PSNR	0.6346	0.6176	0.4301	0.1900	0.1607	0.7605	0.8053	0.6081	0.1705	0.1339
SSIM	0.8019	0.8067	0.6063	0.1469	0.1209	0.8569	0.8718	0.6858	0.1354	0.1007
MS-SSIM	0.8604	0.8415	0.6483	0.1253	0.1007	0.8903	0.9061	0.7298	0.1196	0.0897
IW-SSIM	0.9035	0.8713	0.6846	0.1054	0.0892	0.9017	0.9108	0.7390	0.1135	0.0841
	IVC Database					LIVE Database				
2D-IQA	PLCC	SRCC	KRCC	RMSE	MAE	PLCC	SRCC	KRCC	RMSE	MAE
PSNR	0.6719	0.6884	0.5218	0.9023	0.7191	0.9302	0.9092	0.7484	8.4819	6.3731
SSIM	0.9119	0.9018	0.7223	0.4999	0.3777	0.9455	0.9496	0.8149	7.5252	6.0247
MS-SSIM	0.9108	0.8980	0.7203	0.5029	0.3813	0.9468	0.9512	0.8181	7.4379	5.9897
IW-SSIM	0.9231	0.9125	0.7339	0.4686	0.3694	0.9515	0.9604	0.8379	7.1116	5.6548
	TID 2008 Database					TID 2013 Database				
2D-IQA	PLCC	SRCC	KRCC	RMSE	MAE	PLCC	SRCC	KRCC	RMSE	MAE
PSNR	0.5235	0.5531	0.4027	1.1434	0.8680	0.6775	0.6394	0.4699	0.9119	0.6801
SSIM	0.7732	0.7749	0.5768	0.8511	0.6546	0.7895	0.7417	0.5588	0.7608	0.5926
MS-SSIM	0.8451	0.5768	0.6568	0.7173	0.5578	0.8329	0.7859	0.6047	0.6861	0.5309
IW-SSIM	0.8579	0.8511	0.6636	0.6895	0.5276	0.8319	0.7779	0.5977	0.6880	0.5290
	Toyama-MICT Database									
2D-IQA	PLCC	SRCC	KRCC	RMSE	MAE					
PSNR	0.7328	0.7221	0.5398	0.8983	0.6999					
SSIM	0.9125	0.9023	0.7250	0.5401	0.4030					
MS-SSIM	0.9154	0.9070	0.7316	0.5315	0.3995					
IW-SSIM	0.9401	0.9289	0.7693	0.4502	0.3437					

2.2 3D Visual Quality Factors

In this section, we will give a brief overview of widely recognized 3D visual quality factors. More specific reviews for subjective 3D-IQA studies and objective 3D-IQA studies will be presented in Section 2.3 and Section 2.4, respectively.

2.2.1 Impact of Eye Dominance

Eye dominance is a common visual phenomenon, referring to the tendency to prefer the input from one eye to the other, depending on the human subject [67]. Several studies were devoted to identify the impact of eye dominance on image quality. In [119, 63, 81], no impact of eye dominance was noticed for image quality assessment. Nevertheless, in [128], it was observed that eye dominance improves the performance of visual search task by aiding visual perception in binocular vision, and the eye dominance effect in 3D perception and asymmetric view coding was also analyzed.

2.2.2 Impact of Image Distortions

Image distortions in stereoscopic images can be grouped into two categories based on their dimensionality: monocular (bi-dimensional) image distortions and binocular (three-dimensional) image distortions. Typical monocular distortions include noise, blurriness, blockiness, transmission errors, mean shift, contrast change and so forth [104], while typical stereoscopic distortions includes crosstalk, keystone distortion, depth-plane curvature, puppet theater effect, crosstalk, cardboard effect, shear distortion, picket-fence effect, and image flipping [84, 32]. In this work, we will focus on the impact of monocular distortions. A short review and discussion on existing findings will be given in Section 2.5.

2.2.3 Impact of Depth Perception

Depth quality is no doubt an essential aspect of human QoE when viewing stereoscopic 3D images. Existing studies on the topic appear to be unconclusive, limited, and sometimes conflicting. In [53], it was reported that the perceived depth performance cannot always be predicted from displaying image geometry alone, while other system factors, such as software drivers, electronic interfaces, and individual participant differences, may also play significant roles. In [119, 26], subjective studies suggested that increasing the degree of binocular depth does increase the perceived depth quantity. In [119, 64], it was suggested that depth quality may need to be considered independently from perceived 3D image quality. The results in [119] showed that increased JPEG coding has no effect on depth perception however a negative effect on image quality. In [26, 137], subjective studies suggested that 3D image quality is not sensitive to variations in the degree of binocular depth.

Other studies pointed out perceptual depth quality as an important component in the holistic 3D QoE. In [177], a blurring filter, where the level of blur depends on the depth of the area where it is applied, is used to enhance the viewing experience. In [116], subjective studies revealed that humans tend to prefer DCT compressed stereopairs over the monoscopic single-views even though the blocking artifacts are annoying. In [69], depth naturalness is shown to be a useful ingredient in the assessment of 3D video QoE.

Similarly, in [56], the added value of depth naturalness has been verified for pristine and blurred stereoscopic images. In [70], stimuli with various stereo depth and image quality were evaluated subjectively in terms of naturalness, viewing experience, image quality, and depth perception, and the experimental results suggested that the overall 3D QoE is approximately 75% determined by image quality and 25% by perceived depth. In [24], Chen *et al.* showed that subjective evaluation of depth quality has a low correlation with that of 3D image quality and verified that the overall 3D QoE can be predicted using a single linear model from 3D image quality and depth quality.

Meanwhile, several studies have been proposed to objectively predict perceived depth quality and subsequently to predict 3D quality by combining depth quality and 2D image quality. A comprehensive review and discussion on existing objective depth quality assessment methods will be given in Section 4.1.

2.2.4 Impact of Visual Discomfort

In the literature, several factors have been found to affect visual comfort negatively, which can be grouped into five categories: (a) accommodation-vergence conflict; (b) parallax distribution; (c) binocular mismatches; (d) depth inconsistencies; and (e) cognitive inconsistencies [136].

In [120], subjective studies showed that increased JPEG compression has a negative effect on visual discomfort, In [24], Chen *et al.* showed that there are diverse opinions in interpreting visual discomfort across different human subjects.

Meanwhile, several studies have been conducted to develop objective quality metrics on predicting visual discomfort. In [68], stereoscopic impairments caused by inappropriate shooting parameters or camera misalignment were detected and the induced excessive horizontal and vertical disparities were measured as a visual discomfort prediction metric. In [29], Choi *et al.* proposed a visual fatigue evaluation method using spatial and temporal complexities, depth position and scene movement of the 3D video. In [28], a visual discomfort prediction model was built on principal component analysis of visual discomfort features including spatial complexity, depth position, temporal complexity, scene movement, crosstalk, brightness, and different degrees of brightness. In [62], Jung *et al.* developed a visual comfort metric that quantifies the level of visual discomfort by fast salient object motion information. In [97], a model-based neuronal and statistical framework was developed to automatically predict visual discomfort be employing two types of features: 1) coarse features derived from the statistics of binocular disparities and 2) fine features derived by estimating the neural activity associated with the processing of horizontal disparities. In [96], a 3D visual discomfort prediction model that accounts for both accommodation and vergence was proposed and shown to perform superior to prior models that rely on a computed disparity distribution only.

2.2.5 Overall 3D Quality-of-Experience

In [119], it was hypothesized that the 3D QoE can be predicted by combining image quality and depth quality. Several studies have been devoted to explore the effect of each 3D visual criterion on the overall 3D QoE. In [70], stimuli with various stereo depth (camera base distances and screen disparity) and various image quality (white Gaussian noise and Gaussian blur) were subjectively evaluated in terms of naturalness, viewing experience, image quality, and depth perception. The experimental results showed that naturalness is the most appropriate concept to evaluate 3D quality of stereoscopic images, which is approximately 75% determined by image quality and approximately 25% by perceived depth. In [24], Chen *et al.* conducted a study on the human perception of image quality, depth quality, visual comfort and overall 3D QoE and verified that the overall 3D QoE can be predicted using a single linear model from 3D image quality and depth quality.

2.3 Subjective 3D Image/Video Quality Assessment

2.3.1 Subjective 3D-IQA Studies

To the best of our knowledge, there are currently 9 subject-rated image databases that are widely recognized in the 3D-IQA research community. Table 2.2 lists these databases with detailed descriptions. Among them LIVE 3D Image Quality Database Phase I, LIVE 3D Image Quality Database Phase II, IRCCyN/IVC 3D Images Database, Ningbo University 3D Image Quality Assessment Database Phase I and Phase II, MCL 3D Image Database, and MMSPG 3D Image Quality Assessment Database are publicly available.

- LIVE 3D Image Quality Database Phase I [90] was developed at The University of Texas at Austin. The LIVE 3D Image Quality Database Phase I only includes symmetrically distorted stereoscopic images. It contains a data set of 385 subject-rated stereoscopic images, including 365 distorted images created from 20 pristine images with five types of distortions at different distortion levels. The distortion types include a) JPEG2000 compression; b) JPEG compression; c) White noise contamination; d) Gaussian blur; and e) fast fading channel distortion of JPEG2000 compressed bitstream. The resolution of the images is 640 × 360 pixels. A single stimulus continuous quality scale (SSCQS) protocol with hidden reference was adopted in the study. The continuous quality scales are labeled by equally spaced adjective terms: bad, poor, fair, good, and excellent. The study was conducted using a iZ3D 22-inch 3D monitor with a native resolution of 800 × 600 and a passive polarized 3D glasses. The viewing distance was four times the screen height. Thirty-two subjects with a male-majority population participated in the experiments.
- LIVE 3D Image Quality Database Phase II [23] was developed at The University of Texas at Austin. The LIVE 3D Image Quality Database Phase II includes both symmetrically and asymmetrically distorted stereoscopic images. It contains a data set of 368 subject-rated stereoscopic images, including 360 distorted images created from 8 pristine images with five types of distortions at different distortion levels. The distortion types include a) JPEG2000 compression; b) JPEG compression; c) White noise contamination; d) Gaussian blur; and e) fast fading channel distortion of JPEG2000 compressed bitstream. For each distortion type, every reference stereopair was processed to create 3 symmetrically distorted stereopairs and 6 asymmetrically distorted stereopairs. The resolution of the images is 640 × 360 pixels. A SSCQS protocol with hidden reference was adopted in the study. The continuous quality scales are labeled by equally spaced adjective terms: bad, poor, fair, good, and

excellent. The study was conducted using a Panasonic 58-inch 3D TV (TC-P58VT25) with a native resolution of 1920 × 1080 and an active shutter glasses. The viewing distance was four times the screen height. Thirty-three subjects participated in the experiments, six females and twenty-seven males, aged from 22 to 42 years.

- IRCCyN/IVC 3D Images Database [10] was developed at Ecole Polytechnique de l'Université de Nantes. The IRCCyN/IVC 3D Images Database only includes symmetrically distorted stereoscopic images. It contains a data set of 96 subject-rated stereoscopic images, including 90 distorted images created from 6 pristine images with three types of distortions at different distortion levels. The distortion types include a) JPEG2000 compression; b) JPEG compression; c) Gaussian blur. The mean resolution of the images is 512 × 448 pixels. A multistimuli continuous quality scale experiment with explicit and hidden references, namely as SAMVIQ, was adopted in the study. The study was conducted using a Samsung SyncMaster 21" 1100 MB monitor with a native resolution of 1024 × 768 and an active shutter glasses. The viewing distance was three meters. Seventeen subjects participated in the experiments, mostly males familiar with subjective quality tests, with an average of 28.2 years and a standard deviation of 6.7.

- MICT 3D Image Quality Evaluation Database [113] was developed at University of Toyama. The MICT 3D Image Quality Evaluation Database includes both symmetrically and asymmetrically distorted stereoscopic images. It contains a data set of 500 subject-rated stereoscopic images, including 490 distorted images created from 10 pristine images with only one type of distortions at different distortion levels. The distortion type is JPEG compression. The resolutions of the images is 640 × 480 pixels. A SSCQS protocol with hidden reference was adopted in the study. The continuous quality scales are labeled by equally spaced adjective terms: bad, poor, fair, good, and excellent. The study was conducted using an autostereoscopic (SANYO) display. Twenty-four non-expert subjects participated in the experiments, twelve females and twelve males, aged from 19 to 32 years.

- Ningbo University 3D Image Quality Assessment Database Phase I [148] was developed at Ningbo University. The Ningbo University 3D Image Quality Assessment

Database Phase I only includes asymmetrically distorted stereoscopic images (right-view distortion only, left-view pristine). It contains a data set of 410 subject-rated stereoscopic images, including 400 distorted images created from 10 pristine images with four types of distortions at different distortion levels. The distortion types include a) JPEG2000 compression; b) JPEG compression; c) White noise contamination; d) Gaussian blur. The resolution of the images is from 1252 × 1110 to 1390 × 1110 pixels. A double stimulus continuous quality scale (DSCQS) protocol was adopted in the study. The continuous quality scales are labeled by equally spaced adjective terms: bad, poor, fair, good, and excellent. The study was conducted by a linear polarization stereoscopic display system. Twenty non-expert subjects participated in the experiments, aged from 20 to 25 years with an average of 23.

- Ningbo University 3D Image Quality Assessment Database Phase II [175] was developed at Ningbo University. The NingBo University 3D Image Quality Assessment Database Phase II only includes symmetrically distorted stereoscopic images. It contains a data set of 324 subject-rated stereoscopic images, including 312 distorted images created from 12 pristine images with five types of distortions at different distortion levels. The distortion types include a) JPEG2000 compression; b) JPEG compression; c) White noise contamination; d) Gaussian blur; e) H.264 compression. A DSCQS protocol was adopted in the study. The continuous quality scales are labeled by equally spaced adjective terms: bad, poor, fair, good, and excellent. The resolution of the images is from 480 × 270 to 1024 × 768 pixels. The study was conducted by a linear polarization stereoscopic display system. The viewing distance was 2 to 2.5 times the screen height. Twenty-six non-expert subjects participated in the experiments, aged from 20 to 25 years.

- Tianjin University 3D Image Quality Assessment Database [164] was developed at Tianjin University. The Tianjin University 3D Image Quality Assessment Database only includes symmetrically distorted stereoscopic images. It contains a data set of 300 subject-rated stereoscopic images, including 270 distorted images created from 30 pristine images with three types of distortions at different distortion levels. The distortion types include a) JPEG2000 compression; b) JPEG compression; c) Gaussian blur. The resolution of the images is from 320 × 240 to 1024 × 768 pixels. A DSCQS protocol was adopted in the study. The study was conducted using a 3D FreeEyE (4210) display with a native resolution of 1380 × 768 and a Philips display with a native resolution of 1024 × 768.

- MMSPG 3D Image Quality Assessment Database [43] was developed at MULTIMEDIA SIGNAL PROCESSING GROUP MMSPG. The MMSPG 3D Image Quality Assessment Database contains a data set of 100 subject-rated stereoscopic images including 10 different scenes. For each of the scenes 6 different stimuli have been considered corresponding to different camera distances. The resolution of the images is 1920 × 1080 pixels. A SSCQS protocol was adopted in the study. The continuous quality scales are labeled by equally spaced adjective terms: bad, poor, fair, good, and excellent. The study was conducted using a 46-inch polarized stereoscopic display (Hyundai S465D) with a native resolution of 1920 × 1080 pixels. The viewing distance is approximatively two meters which is equal to the height of the screen multiplied by factor 3. Seventeen non-expert subjects participated in the test, one female and sixteen males, aged from 22 to 53 years with an average of 30.

- MCL 3D Image Database [132] was developed at University of Southern California. The MCL 3D Image Database contains a data set of 693 subject-rated stereoscopic images from 9 pristine image-plus-depth source contents with seven types of distortions at different distortion levels. The distortion types include a) JPEG2000 compression; b) JPEG compression; c) White noise contamination; d) Gaussian blur; e)Transmission error; and f) Rendering distortions. The resolution of the images is 1024 × 768 and 1920 × 1088 pixels. A pairwise comparison experiment was adopted in the subjective study. The study was conducted using a 46.9-inch polarized stereoscopic display (LG 47LW5600) with a native resolution of 1920 × 1080 pixels. The viewing distance was 3.2 meters away from the display screen. 270 subjects participated in the experiments, 100 females and 170 males, aged mostly from 21 to 30 years. Among them, 34 (or 13%) were experts and 236 (or 87%) were non-experts.

Subjective data is essential in understanding the impact of various distortions on the perceptual quality of stereoscopic images. Ideally, we would need a complete set of sub-

Table 2.2: Summary of existing 3D image quality databases

Database	Subjects	Protocol	Display	Images	Resolutions	Distortions
LIVE Phase I [90]	32	SSCQS	Passive	385	640 × 360	JPEG2000, JPEG, white noise, gaussian blur, fast fading
LIVE Phase II [23]	33	SSCQS	Active	368	640 × 360	JPEG2000, JPEG, white noise, gaussian blur, fast fading
IRCCyN/IVC [10]	17	SAMVIQ	Active	96	512 × 448	JPEG2000, JPEG, gaussian blur
MICT [113]	24	SSCQS	Auto	500	640 × 480	JPEG
Ningbo Phase I [148]	20	DSCQS	Passive	410	1252×1110 to 1390 × 1110	JPEG2000, JPEG, white noise, gaussian blur
Ningbo Phase II [175]	26	DSCQS	Passive	324	480 × 270 to 1024 × 768	JPEG2000, JPEG, white noise, gaussian blur, H.264 compression
Tianjin [164]	N/A	DSCQS	Auto	300	320 × 240 to 1024 × 768	JPEG2000, JPEG, white noise
MMSPG [43]	17	SSCQS	Passive	100	1920 × 1080	Different camera distances
MCL [132]	270	pairwise	Passive	693	1024 × 768 & 1920 × 1080	JPEG2000, JPEG, white noise, gaussian blur, transmission loss, rendering distortions

jective test on an image database that contains both 2D (single-view) and stereoscopic 3D images, both symmetrically and asymmetrically distorted images at different distortion levels, as well as both single- and mixed-distortion images. The above-mentioned existing 3D image quality databases are highly valuable but limited in one aspect or another. Specifically, IRCCyN/IVC 3D Images Database, Tianjin University Database, Ningbo University Database Phase II, and LIVE 3D Image Quality Database Phase I only include symmetrically distorted stereoscopic images. Ningbo University Database Phase I only includes asymmetrically distorted stereoscopic images. MICT 3D Image Quality Evaluation Database contains both cases but only for JPEG compressed images. The most recent LIVE 3D Image Quality Database Phase II includes both symmetric and asymmetric cases as well as five distortion types. Unfortunately, 2D-IQA of single-view images are missing, making it difficult to directly examine the relationship between the perceptual quality of single-views and stereoscopic images. In addition, asymmetric distortions with mixed distortion types are missing in all existing databases, making it hard to validate the generalization capability of 3D quality prediction models.

2.3.2 Subjective 3D-VQA Studies

To the best of our knowledge, there are currently 6 subject-rated video databases that are commonly recognized in the 3D-VQA research community. Table 2.3 lists these databases with detailed descriptions.

- LIVE 3D Video Quality Database [24] was developed at The University of Texas at Austin. The LIVE 3D Video Quality Database contains a data set of 54 subject-rated stereoscopic videos created from 6 pristine videos. The resolution of the videos is 720 × 480 pixels. Two of these videos are fifteen seconds long, while the rest are ten seconds long. The frame rate of the videos is 25 frames per second. H.264/AVC coding was chosen as the compression method and every reference stereoscopic video was processed to create 3 symmetrically compressed stereoscopic videos and 6 asymmetrically compressed stereoscopic videos. A SSCQS experiment with hidden reference protocol was adopted in the study. The continuous quality scales are labeled by equally spaced adjective terms: bad, poor, fair, good, and excellent. The spatial quality, depth quality, visual comfort and overall 3D quality were graded individually. The study was conducted using an Alienware OptX AW2310 full HD 3D monitor and an NVIDIA 3D Vision$^{\text{TM}}$2 active shutter glasses. The viewing distance from subjects to screen was fixed at 23 inches, which is 3 times the screen height. Twenty-seven subjects participated in the test, four females and twenty-three males, aged from 24 to 50 years.
- StSD 3D Video Database [34] was developed at University of Surrey. The StSD 3D Video Database contains a data set of 116 subject-rated stereoscopic videos create from 14 pristine videos. The resolution of the videos is 960 × 1080 pixels. All videos are eight seconds long. The frame rate of the videos is 25 frames per second. H.264/AVC and HEVC coding were chosen as the compression method and every reference stereoscopic video was processed to create symmetrically compressed stereoscopic videos and asymmetrically compressed stereoscopic videos. A DSCQS protocol was adopted in the study. The continuous quality scales are labeled by equally spaced adjective terms: bad, poor, fair, good, and excellent. The study

was conducted using a 46-inch JVC GD-463D10 passive stereoscopic display with a 2D equivalent resolution of 1920 × 1080. The viewing distance is 3.75 meters away from the screen, which is approximately 5.5 times the picture height. 16 non-expert observers participated in the test. The average age of the subjects was 33 years.

- Tampere 3D Video Database [61] was developed at Tampere University of Technology. The Tampere 3D Video Database contains a data set of 60 subject-rated stereoscopic videos create from 4 pristine videos. The resolution of the videos is 720 × 480 pixels. All videos are ten seconds long. The frame rate of the videos is 30 frames per second. H.264/AVC coding was chosen as the compression method. Every reference stereoscopic video was processed to create 15 symmetrically compressed stereoscopic videos with 3 depth levels and 5 QP levels. An absolute category rating (ACR) protocol with 11-points unlabelled scales was adopted in the study. The study was conducted using a portable autostereoscopic LCD display with horizontal double density pixel (HDDP) arrangement produced by NEC LCD. The physical size of the display is 3.5 inch and the resolution is 427 × 240 pixels at 155 DPI. The viewing distance between the display and the viewer was set to 40 cm. Thirty subjects participated in the test, 15 females and 15 males, aged from 18 to 45 years.

- NAMA3DS1-COSPAD1 3D Video Database [139] was developed at l'Université de Nantes and Universidad Politécnica de Madrid. The NAMA3DS1-COSPAD1 3D Video Database contains a data set of 110 subject-rated stereoscopic videos create from 10 pristine videos. The resolution of the videos is 1920 × 1080 pixels. 99 sequences are sixteen seconds long and 11 sequences are thirteen seconds long. The frame rate of the videos is 25 frames per second. JPEG 2000 and H.264/AVC coding were chosen as the compression method. Every reference stereoscopic video was processed to create symmetrically compressed stereoscopic videos. Besides, one pre-processing (downsampling) and one post-processing (sharpening) were included to generate test sequences. An absolute category rating with hidden reference (ACR-HR) protocol with 5-points labelled from 1 (bad) to 5 (excellent) was adopted in the study. The study was conducted using a Philips 46PFL9705H 46-inch stereoscopic display with an active shutter glasses. The viewing distance is 172 cm, which is three times the height of the screen. Twenty-nine subjects participated in the test, 12 females and 17 males, aged from 18 to 63 years.

- MMSPG 3D Video Quality Assessment Database [42] was developed at MULTIMEDIA SIGNAL PROCESSING GROUP MMSPG. The MMSPG 3D Video Quality Assessment Database contains a data set of 36 subject-rated stereoscopic videos including 6 different scenes. The resolution of the videos is 1920 × 1080 pixels. The frame rate of the videos is 25 frames per second. For each of the scenes 6 different stimuli have been considered corresponding to different camera distances. A SSCQS protocol was adopted in the study. The continuous quality scales are labeled by equally spaced adjective terms: bad, poor, fair, good, and excellent. The viewing distance is approximatively two meters which is equal to the height of the screen multiplied by factor 3. Twenty non-expert subjects participated in the test, six females and sixteen males, aged from 24 to 37 years with an average of 27.

- UBC Digital Multimedia Lab 3D Video Database [6, 7] was developed at University of British Columbia. The UBC Digital Multimedia Lab 3D Video Database contains a data set of 64 subject-rated stereoscopic videos create from 5 pristine videos. The resolution of the videos is 1920 × 1080 pixels. The frame rates of the videos are $24, 30, 48$ and 60 frames per second. 3D HEVC coding standard was chosen as the compression method. Every reference stereoscopic video was processed to create compressed stereoscopic videos at four QP levels. A single stimulus procedure using an 11-grade numerical categorical scale protocol was adopted in the study. The study was conducted using a 64-inch full HD 3D TV with a circularly passive polarized glasses. Sixteen subjects participated in the test, aged from 19 to 37 years.

- 3DVCL@FER Video Database [36] was developed at University North in Croatia and University of Coimbra in Portugal. The 3DVCL@FER Video Database contains a data set of 184 subject-rated stereoscopic videos create from 8 pristine videos. The resolution of the videos is 1920 × 1080 pixels. All videos are sixteen seconds long. The frame rate of the videos is 25 frames per second. A total of 22 degradation types are adopted to generate test sequences including 3D HEVC coding, H.264/AVC coding, JPEG2000 compression, geometric distortion, horizontal and vertical disparity, packet losses, frame rate reduction, frame-freeze, etc. An ACR-HR protocol with a pseudo-continuous scale from 0 to 5 with a step of 0.1 was adopted in the study. The picture quality, depth quality and visual comfort were graded individually. The test in Croatia was conducted using a 32 inch Samsung UE32H6400 with an active shutter glasses; while the test in Portugal was conducted using a 27-inch ASUS VG278 HR with active shutter glasses. The average viewing distance is 1.2 meters. Thirty-five subjects participated in the test, aged from 20 to 48 years.

Table 2.3: Summary of existing 3D video quality databases

Database	Subjects	Protocol	Display	Videos	Resolutions	FPS	Controlling Parameters
LIVE [24]	27	DSCQS	Active	54	720×480	25	H.264 compression
StSD [34]	14	DSCQS	Passive	116	960×1080	25	H.264 & HEVC compressions
Tampere [61]	30	ACR	Auto	60	1024×768	30	Depth levels & H.264 compression
MMSPG [42]	20	SSCQS	Passive	36	1920×1080	25	Different camera distances
NAMA3DS1-COSPAD1 [139]	29	ACR-HR	Active	110	1920×1080	25	H.264 & JPEG2000 compression, downsampling & sharpening
UBC [6, 7]	16	SSNCS	Passive	64	1920×1080	$24, 30, 48, 60$	3D HEVC compressions & Frame rates
3DVCL@FER [36]	35	ACR-HR	Active	184	1920×1080	25	3D HEVC, H.264, JPEG2000 compression, Geometric distortion, Packet losses, Frame rates & Frame-freeze

2.4 Objective 3D Image/Video Quality Assessment

Existing objective 3D-IQA/VQA models may be grouped into two categories. The first type of approaches are built directly upon successful 2D-IQA/VQA methods. These approaches can be further divided into two subcategories, depending on the use of depth or disparity information. Methods in the first subcategory do not explicitly use depth information.

2.4.1 Objective 3D-IQA Studies

In [19], four 2D-IQA metrics, namely SSIM, Universal Quality Index (UQI) [149], C4 [21] and Reduced Reference Image Quality Assessment (RRIQA) [157] as well as three approaches, called average approach, main eye approach, and visual acuity approach, were tested for measuring the perceptual quality of stereoscopic images. The experimental results show that C4 outperforms the other three metrics on IRCCyN/IVC 3D Images Database.

The second subcategory of methods incorporates depth information with 2D-IQA. In [10, 11], disparity maps between left- and right-views were estimated, followed by 2D quality assessment of disparity quality using SSIM and C4, which was subsequently combined with 2D image quality to produce an overall 3D image quality score. The results claimed that C4 outperforms SSIM on both evaluating stereoscopic image pairs and disparity maps on IRCCyN/IVC 3D Images Database and also suggested that the 3D-IQA performance of SSIM can be improved when adding depth quality. You *et al.* [168] investigated the capabilities of evaluating stereopairs as well as disparity maps with respect to ten well-known 2D-IQA metrics, i.e., PSNR, SSIM, MS-SSIM [158], UQI, Visual Information Fidelity (VIF) [123], etc. Their results suggested that an improved performance can be achieved when stereo image quality and depth quality are combined appropriately. Similarly, Yang *et al.* [164, 163] proposed a 3D-IQA algorithm based on the average PSNR of left- and right-views and the absolute difference with respect to disparity map. However, none of these more sophisticated 3D-IQA models perform better than or in most cases, even as good as, direct averaging 2D-IQA measures of both views [23].

The second type of 3D-IQA approaches focus on building 3D quality models directly without relying on existing 2D-IQA algorithms. In [45], Gorley *et al.* computed quality scores on matched feature points delivered by SIFT [76] and RANSAC [40]. The experimental results showed that the stereo band limited contrast model performs better than PSNR. In [127], an estimation of stereo image quality was proposed based on a multiple channel HVS model.

Of particular interests are several models that consider binocular visual characteristics. The quality metric presented in [12] is based on binocular energy contained in the left- and right-views calculated by complex wavelet transform and Bandelet transform and the results showed that it outperforms the no-reference 3D-IQA model proposed in [113]. In [108], a new quality metric for stereoscopic images based on binocular perception model and three components of SSIM was proposed. The experimental results showed that this binocular perception model performs better than PSNR, SSIM as well as 3D-IQA models [10, 163] on IRCCyN/IVC 3D Images Database. In [74], a quality assessment algorithm of stereoscopic image compression based on binocular combination and binocular frequency integration was proposed. The experimental results showed that the binocular integration model performs better than 3D-IQA models [45] and [50], but not good as [10], [168] and [25] on LIVE 3D Image Quality Database Phase II. In [121], a stereoscopic image is separated into different binocular regions, each evaluated independently by considering their visual properties, followed by an integration step to produce an overall quality score. The experimental results showed that the region-based model outperforms PSNR, MS-SSIM and VIF as well as 3D-IQA models [10, 168] on Ningbo University 3D Image Quality Assessment Database Phase II. In [25], a "cyclopean" 3D-IQA model accounting for binocular rivalry was proposed and the experimental results showed that the framework significantly outperforms conventional 2D-IQA metrics PSNR, SSIM, MS-SSIM and VIF as well as [10], [168], [45], and [50] on LIVE 3D Image Quality Database Phase II.

2.4.2 Objective 3D-VQA Studies

Similarly, existing objective 3D-VQA methods may be grouped into two categories. The first type of approaches are built directly upon successful 2D-IQA/VQA methods. In [51, 166], 2D-IQA measures, including PSNR, SSIM and video quality metric (VQM) [102], were applied to the left- and right-view images/videos of 3D videos separately and then combined to a 3D quality score. Both experimental results showed that VQM performs better than PSNR and SSIM. In [138], PSNR and VSSIM [155], which is a version of SSIM adapted for video, were compared to measure the perceptual 3D quality and the VSSIM was found to be closer to the subjective evaluation results. In [24], PSNR and MS-SSIM were applied to estimate 3D image quality and overall 3D quality-of-experience. The subjective testing results showed that MS-SSIM slightly outperforms PSNR with respect to both 3D visual experience criteria.

The second type of 3D-VQA approaches focus on building 3D quality models directly without relying on existing 2D-IQA/VQA algorithms. In [176], Zhu *et al.* proposed a 3D-VQA model by considering depth perception and their experimental results showed that it performs better than MSE and PSNR. In [58, 59], Jin *et al.* proposed a 3D-VQA model based on 3D-DCT transform. Similar blocks from left- and right-views are found by block-matching, grouped into 3D stacks and then analyzed by 3D-DCT. The experimental results showed that the model outperforms PSNR, SSIM, MS-SSIM and UQI on Tampere 3D Video Database [61]. In [111], a SSIM-inspired 3D-VQA model using depth map segmentation was proposed followed by an extensive subjective test. The results indicated that the model can predict perceived 3D video quality effectively. In [118], an objective 3D-VQA algorithm using blocking artifacts, blurring in edge regions, and video quality difference between two views was proposed. The subjective testing results showed that the model outperforms SSIM and VQM. In [34], a binocular suppression inspired StSD metric was proposed based on a comprehensive subjective study. The results indicated that the StSD model significantly outperforms SSIM and the aforementioned 3D-VQA models [58, 118] on StSD 3D Video Database [34].

2.5 Discussion on existing 3D-IQA and 3D-VQA studies

Recent subjective studies suggested that in the case of symmetric distortion of both views (in terms of both distortion types and levels), simply averaging state-of-the-art 2D-IQA measures of both views is sufficient to provide reasonably accurate image quality predictions of stereoscopic images. In particular, in [90], it was shown that averaging PSNR, SSIM, MS-SSIM, UQI and VIF measurements of left- and right-views performs equally well or better than the advanced 3D-IQA models [10, 164, 168, 163, 45, 127, 176, 50], and [4] on LIVE 3D Image Quality Database Phase I. Similar results were also observed in [23], where averaging SSIM and MS-SSIM measurements of both views outperformed advanced 3D-IQA models [10, 168, 45, 74, 50], and [4] on LIVE 3D Image Quality Database Phase II. In [121], it was reported that directly averaging MS-SSIM outperformed 3D-IQA models [10, 168] on Ningbo University 3D Image Quality Assessment Database Phase II.

The performance comparison of direct averaging PSNR, SSIM, MS-SSIM, IW-SSIM and VQM on the above-mentioned two 2D-VQA databases, i.e., NAMA3DS1 and StSD

3D video databases, is conducted and reported in Table 2.5. We can observe that in the cases of symmetric compressions or slightly asymmetric compressions, simply averaging state-of-the-art 2D-IQA/VQA measures of both views is sufficient to provide reasonably accurate video quality predictions of stereoscopic videos.

Table 2.4: Performance comparison of 2D-to-3D prediction models on NAMA3DS1-COSPAD1 Video Database

2D-IQA/VQA	PLCC	SRCC	KRCC	RMSE	MAE
Average PSNR	0.7104	0.6931	0.5102	0.8231	0.6391
Average SSIM	0.8037	0.7753	0.5916	0.6959	0.5348
Average MS-SSIM	0.8060	0.7942	0.6031	0.6922	0.5322
Average IW-SSIM	0.9082	0.8902	0.7126	0.4895	0.3798
Average VQM	0.8919	0.8782	0.6990	0.5289	0.3997

Table 2.5: Performance comparison of 2D-to-3D prediction models on StSD 3D Video Database

2D-IQA/VQA	PLCC	SRCC	KRCC	RMSE	MAE
Average PSNR	0.8016	0.9346	0.8981	0.9584	0.9098
Average SSIM	0.8200	0.9456	0.9250	0.9569	0.9229
Average MS-SSIM	0.6157	0.7935	0.7549	0.8315	0.7542
Average IW-SSIM	0.1311	0.2193	0.2193	0.2193	0.0910
Average VQM	0.1025	0.1891	0.1891	0.1891	0.0666

Compared with the case of symmetric distortions, quality assessment of asymmetrically distorted stereoscopic images is a much more challenging problem. In [23], it was reported that there is a large drop in the performance of both 2D-IQA and 3D-IQA models from quality predictions of symmetrically to asymmetrically distorted stereoscopic images on LIVE 3D Image Quality Database Phase II.

It is worth noting that previous studies exhibit somewhat conflicting observations and opinions regarding the effect of asymmetric distortions. For image blur, evidence in [83] shows that the quality of asymmetrically blurred images is more affected by the higher quality view, which is generally consistent with the results given in [148]. For image blockiness, it was reported in [120] that 3D image quality should be approximated by averaging the quality of high quality and low quality views but there is a tendency towards the low quality view and this tendency becomes stronger when compression levels are high and images contain homogeneous areas. In [83], an under-weighting when direct averaging the quality of both views is found for monocular blockiness from MPEG-2 codec. In [148], it was suggested that the best strategy of asymmetric quality assessment for JPEG compressed images should be content and texture dependent.

Chapter 3

Quality prediction of asymmetrically distorted stereoscopic 3D images

In this chapter, we first build a database that contains both single-view and symmetrically and asymmetrically distorted stereoscopic images. We then carry out a subjective test, where we find that the quality prediction bias of asymmetrically distorted images could lean towards opposite directions (overestimate or underestimate), depending on the distortion types and levels. Our subjective test also suggests that eye dominance effect does not have strong impact on visual quality decisions of stereoscopic images. We then develop an information content and divisive normalization based pooling scheme that improves upon SSIM in estimating the quality of single-view images. Finally, we propose a binocular rivalry inspired multi-scale model to predict the quality of stereoscopic images from that of the single-view images. Our results show that the proposed model, without explicitly identifying image distortion types, successfully eliminates the prediction bias, leading to significantly improved quality prediction of stereoscopic images.

3.1 Introduction

Objective quality assessment of distorted stereoscopic images is a challenging problem, especially when the distortions in the left- and right-views are asymmetric. As we discussed in Section 2.4, existing studies suggest that simply averaging the quality of the left- and right-views well predicts the quality of symmetrically distorted stereoscopic images, but generates substantial prediction bias when applied to asymmetrically distorted stereoscopic images.

In this chapter, we focus on how to predict the quality of a stereoscopic 3D image from that of the 2D single-view images. First, we carry out a subjective quality assessment experiment on a database that contains both single-view images and stereoscopic images with symmetric and asymmetric distortion types and levels. This database allows us to directly study the quality prediction performance from single-view images to stereoscopic images, for which we observe that simply averaging the quality of both views creates substantial bias on asymmetrically distorted stereoscopic images, and interestingly, the bias could lean towards opposite directions, largely depending on the distortion types. We then develop an information content and divisive normalization based pooling scheme that improves upon SSIM in estimating the quality of single-view images. Furthermore, by incorporating spatial frequency tuned mechanisms of the HVS, we propose a series of binocular rivalry inspired models (FR and NR) to account for the bias, which not only results in better quality prediction of stereoscopic images with asymmetric distortion levels, but also well generalizes to the case of asymmetric distortions with mixed distortion types.

3.2 Subjective Database

3.2.1 Waterloo-IVC 3D Image Quality Database Phase I

The new Waterloo-IVC 3D Image Quality Database Phase I is created from 6 pristine stereoscopic image pairs (and thus their corresponding single-view images) shown in Figure 3.1, all collected from the Middlebury Stereo 2005 Datasets [114]. The original res-

(a) (b) (c)

(d) (e) (f)

Figure 3.1: The 6 pristine images in Waterloo-IVC 3D Image Database Phase I. Only the right-views are shown here. (a) Art. (b) Books. (c) Dolls. (d) Moebius. (e) Laundry. (f) Reindeer.

olution of single-view images is 1390 × 1100 or 1342 × 1100. All single-view images and stereopairs were slightly cropped to fit a display of 1920 × 1080 resolution. Table 3.1 lists the original resolution and the displayed resolution of all training and testing images.

Each single-view image was altered by three types of distortions: additive white Gaussian noise contamination, Gaussian blur, and JPEG compression. Each distortion type had four distortion levels, where the distortion control parameters were decided to ensure a good perceptual separation between distortion levels as reported in Table 3.2 and Table 3.3. More specially, additive white Gaussian noise was applied equally across the R, G and B color channels. Similarly, Gaussian blur was simulated by applying a Gaussian low-pass filter to each of the color channels. For both noise and blur, the control parameter

Table 3.1: Resolutions of training and testing images in Waterloo-IVC 3D Image Database Phase I

Image	Original Resolution	Adjusted Resolution	Usage
Art	1390 × 1100	1390 × 1080	Testing
Books	1390 × 1100	1390 × 1080	Testing
Dolls	1390 × 1100	1390 × 1080	Testing
Moebius	1390 × 1100	1390 × 1080	Testing
Laundry	1342 × 1100	1342 × 1080	Testing
Reindeer	1342 × 1100	1342 × 1080	Testing
Computer	1330 × 1100	1330 × 1080	Training
Drumsticks	1390 × 1100	1390 × 1080	Training

was the variance of the Gaussian. JPEG Compression was simulated using MATLAB®'s JPEG compression utility `imwrite` (Write True Color Image to JPEG). Consequently, 72 distorted single-view 2D images (24 each for Noise, Blur and JPEG) were created, which were named as W1 to W4, G1 to G4 and J1 to J4, respectively. PSNR and SSIM [153] evaluations of the simulated distorted single-view 2D images are reported in Table 3.4 and Table 3.5, respectively.

Table 3.2: Value ranges of control parameters for distortion simulation for Waterloo-IVC 3D Image Database Phase I

Distortion	Control Parameter	Range
White Noise	Variance of Gaussian	[0.105 0.390]
Gaussian Blur	Variance of Gaussian	[2.40 20.00]
JPEG Compression	Quality Parameter	[3 10]

The single-view images were employed to generate distorted stereopairs, either symmetrically or asymmetrically. Altogether, there are totally 78 single-view images and 330 stereoscopic images. Table 3.6 categorizes these images into seven groups with detailed descriptions. Group 3D.1, Group 3D.2 and Group 3D.3 cover all combinations while Group 3D.4 includes a random subset from all possible fusions. For mixed distortion types and levels, there are 48 possible combinations for each source image and 288 in total. These are too many combinations in the subjective test that could raise major concerns on visual

Table 3.3: Values of control parameters for distortion simulation for Waterloo-IVC 3D Image Database Phase I

Art						
White Noise	σ	**Gaussian Blur**	σ	**Width**	**JPEG**	**Quality**
W1	0.115	G1	3.00	21	J1	10
W2	0.170	G2	5.10	31	J2	7
W3	0.250	G3	8.10	45	J3	5
W4	0.350	G4	13.00	71	J4	3
Books						
White Noise	σ	**Gaussian Blur**	σ	**Width**	**JPEG**	**Quality**
W1	0.125	G1	2.40	15	J1	10
W2	0.190	G2	4.00	27	J2	7
W3	0.280	G3	6.80	41	J3	5
W4	0.350	G4	11.50	61	J4	3
Dolls						
White Noise	σ	**Gaussian Blur**	σ	**Width**	**JPEG**	**Quality**
W1	0.120	G1	2.80	19	J1	10
W2	0.190	G2	4.70	29	J2	7
W3	0.275	G3	7.40	41	J3	5
W4	0.390	G4	11.75	61	J4	3
Laundry						
White Noise	σ	**Gaussian Blur**	σ	**Width**	**JPEG**	**Quality**
W1	0.120	G1	2.60	17	J1	10
W2	0.185	G2	4.40	27	J2	7
W3	0.270	G3	7.00	41	J3	5
W4	0.390	G4	11.75	61	J4	3
Moebius						
White Noise	σ	**Gaussian Blur**	σ	**Width**	**JPEG**	**Quality**
W1	0.105	G1	2.90	21	J1	10
W2	0.155	G2	5.10	31	J2	7
W3	0.215	G3	8.60	47	J3	5
W4	0.300	G4	15.00	81	J4	3
Reindeer						
White Noise	σ	**Gaussian Blur**	σ	**Width**	**JPEG**	**Quality**
W1	0.105	G1	2.90	21	J1	10
W2	0.155	G2	5.70	33	J2	7
W3	0.215	G3	10.00	55	J3	5
W4	0.300	G4	20.00	91	J4	3

Table 3.4: PSNR evaluations of distorted 2D images on Waterloo-IVC 3D Image Database Phase I

White Noise				
LEVEL	**W1**	**W2**	**W3**	**W4**
Art	22.57	19.46	16.57	14.32
Books	21.85	18.63	15.83	13.70
Dolls	21.94	18.59	15.87	13.65
Moebius	23.20	19.98	17.41	15.09
Laundry	22.04	18.51	15.75	13.56
Reindeer	23.65	20.47	17.88	15.42
Gaussian Blur				
LEVEL	**G1**	**G2**	**G3**	**G4**
Art	28.85	26.13	23.99	22.11
Books	25.97	23.65	21.59	19.71
Dolls	28.33	25.73	23.66	21.68
Moebius	29.15	26.76	24.83	22.80
Laundry	27.26	24.78	22.70	20.77
Reindeer	28.63	25.96	23.83	21.24
JPEG Compression				
LEVEL	**J1**	**J2**	**J3**	**J4**
Art	32.41	30.69	28.94	26.27
Books	30.67	29.18	27.58	25.02
Dolls	31.58	29.88	28.18	25.57
Moebius	32.03	30.48	28.89	26.33
Laundry	32.04	30.26	28.57	25.98
Reindeer	32.40	30.72	29.04	25.93

Table 3.5: SSIM evaluations of distorted 2D images on Waterloo-IVC 3D Image Database Phase I

White Noise				
LEVEL	**W1**	**W2**	**W3**	**W4**
Art	0.8003	0.7037	0.5986	0.5030
Books	0.8146	0.7203	0.6188	0.5275
Dolls	0.8050	0.7032	0.6001	0.4990
Moebius	0.8081	0.7018	0.6025	0.5013
Laundry	0.8038	0.6983	0.5986	0.4977
Reindeer	0.8011	0.6996	0.6025	0.5032

Gaussian Blur				
LEVEL	**G1**	**G2**	**G3**	**G4**
Art	0.8521	0.7054	0.5509	0.4056
Books	0.8578	0.7057	0.5396	0.3857
Dolls	0.8562	0.7089	0.5582	0.4078
Moebius	0.8515	0.7070	0.5566	0.4102
Laundry	0.8546	0.7116	0.5574	0.4050
Reindeer	0.8571	0.7023	0.5559	0.4017

JPEG Compression				
LEVEL	**J1**	**J2**	**J3**	**J4**
Art	0.9211	0.8812	0.8308	0.7388
Books	0.9261	0.8908	0.8482	0.7568
Dolls	0.9239	0.8830	0.8321	0.7318
Moebius	0.8994	0.8469	0.7954	0.6943
Laundry	0.9362	0.9012	0.8606	0.7686
Reindeer	0.9010	0.8546	0.8224	0.7132

fatigue. Thus we divided all possible combinations into 12 clusters based on distortion types and levels and randomly selected one from each cluster. Table 3.7 lists all possible combinations of mixed distortion types and levels and how we define these clusters. The details of the random selections in Group 3D.4 of Waterloo-IVC 3D Image Database Phase I can be found in Table 3.8.

Table 3.6: Categories of test images on Waterloo-IVC 3D Image Database Phase I

Group	# of images	Description
2D.0	6×1	Pristine single-view images
2D.1	6×12	Distorted single-view images
3D.0	6×1	Pristine stereopairs
3D.1	6×12	Symmetrically distorted stereopairs with the same distortion type and distortion level
3D.2	6×12	Asymmetrically distorted stereopairs with distortion on one view only
3D.3	6×18	Asymmetrically distorted stereopairs with the same distortion type but different levels
3D.4	6×12	Asymmetrically distorted stereopairs with mixed distortion types and levels

3.2.2 Waterloo-IVC 3D Image Quality Database Phase II

The new Waterloo-IVC 3D Image Quality Database Phase II with more diverse image content is created from 10 pristine stereoscopic image pairs (and thus their corresponding single-view images) shown in Figure 3.2. All images were collected from previous subjective 3D quality studies [139, 100]. The original resolution of single-view images is 1920×1080 or 1920×1088. All single-view images and stereopairs were slightly cropped to fit a display of 1920×1080 resolution. Table 3.9 lists the original resolution and the displayed resolution of all training and testing images. Each single-view image was altered by the same three types of distortions and each distortion type had the same four distortion levels. Values of control parameters for distortion simulation are reported in Table 3.10 to Table 3.12. PSNR and SSIM evaluations of the simulated distorted single-view 2D images are reported in Table 3.13 and Table 3.14, respectively.

Table 3.7: All possible combinations of mixed distortion types and levels

Cluster 1		Cluster 4		Cluster 7		Cluster 10	
LeftG1	RightW1	LeftG2	RightW1	LeftG3	RightW1	LeftG4	RightW1
LeftG1	RightW2	LeftG2	RightW2	LeftG3	RightW2	LeftG4	RightW2
LeftG1	RightW3	LeftG2	RightW3	LeftG3	RightW3	LeftG4	RightW3
LeftG1	RightW4	LeftG2	RightW4	LeftG3	RightW4	LeftG4	RightW4
Cluster 2		Cluster 5		Cluster 8		Cluster 11	
LeftG1	RightJ1	LeftG2	RightJ1	LeftG3	RightJ1	LeftG4	RightJ1
LeftG1	RightJ2	LeftG2	RightJ2	LeftG3	RightJ2	LeftG4	RightJ2
LeftG1	RightJ3	LeftG2	RightJ3	LeftG3	RightJ3	LeftG4	RightJ3
LeftG1	RightJ4	LeftG2	RightJ4	LeftG3	RightJ4	LeftG4	RightJ4
Cluster 3		Cluster 6		Cluster 9		Cluster 12	
LeftW1	RightJ1	LeftW2	RightJ1	LeftW3	RightJ1	LeftW4	RightJ1
LeftW1	RightJ2	LeftW2	RightJ2	LeftW3	RightJ2	LeftW4	RightJ2
LeftW1	RightJ3	LeftW2	RightJ3	LeftW3	RightJ3	LeftW4	RightJ3
LeftW1	RightJ4	LeftW2	RightJ4	LeftW3	RightJ4	LeftW4	RightJ4

Table 3.8: Group 3D.4 of Waterloo-IVC 3D Image Database Phase I

Art	Books	Dolls	Laundry	Moebius	Reindeer
ArtW1G1	BooksG1W2	DollsW1G1	LaundryG1W2	MoebiusG1W3	ReindeerW1G1
ArtJ3G1	BooksJ1G1	DollsJ1G1	LaundryJ1G1	MoebiusG1J3	ReindeerG1J3
ArtJ2W1	BooksW1J4	DollsW1J3	LaundryW1J4	MoebiusJ2W1	ReindeerW1J1
ArtG2W2	BooksG2W4	DollsW1G2	LaundryG2W1	MoebiusW2G2	ReindeerW1G2
ArtG2J2	BooksJ1G2	DollsG2J1	LaundryJ3G2	MoebiusG2J1	ReindeerJ3G2
ArtJ2W2	BooksW2J4	DollsJ3W2	LaundryJ1W2	MoebiusW2J2	ReindeerJ1W2
ArtW3G3	BooksW2G3	DollsG3W4	LaundryG3W3	MoebiusG3W4	ReindeerW3G3
ArtJ1G3	BooksJ4G3	DollsG3J3	LaundryG3J1	MoebiusG3J3	ReindeerG3J1
ArtJ1W3	BooksW3J2	DollsJ2W3	LaundryJ2W3	MoebiusJ1W3	ReindeerW3J3
ArtG4W2	BooksG4W2	DollsG4W4	LaundryG4W3	MoebiusG4W2	ReindeerG4W2
ArtG4J1	BooksJ2G4	DollsG4J1	LaundryG4J2	MoebiusG4J2	ReindeerG4J2
ArtJ3W4	BooksW4J2	DollsJ2W4	LaundryJ4W4	MoebiusW4J2	ReindeerJ3W4

Table 3.9: Resolutions of training and testing images in Waterloo-IVC 3D Image Database Phase II

Image	Original Resolution	Adjusted Resolution	Usage
Barrier	1920 × 1080	1920 × 1080	Testing
CraftLoom	1920 × 1080	1920 × 1080	Testing
Dancer	1920 × 1088	1920 × 1080	Testing
Hall	1920 × 1080	1920 × 1080	Testing
Laboratory	1920 × 1080	1920 × 1080	Testing
OldTownCar	1920 × 1080	1920 × 1080	Testing
Persons	1920 × 1080	1920 × 1080	Testing
Soccer	1920 × 1080	1920 × 1080	Testing
Tree	1920 × 1080	1920 × 1080	Testing
Umbrella	1920 × 1080	1920 × 1080	Testing
Basketball	1920 × 1080	1920 × 1080	Training
Gym	1920 × 1080	1920 × 1080	Training

Table 3.10: Value ranges of control parameters for distortion simulation for Waterloo-IVC 3D Image Database Phase II

Distortion	Control Parameter	Range
White Noise	Variance of Gaussian	[0.080 0.530]
Gaussian Blur	Variance of Gaussian	[2.00 20.00]
JPEG Compression	Quality Parameter	[3 10]

Table 3.11: Values of control parameters for distortion simulation for Waterloo-IVC 3D Image Database Phase II

Barrier						
White Noise	σ	**Gaussian Blur**	σ	**Width**	**JPEG**	**Quality**
W1	0.110	G1	2.50	15	J1	10
W2	0.160	G2	4.60	31	J2	7
W3	0.240	G3	7.50	43	J3	5
W4	0.340	G4	11.75	61	J4	3
CraftLoom						
White Noise	σ	**Gaussian Blur**	σ	**Width**	**JPEG**	**Quality**
W1	0.110	G1	2.70	21	J1	10
W2	0.170	G2	4.70	31	J2	7
W3	0.250	G3	7.50	45	J3	5
W4	0.360	G4	12.00	65	J4	3
Dancer						
White Noise	σ	**Gaussian Blur**	σ	**Width**	**JPEG**	**Quality**
W1	0.110	G1	2.50	15	J1	10
W2	0.160	G2	4.60	29	J2	7
W3	0.230	G3	8.00	45	J3	5
W4	0.320	G4	15.00	81	J4	3
Hall						
White Noise	σ	**Gaussian Blur**	σ	**Width**	**JPEG**	**Quality**
W1	0.140	G1	2.20	13	J1	10
W2	0.220	G2	3.50	23	J2	7
W3	0.320	G3	5.30	31	J3	5
W4	0.440	G4	8.70	47	J4	3
Laboratory						
White Noise	σ	**Gaussian Blur**	σ	**Width**	**JPEG**	**Quality**
W1	0.100	G1	3.20	21	J1	10
W2	0.160	G2	5.70	33	J2	7
W3	0.230	G3	9.00	51	J3	5
W4	0.320	G4	14.00	75	J4	3

Table 3.12: Values of control parameters for distortion simulation for Waterloo-IVC 3D Image Database Phase II

<table>
<tr><th colspan="7">OldTownCar</th></tr>
<tr><th>White Noise</th><th>σ</th><th>Gaussian Blur</th><th>σ</th><th>Width</th><th>JPEG</th><th>Quality</th></tr>
<tr><td>W1</td><td>0.110</td><td>G1</td><td>2.90</td><td>21</td><td>J1</td><td>10</td></tr>
<tr><td>W2</td><td>0.170</td><td>G2</td><td>4.90</td><td>31</td><td>J2</td><td>7</td></tr>
<tr><td>W3</td><td>0.240</td><td>G3</td><td>7.60</td><td>43</td><td>J3</td><td>5</td></tr>
<tr><td>W4</td><td>0.340</td><td>G4</td><td>11.20</td><td>61</td><td>J4</td><td>3</td></tr>
<tr><th colspan="7">Persons</th></tr>
<tr><th>White Noise</th><th>σ</th><th>Gaussian Blur</th><th>σ</th><th>Width</th><th>JPEG</th><th>Quality</th></tr>
<tr><td>W1</td><td>0.080</td><td>G1</td><td>3.30</td><td>21</td><td>J1</td><td>10</td></tr>
<tr><td>W2</td><td>0.130</td><td>G2</td><td>6.00</td><td>35</td><td>J2</td><td>7</td></tr>
<tr><td>W3</td><td>0.190</td><td>G3</td><td>10.80</td><td>61</td><td>J3</td><td>5</td></tr>
<tr><td>W4</td><td>0.260</td><td>G4</td><td>20.00</td><td>91</td><td>J4</td><td>3</td></tr>
<tr><th colspan="7">Soccer</th></tr>
<tr><th>White Noise</th><th>σ</th><th>Gaussian Blur</th><th>σ</th><th>Width</th><th>JPEG</th><th>Quality</th></tr>
<tr><td>W1</td><td>0.170</td><td>G1</td><td>2.00</td><td>13</td><td>J1</td><td>10</td></tr>
<tr><td>W2</td><td>0.250</td><td>G2</td><td>3.10</td><td>21</td><td>J2</td><td>7</td></tr>
<tr><td>W3</td><td>0.370</td><td>G3</td><td>5.10</td><td>31</td><td>J3</td><td>5</td></tr>
<tr><td>W4</td><td>0.500</td><td>G4</td><td>7.80</td><td>45</td><td>J4</td><td>3</td></tr>
<tr><th colspan="7">Tree</th></tr>
<tr><th>White Noise</th><th>σ</th><th>Gaussian Blur</th><th>σ</th><th>Width</th><th>JPEG</th><th>Quality</th></tr>
<tr><td>W1</td><td>0.170</td><td>G1</td><td>2.20</td><td>13</td><td>J1</td><td>10</td></tr>
<tr><td>W2</td><td>0.260</td><td>G2</td><td>3.40</td><td>21</td><td>J2</td><td>7</td></tr>
<tr><td>W3</td><td>0.380</td><td>G3</td><td>5.00</td><td>31</td><td>J3</td><td>5</td></tr>
<tr><td>W4</td><td>0.530</td><td>G4</td><td>7.60</td><td>45</td><td>J4</td><td>3</td></tr>
<tr><th colspan="7">Umbrella</th></tr>
<tr><th>White Noise</th><th>σ</th><th>Gaussian Blur</th><th>σ</th><th>Width</th><th>JPEG</th><th>Quality</th></tr>
<tr><td>W1</td><td>0.140</td><td>G1</td><td>2.20</td><td>13</td><td>J1</td><td>10</td></tr>
<tr><td>W2</td><td>0.210</td><td>G2</td><td>3.60</td><td>23</td><td>J2</td><td>7</td></tr>
<tr><td>W3</td><td>0.300</td><td>G3</td><td>5.70</td><td>33</td><td>J3</td><td>5</td></tr>
<tr><td>W4</td><td>0.400</td><td>G4</td><td>9.20</td><td>51</td><td>J4</td><td>3</td></tr>
</table>

Table 3.13: PSNR evaluations of distorted 2D images on Waterloo-IVC 3D Image Database Phase II

White Noise				
LEVEL	**W1**	**W2**	**W3**	**W4**
Barrier	22.97	19.99	16.94	14.54
CraftLoom	23.02	19.54	16.64	14.17
Dancer	23.10	20.03	17.20	14.86
Hall	21.04	17.55	14.91	12.98
Laboratory	24.08	20.27	17.41	14.97
OldTownCar	23.21	19.80	17.21	14.73
Persons	25.97	22.18	19.22	16.79
Soccer	19.14	16.27	13.85	12.42
Tree	19.18	16.01	13.68	12.11
Umbrella	21.04	17.96	15.41	13.55
Gaussian Blur				
LEVEL	**G1**	**G2**	**G3**	**G4**
Barrier	27.80	25.27	23.67	22.35
CraftLoom	28.77	26.33	24.62	23.06
Dancer	27.09	24.78	22.78	20.45
Hall	25.24	23.13	21.81	20.57
Laboratory	28.60	25.37	23.07	21.07
OldTownCar	27.95	25.67	24.08	22.81
Persons	27.60	24.67	21.95	19.18
Soccer	25.60	23.72	22.07	20.84
Tree	24.07	22.04	20.67	19.48
Umbrella	26.30	24.38	23.11	21.99
JPEG Compression				
LEVEL	**J1**	**J2**	**J3**	**J4**
Barrier	31.50	29.88	28.22	25.83
CraftLoom	32.18	30.42	28.66	25.91
Dancer	30.15	28.92	27.78	25.47
Hall	29.64	28.09	26.62	24.45
Laboratory	34.25	32.29	30.15	27.13
OldTownCar	32.35	30.59	28.79	26.10
Persons	34.59	32.74	30.34	27.05
Soccer	29.48	27.86	26.29	23.85
Tree	28.34	26.85	25.41	23.17
Umbrella	29.27	27.80	26.39	24.24

Table 3.14: SSIM evaluations of distorted 2D images on Waterloo-IVC 3D Image Database Phase II

White Noise				
LEVEL	**W1**	**W2**	**W3**	**W4**
Barrier	0.8008	0.7061	0.5921	0.4913
CraftLoom	0.8189	0.7141	0.6084	0.4998
Dancer	0.8035	0.7069	0.6030	0.5051
Hall	0.8140	0.7105	0.6087	0.5179
Laboratory	0.8190	0.7023	0.5990	0.5030
OldTownCar	0.8093	0.7025	0.6066	0.5037
Persons	0.8250	0.7079	0.6038	0.5143
Soccer	0.7913	0.7037	0.5995	0.5094
Tree	0.8110	0.7175	0.6180	0.5196
Umbrella	0.8067	0.7074	0.6059	0.5187
Gaussian Blur				
LEVEL	**G1**	**G2**	**G3**	**G4**
Barrier	0.8622	0.7026	0.5670	0.4575
CraftLoom	0.8571	0.6994	0.5496	0.4157
Dancer	0.8566	0.7049	0.5584	0.4204
Hall	0.8562	0.7056	0.5657	0.4167
Laboratory	0.8597	0.7028	0.5553	0.4291
OldTownCar	0.8504	0.7054	0.5677	0.4550
Persons	0.8554	0.7084	0.5505	0.4295
Soccer	0.8686	0.7294	0.5500	0.4052
Tree	0.8524	0.7024	0.5522	0.3992
Umbrella	0.8591	0.7103	0.5629	0.4183
JPEG Compression				
LEVEL	**J1**	**J2**	**J3**	**J4**
Barrier	0.8980	0.8526	0.7988	0.7143
CraftLoom	0.9226	0.8822	0.8292	0.7296
Dancer	0.8772	0.8464	0.8020	0.7246
Hall	0.9226	0.8848	0.8402	0.7644
Laboratory	0.9270	0.8917	0.8466	0.7620
OldTownCar	0.9179	0.8763	0.8198	0.7279
Persons	0.9197	0.8904	0.8581	0.7976
Soccer	0.9475	0.9181	0.8765	0.7730
Tree	0.9368	0.9033	0.8597	0.7720
Umbrella	0.9085	0.8577	0.8019	0.6974


(a) (b) (c) (d) (e) (f) (g) (h) (i) (j)

Figure 3.2: The 10 pristine images in Waterloo-IVC 3D Image Database Phase II. Only the right-views are shown here. (a) Barrier. (b) CraftLoom. (c) Dancer. (d) Hall. (e) Laboratory. (f) OldTownCar. (g) Persons. (h) Soccer. (i) Tree. (j) Umbrella.

The single-view images were employed to generate distorted stereopairs, either symmetrically or asymmetrically. Altogether, there are totally 130 single-view images and 460 stereoscopic images. Table 3.15 categorizes these images into seven groups with detailed descriptions. Group 3D.1 and Group 3D.2 cover all combinations. Group 3D.3 includes those combinations that the difference of the distortion levels is higher than ONE while Group 3D.4 includes a random subset from all possible fusions. For mixed distortion types and levels, there are 48 possible combinations for each source image and 480 in total. Similarly, we divided all possible combinations into 12 clusters based on distortion types and levels and randomly selected one from each cluster. Table 3.7 lists all possible combinations of mixed distortion types and levels and how we define these clusters. The details of the random selections in Group 3D.4 of Waterloo-IVC 3D Image Database Phase II can be found in Table 3.16.

Table 3.15: Categories of test images on Waterloo-IVC 3D Image Database Phase II

Group	# of images	Description
2D.0	10×1	Pristine single-view images
2D.1	10×12	Distorted single-view images
3D.0	10×1	Pristine stereopairs
3D.1	10×12	Symmetrically distorted stereopairs with the same distortion type and distortion level
3D.2	10×12	Asymmetrically distorted stereopairs with distortion on one view only
3D.3	10×9	Asymmetrically distorted stereopairs with the same distortion type but different levels
3D.4	10×12	Asymmetrically distorted stereopairs with mixed distortion types and levels

3.2.3 Summary

To the best of our knowledge, there are two unique features of the current database when compared with existing publicly known 3D-IQA databases. First, this is the only database that allows us to perform subjective test on both 2D and 3D images. The inclusion of 2D images allows us to directly examine the relationship between the perceptual quality of stereoscopic images and that of its single-view images. This is advantageous against previous studies which do not have ground truth of 2D image quality but have to rely on objective 2D-IQA measures to provide estimates. Second, this is the only database that contains mixed distortion types in asymmetrically distorted images.

The motivation of including different asymmetrical distortion levels and various mixed distortion types is threefold. First, purely for scientific curiosity, we are interested in knowing how the HVS behaves in the cases of asymmetrical/mixed distortions. Second, asymmetrical/mixed distortions are realistic in practice. For example, in the case of multi-exposure stereo images [135], because of the different exposure levels being used on different views, the amount of noise coming into the left- and right-view image sensors is different. For another example, asymmetric blur distortions and asymmetric blocking artifacts can be found in the case of mixed-resolution coding and asymmetric transform-domain quanti-

Table 3.16: Group 3D.4 of Waterloo-IVC 3D Image Database Phase II

Barrier	CraftLoom	Dancer	Hall	Laboratory
BarrierW2G1	CraftLoomG1W1	DancerG1W4	HallW4G1	LaboratoryW4G1
BarrierJ2G1	CraftLoomG1J2	DancerG1J1	HallJ4G1	LaboratoryG1J3
BarrierW1J4	CraftLoomJ1W1	DancerJ1W1	HallJ4W1	LaboratoryW1J2
BarrierW1G2	CraftLoomG2W1	DancerW2G2	HallG2W2	LaboratoryG2W4
BarrierG2J1	CraftLoomJ4G2	DancerJ1G2	HallG2J4	LaboratoryG2J1
BarrierJ3W2	CraftLoomW2J1	DancerJ1W2	HallW2J3	LaboratoryW2J2
BarrierG3W3	CraftLoomG3W1	DancerW3G3	HallG3W3	LaboratoryG3W2
BarrierJ2G3	CraftLoomJ3G3	DancerG3J4	HallG3J2	LaboratoryJ3G3
BarrierW3J4	CraftLoomW3J2	DancerJ3W3	HallW3J4	LaboratoryJ2W3
BarrierW2G4	CraftLoomG4W1	DancerW4G4	HallW4G4	LaboratoryW4G4
BarrierG4J3	CraftLoomG4J2	DancerG4J3	HallG4J4	LaboratoryJ3G4
BarrierJ3W4	CraftLoomW4J3	DancerW4J4	HallJ2W4	LaboratoryJ4W4
OldTownCar	Persons	Soccer	Tree	Umbrella
OldTownCarW1G1	PersonsG1W4	SoccerG1W2	TreeG1W1	UmbrellaW3G1
OldTownCarG1J1	PersonsG1J1	SoccerJ3G1	TreeJ1G1	UmbrellaJ2G1
OldTownCarW1J1	PersonsJ3W1	SoccerJ1W1	TreeW1J4	UmbrellaW1J3
OldTownCarG2W4	PersonsG2W3	SoccerG2W3	TreeG2W3	UmbrellaW3G2
OldTownCarG2J1	PersonsG2J1	SoccerJ4G2	TreeG2J3	UmbrellaG2J3
OldTownCarW2J3	PersonsW2J1	SoccerW2J4	TreeW2J1	UmbrellaJ1W2
OldTownCarW3G3	PersonsW3G3	SoccerG3W4	TreeW2G3	UmbrellaG3W3
OldTownCarJ3G3	PersonsG3J1	SoccerJ2G3	TreeJ4G3	UmbrellaJ1G3
OldTownCarW3J3	PersonsW3J4	SoccerJ3W3	TreeJ3W3	UmbrellaJ4W3
OldTownCarG4W3	PersonsW1G4	SoccerW2G4	TreeW2G4	UmbrellaG4W2
OldTownCarJ3G4	PersonsG4J3	SoccerG4J4	TreeJ1G4	UmbrellaJ3G4
OldTownCarW4J4	PersonsW4J4	SoccerJ1W4	TreeW4J2	UmbrellaW4J4

zation coding, and such distortions could have mixed types when postprocessing techniques (deblocking or blurring) are employed. Moreover, many 3D images are captured by a texture image view plus a depth map, where the texture image, which could contain noise, is used as one view, and the other view can be synthesized by combining the texture image with the depth map. Such a stereoscopic image contains both noise and mixed types of distortions. Third, the inclusion of these images provides the potential of a much stronger test on 3D-IQA models on their generalizability. Such test has been largely lacking in previous studies where the development of objective 3D-IQA models only took into account asymmetric distortions of specific and very limited distortion types such as compression only.

3.3 Subjective Test

The subjective test was conducted in the Lab for Image and Vision Computing at University of Waterloo. The test environment has no reflecting ceiling walls and floor, and was not insulated by any external audible and visual pollution. An ASUS 27" VG278H 3D LED monitor with an NVIDIA 3D Vision$^{\mathrm{TM}}$2 active shutter glasses shown in Figure 3.3 is used for the test. The default viewing distance was 3.5 times the screen height. In the actual experiment, some subjects did not feel comfortable with the default viewing distance and were allowed to adjust the actual viewing distance around it. The details of viewing conditions are given in Table 3.17. Figure 3.4 shows the experimental setup we have used.

In Phase I, twenty-four näive subjects, 14 males and 10 females aged from 22 to 45, participated in the study. In Phase II, Twenty-two näive subjects, 11 males and 11 females aged from 21 to 34, participated in the study . A 3D vision test was conducted first to verify their ability to view stereoscopic 3D content. In Phase I, three of them (1 male, 2 females) failed the vision test and did not continue with the subsequent experiment; in Phase II, no one failed the vision test. As a result, a total of twenty-one and twenty-two subjects proceeded to the formal test in Phase I and Phase II, respectively. While a visual acuity test was not performed in this study, a verbal confirmation was obtained prior to the experiment and subjects were asked to use their eyeglasses or contact lenses to correct

Figure 3.3: ASUS 27" VG278H 3D LED monitor with NVIDIA 3D Vision$^{\text{TM}}$2 active shutter glasses.

Figure 3.4: Experimental setup: the subject is facing the 3D LED monitor with active shutter glasses using a holder to control the viewing distance.

Table 3.17: Viewing conditions of the subjective test

Parameter	Value
Subjects Per Monitor	1
Screen Resolution	1920 × 1080
Screen Diameter	27.00”
Screen Width	23.53”
Screen Height	13.24”
Viewing Distance	45.00”
Viewing Angle	29.3°
Pixels Per Degree	65.5 pixels

their visual acuities.

Following previous works [119, 26], and [24], the subjects were asked to evaluate four aspects of their 3D viewing experience, including the perception of 3D image quality (3DIQ), Depth quality (DQ), Visual comfort (VC) and 3D Quality-of-Expeirence (3DQoE). The detailed descriptions of each aspects of visual experience including 2D image quality (2DIQ) are elaborated in Table 3.18. Since to visualize every 3D stereoscopic image, the subjects need to readjust their eyes so as to adapt to the content of the scene and establish 3D perception, using a double stimulus approach leads to interruptions of the viewing experience. Therefore, to reduce this effect, we choose to use the single stimulus procedure using an 11-grade numerical categorical scale protocol.

Table 3.18: Description of visual experience criteria

Criterion	Description
2DIQ	The single-view image content quality
3DIQ	The image content quality without considering depth and comfortness
DQ	The amount, naturalness and clearness of depth perception experience
VC	The comfortness when viewing stereoscopic images
3DQoE	The overall 3D viewing experience

Our pilot tests showed that one-pass experiment (where a subject gives 2DIQ, 3DIQ, DQ, VC, and 3DQoE scores to each stereoscopic image in one trial) may cause significant visual fatigue of the human subjects within a short period of time. To avoid this problem, we resorted to a multi-pass approach [26] in the formal test, where within each pass, the subject gives one of the four scores. We also found that the 2D perceptual quality of left- and right-views are so close to each other at the same distortion types and levels, so that the difference in their Mean Opinion Scores (MOS) is negligible. Thus in order to control the scale of this subjective experiment, only one of the views were tested (randomly picked) in Group 2D.0 and Group 2D.1 in the formal test. The sequence of these 2D and 3D sub-tests was determined based on the following considerations:

- The 2DIQ test should be scheduled prior to any 3D sub-tests;
- The 3DIQ test could be scheduled at any place as its goal is the most straightforward to subjects;
- DQ and VC are relatively weaker visual experience criteria, which should be not scheduled after the 3DIQ test;
- The VC test should be scheduled at a more comfortable position.
- The 3DQoE test should not be scheduled prior to any other 2D or 3D sub-tests.

Figure 3.5 shows the detailed procedure of our formal subjective test. For Phase I, the test was scheduled on two consecutive days for each subject. Day 1 (2 hours) was dedicated to 2DIQ, VC and 3DIQ tests, and Day 2 (2 hours) to DQ and 3DQoE tests. Currently, only 2DIQ and 3DIQ sub-tests were conducted on Phase II. Thus all sub-tests for Phase II were finished in a single day. Figure 3.6 shows a test image in visual comfort test. The left side "Comfort" reminds subjects of the purpose of the test.

There are two kinds of training in our subjective test. First, a general introduction was given after the 3D vision test. All 3D visual experience criteria (3DIQ, DQ, VC and 3DQoE) were introduced and their definitions were given to subjects in both written and oral forms. After this general introduction, a preliminary understanding of four 3D visual experience criteria was expected for the subjects. Second, specific instructions and training sessions were given before each sub-test (2DIQ, 3DIQ, DQ, VC and 3DQoE). In each sub-test, the corresponding rating strategy was first introduced and the subjects were

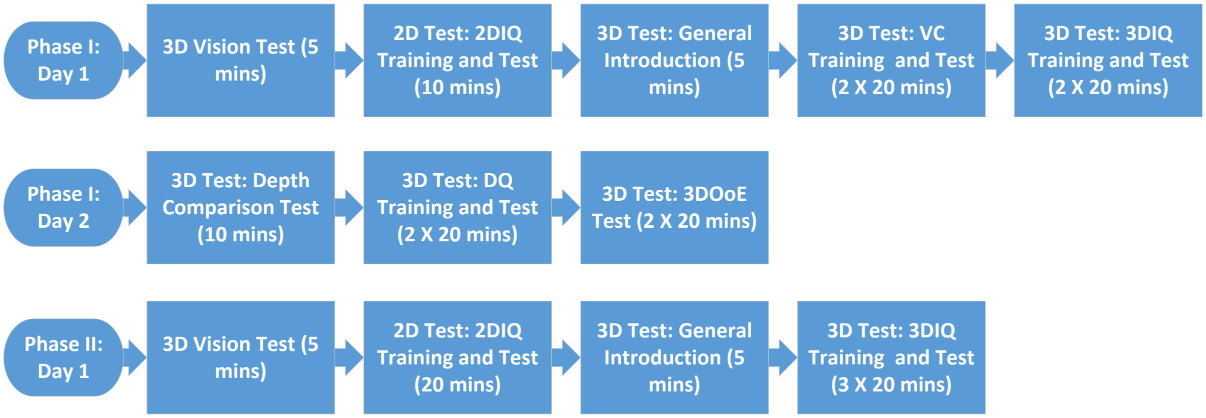


Figure 3.5: The procedure of the subjective test in Waterloo-IVC 3D Image Database Phase I and Phase II.

Figure 3.6: A test image in VC test: the right-view is JPEG distorted while the left-view is noise distorted.

then required to practice by giving scores to training images until they fully understood the criteria and built their own scoring strategies. Note that the training processes were different depending on the characteristic of each criterion:

For both 2DIQ and 3DIQ sub-tests, we use three types of images in the training phase: pristine images, moderately distorted images, and highly-distorted images. The subjects were told to give scores at the high end (close to 10 pts) to the pristine images, at the mid-range to the moderately distorted images, and at the low end (close to 0 pts) to the highly-distorted images.

For DQ sub-test, a self-training process was employed to help the subjects establishing their own rating strategies with the help of a depth comparison test (stimuli with the same source image but different depth levels were presented to help subjects establish the concept on the amount of depth), and subjects were introduced to build their own rating strategies. Previous works reported that the perception of depth information is both highly content and texture dependent [120] and subject dependent [26, 24]. Therefore, it is not desirable to educate the subjects to use the same given rating strategy. Thus after a depth comparison test (Please refer to Section 4.2 for more details), the 3D pristine stereopairs were first presented and the subjects were instructed to give high scores (close to 10 pts) to such images, and the 2D pristine images (with no depth from stereo cues) were presented and the subjects were instructed to give low scores (close to 0 pts). Next, stereopairs of different types/levels of distortions were presented and the subjects were asked to practice by giving their ratings on DQ between 0 to 10 pts. During this process, the instructor also repeated the definition of DQ and emphasized that there is not necessarily any correlation between DQ and the type/level of distortions.

For VC sub-test, similar to DQ, it is not desirable to educate the subjects to use the same given rating strategy, because previous work found that visual comfort is highly subject dependent [24]. A similar self-training process is employed for this training session. In the training process, different types/levels of distorted stereopairs including pristine stereopairs were presented and the subjects were asked to practice by giving their ratings on VC between 0 pts to 10 pts. During this process, the instructor repeated the definition of VC and emphasized that there is not necessarily any correlation between VC and the type/level of distortions.

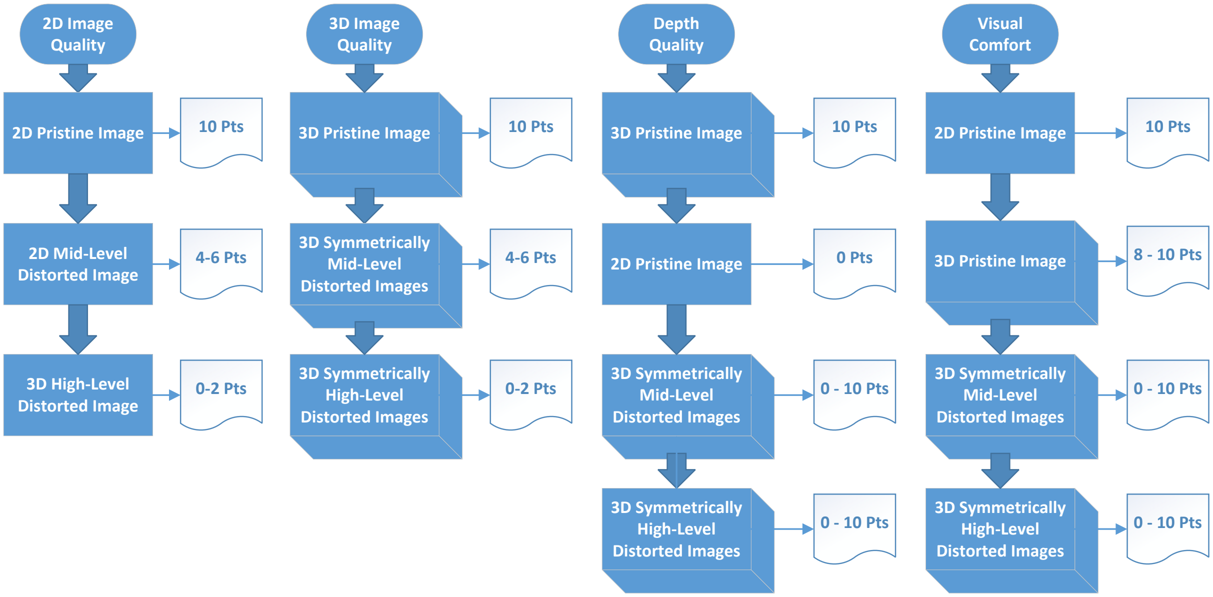


Figure 3.7: Rating strategies for 2DIQ, 3DIQ, DQ, VC and 3DQoE.

For the overall 3DQoE sub-test, there is no training session. The subjects were asked to rate each stereopair based on their overall impression. They were asked to consider the previously introduced visual experience criteria 3DIQ, DQ and VC and were encouraged to use their own strategies to use these criteria. For example, some subjects focus on image content while some others consider depth information to be more important. There are also subjects who seem to dislike 3D images in general and put visual comfort on top priority.

Figure 3.7 summaries all these rating instructions. It is worth mentioning that we carried out a supplementary depth subjective experiment with the same type/level of distortions in order to single out the contributions of stereo information from those of monocular cues, which provide useful insights in the future development of comprehensive 3D QoE models that aim to achieve a good balance between perceptual 3D image quality, depth quality, and visual comfort. This new depth subjective experiment will be discussed in Chapter 4.

All stimuli were shown once in each sub-test. However, in Phase I, there were 6 repetitions for single-view images and 12 repetitions for stereopairs, which means that for each subject, her/his first 6 single-view images and first 12 stereopairs were shown twice; in Phase II, there were 20 repetitions for single-view images and 20 repetitions for stereopairs. Table 3.19 reports the mean absolute difference (MAD) values of the repeated scores (on the scale between 0 and 10) with respect to different visual experience criteria. The small MAD values suggest that the current experiment, which does not repeat most of the stimuli, obtains reasonably reliable scores.

Table 3.19: Mean MAD values on repetitions

Phase I	2DIQ	3DIQ	DQ	VC	3DQoE
$\mu_{\mathbf{MAD}}$	1.0089	0.9588	1.0884	1.0395	0.9063
Phase II	2DIQ	3DIQ	DQ	VC	3DQoE
$\mu_{\mathbf{MAD}}$	1.0099	1.0272	N/A	N/A	N/A

The order of stimuli was randomized and the consecutive testing single-view images or stereopairs were from different source images. All single-view images or around 160 stereopairs were evaluated in one session. In Phase I, the 2DIQ sub-test, including 84 testing single-view images with 6 repetitions, was finished under 10 minutes. For 3DIQ , DQ , VC and 3DQoE sub-tests, 342 testing stereopairs with 12 repetitions were partitioned into two sessions and each single session (171 stereopairs) was finished in 15 to 20 minutes. Sufficient relaxation periods (5 minutes or more) were given between sessions. The test duration on each single day is 2 hours.

In Phase II, the 2DIQ sub-test, including 150 testing single-view images with 20 repetitions, was finished under 15 minutes. For 3DIQ sub-tests, 480 testing stereopairs with 20 repetitions were partitioned into three sessions and each single session (160 stereopairs) was finished in 15 to 20 minutes. The total test duration for Phase II is 2 hours.

Moreover, we found that repeatedly switching between viewing 3D images and grading on a piece of paper or a computer screen is a tiring experience. To overcome this problem, we asked the subject to speak out a score between 0 and 10, and a customized graphical user interface shown in Figure 3.8 on another computer screen was used by the instructor to record the score. All these efforts were intended to reduce visual fatigue and discomfort of the subjects and to reduce the interference between different visual experience criteria.

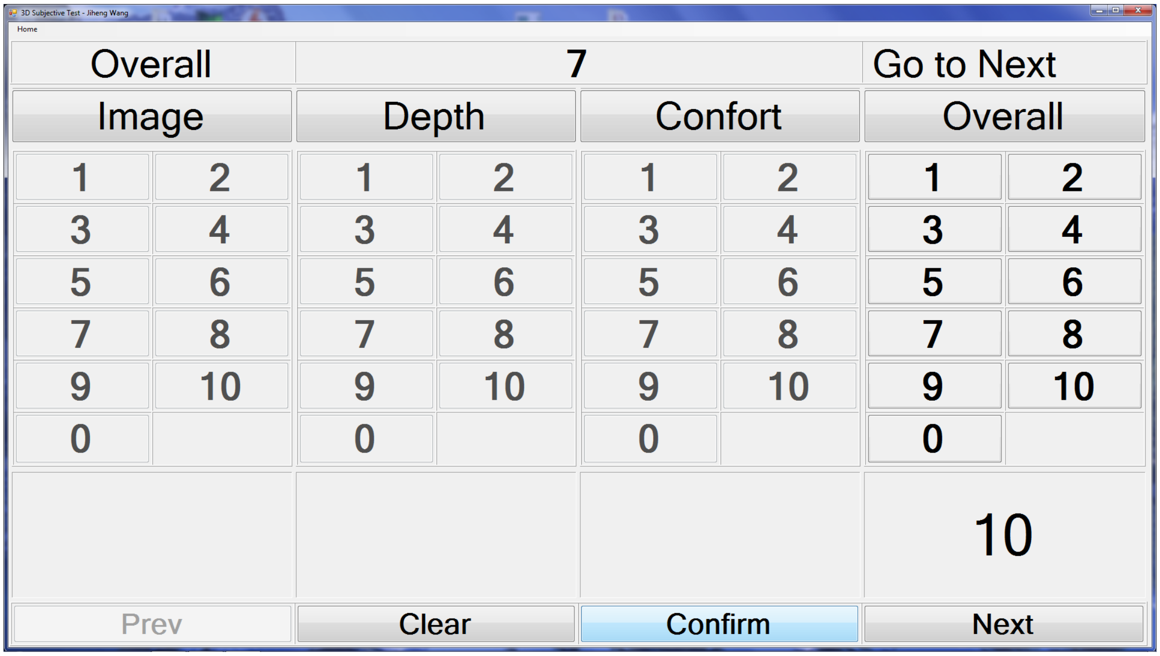


Figure 3.8: The customized GUI application for recording scores

3.4 Impact of Eye Dominance

Eye dominance is a common visual phenomenon, referring to the tendency to prefer the input from one eye to the other, depending on the human subject [67]. When studying visual quality of asymmetrically distorted images, it is important to understand if eye dominance plays a significant role in the subjective test results. For this purpose, we carried out a separate study on the impact of eye dominance in the perception of asymmetrically distorted stereoscopic images.

Twenty subjects (12 males and 8 females) participated in the experiment. The side of the dominant eye under static conditions was checked first by Rosenbach's test [107]. This test examines which eye determines the position of a finger when the subject is asked to point to an object. Ten subjects (7 males, 3 females) had a dominant left eye, and the others (5 males, 5 females) are right-eye dominant. A subjective test was conducted with the same test settings and viewing conditions as described in Section 3.3. All test images are selected from Waterloo-IVC 3D Image Database Phase I, which are the case of strong asymmetric distortions (Group 3D.2) and mixed distortions (Group 3D.4). Each asymmetric image creates two test cases, with the left- and right-views exchanged. Altogether, there are totally 78 symmetric stereoscopic images and 144 pairs (288 singles) of asymmetrically distorted stereoscopic images.

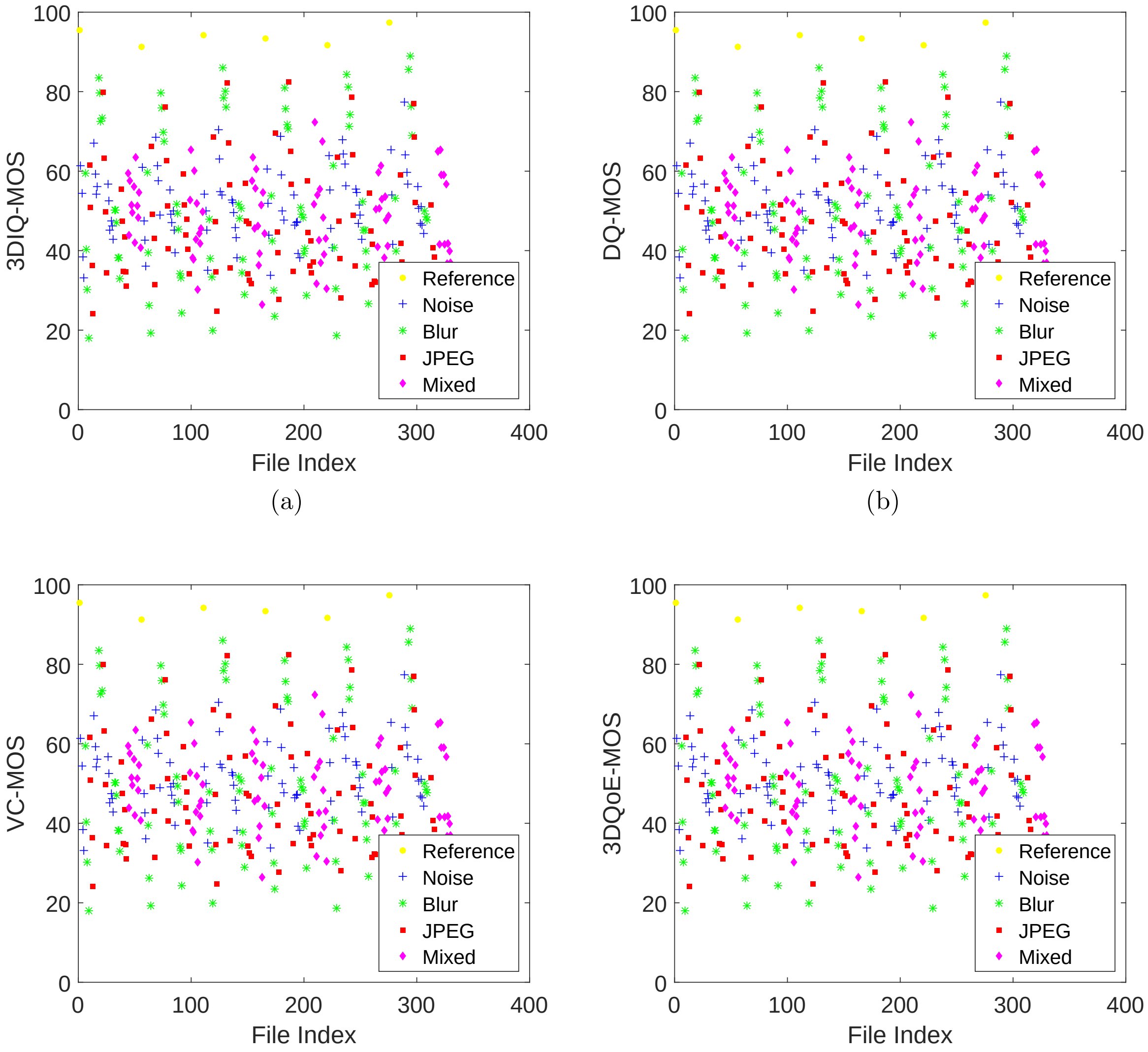
The 3DIQ MOS scores for each image were computed for left-eye dominant subjects and right-eye dominant subjects, denoted as $3\mathrm{DIQ}_{\mathrm{L}}$ and $3\mathrm{DIQ}_{\mathrm{R}}$, respectively. We employed the one-sample t-test to obtain a test decision for the null hypothesis that the difference between $3\mathrm{DIQ}_{\mathrm{L}}$ and $3\mathrm{DIQ}_{\mathrm{R}}$, i.e., $3\mathrm{DIQ}_{\mathrm{D}} = 3\mathrm{DIQ}_{\mathrm{L}} - 3\mathrm{DIQ}_{\mathrm{R}}$, comes from a normal distribution with of zero-mean and unknown variance. The alternative hypothesis is that the population distribution does not have a mean equaling zero. The result h is 1 if the test rejects the null hypothesis at the 5% significance level, and 0 otherwise. The returned p-values for symmetric and asymmetric images are 0.3801 and 0.1322, respectively, thus the null hypothesis cannot be rejected at the 5% significance level, which indicates that the impact of eye dominance in the perception of asymmetrically distorted stereoscopic images is not considered significant. This is consistent with the "stimulus" view of rivalry that is widely accepted in the field of visual neuroscience [15]. A comprehensive review and discussion on the question of "stimulus" rivalry versus "eye" rivalry can also be found in [15, 81].

3.5 Analysis and Key Observations

3.5.1 Subjective data analysis

The raw scores 2DIQ, 3DIQ, DQ, VC and 3DQoE given by each subject were converted to Z-scores and the entire data set was rescaled to fill the range from 1 to 100. The MOS scores with each 2D/3D visual criterion for each 2D and 3D image were then computed after removing outliers [101]. Figure 3.9 and Figure 3.10 show the distributions of MOS scores for each 3D visual criterion on Phase I and Phase II, respectively.

For each single-view image or stereopair, the standard deviation of Z-scores represents the degree of variation and the means of these standard deviations are reported in Ta-



(a) (b) (c) (d)

Figure 3.9: MOS scores of all stimuli in Waterloo-IVC 3D Image Database Phase I. (a) 3DIQ. (b) DQ. (c) VC. (d) 3DQoE.

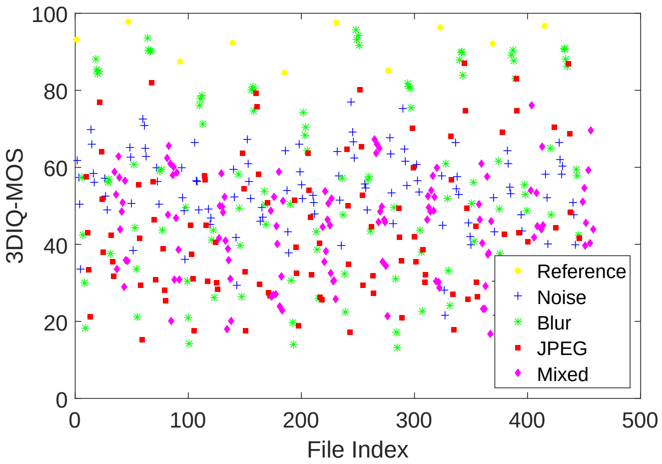


Figure 3.10: 3DIQ-MOS scores of all stimuli in Waterloo-IVC 3D Image Database Phase II.

ble 3.20, which indicate large variations in DQ, VC and 3DQoE scores. Table 3.21 reports the PLCC, SRCC and KRCC values between the MOS scores and the individual scores given by each subject, which reflect the degree of agreement among the subjects, where it can be observed that DQ and VC scores show less correlation with MOS compared with 2DIQ and 3DIQ scores and 3DQoE scores show more correlation than DQ and VC scores but less than 2DIQ and 3DIQ scores as well.

Table 3.20: Mean of Standard deviations of individual scores

	Criterion	Mean of standard deviation of individual scores
Phase I	2DIQ	9.7338
	3DIQ	12.0041
	DQ	20.0122
	VC	19.4743
	3DQoE	17.5495
Phase II	2DIQ	9.2411
	3DIQ	11.2551

Table 3.22 reports PLCC, SRCC and KRCC values between 3D visual experience criteria on individual scores, where it can be observed that 3DIQ and DQ have higher correla-

Table 3.21: Mean and Standard deviation of correlations between individual scores and MOS scores

Waterloo-IVC 3D Image Database Phase I						
	PLCC		SRCC		KRCC	
Criterion	Mean	Std.	Mean	Std.	Mean	Std.
2DIQ	0.9014	0.0241	0.8645	0.0341	0.7263	0.0410
3DIQ	0.8045	0.0675	0.7762	0.0745	0.6315	0.0713
DQ	0.7414	0.1447	0.7298	0.1481	0.5838	0.1292
VC	0.7222	0.0941	0.7068	0.0977	0.5623	0.0881
3DQoE	0.7732	0.1281	0.7603	0.1314	0.6148	0.1150

Waterloo-IVC 3D Image Database Phase II						
	PLCC		SRCC		KRCC	
Criterion	Mean	Std.	Mean	Std.	Mean	Std.
2DIQ	0.9111	0.0222	0.9086	0.0269	0.7825	0.0330
3DIQ	0.8789	0.0438	0.8847	0.0462	0.7410	0.0496

tions with 3DQoE, which is consistent with previous studies that the 3D visual experience can be predicted by a combination of image quality and depth quality [119]. Second, VC has the lowest correlations with other 3D visual experience criteria, indicating that VC is a more independent 3D visual criterion.

3.5.2 Key observation: Distortion type dependency

The raw 2DIQ and 3DIQ scores given by each subject were converted to Z-scores, respectively. Then the entire data sets were rescaled to fill the range from 1 to 100 and the MOS scores for each 2D and 3D image was computed after removing outliers. Given the subjective data, the main question we would like to ask in the current work is how the single-view 2D image quality predicts the 3D image quality (3DIQ scores in the subjective test), especially for the case of asymmetric distortions. The most straightforward 2D-to-3D quality prediction method is to average the MOSs of the left- and right-view images. Figure 3.11 (a) to Figure 3.15 (a) show the corresponding scatter plots for Waterloo-IVC database Phase I while Figure 3.11 (c) to Figure 3.15 (c) show the scatter plots for Waterloo-IVC database Phase II. Table 3.23 reports PLCC, SRCC, KRCC, RMSE and MAE between

Table 3.22: Correlations between 3D visual experience criteria on individual scores in Waterloo-IVC 3D Image Database Phase I

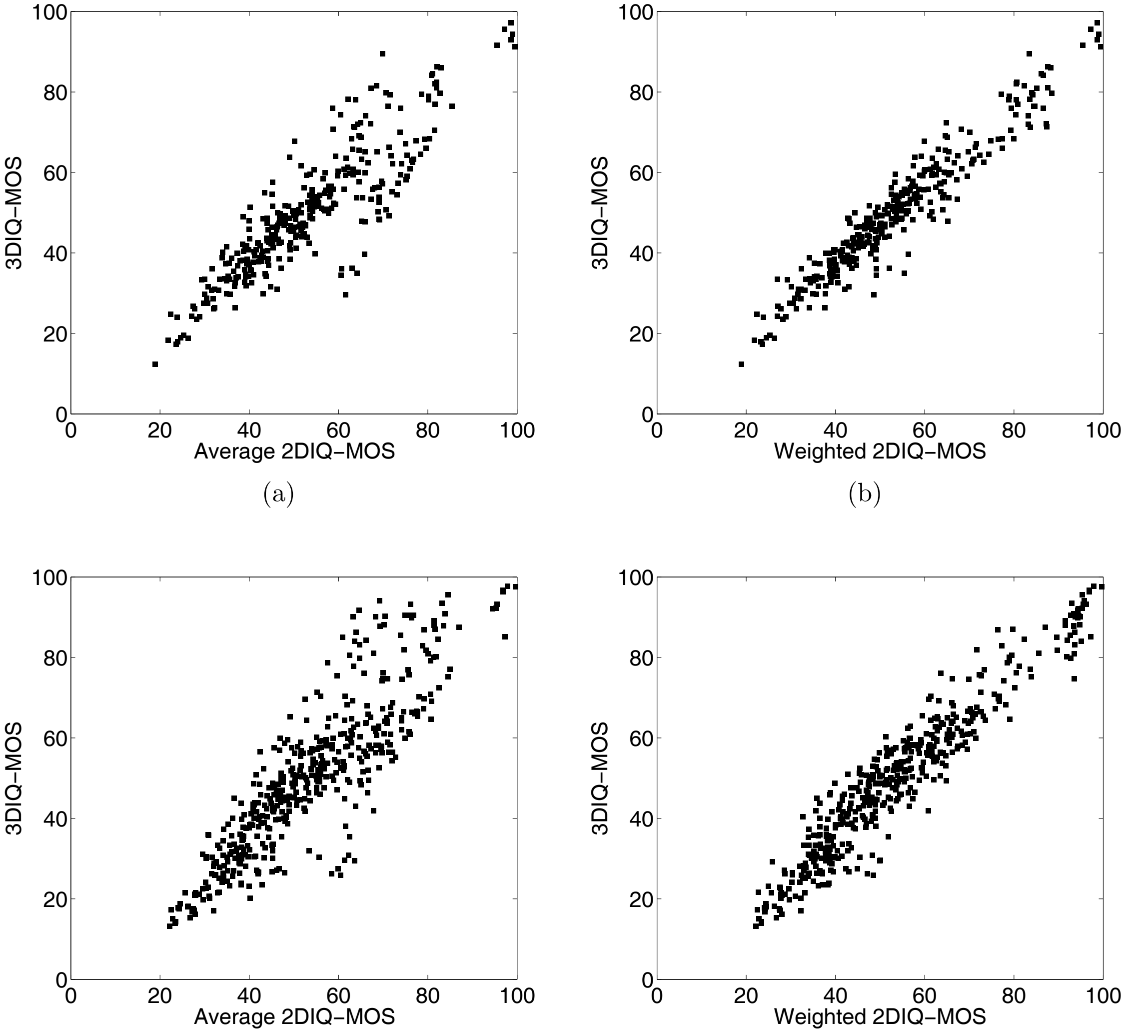
<table>
<tr><th colspan="5">PLCC</th></tr>
<tr><td></td><td>3DIQ</td><td>DQ</td><td>VC</td><td>3DQoE</td></tr>
<tr><td>3DIQ</td><td>1</td><td>0.6749</td><td>0.6590</td><td>0.7326</td></tr>
<tr><td>DQ</td><td>0.6749</td><td>1</td><td>0.6555</td><td>0.7235</td></tr>
<tr><td>VC</td><td>0.6590</td><td>0.6555</td><td>1</td><td>0.6579</td></tr>
<tr><td>3DQoE</td><td>0.7326</td><td>0.7235</td><td>0.6579</td><td>1</td></tr>
<tr><th colspan="5">SRCC</th></tr>
<tr><td></td><td>3DIQ</td><td>DQ</td><td>VC</td><td>3DQoE</td></tr>
<tr><td>3DIQ</td><td>1</td><td>0.6631</td><td>0.6427</td><td>0.7188</td></tr>
<tr><td>DQ</td><td>0.6631</td><td>1</td><td>0.6379</td><td>0.7151</td></tr>
<tr><td>VC</td><td>0.6427</td><td>0.6379</td><td>1</td><td>0.6503</td></tr>
<tr><td>3DQoE</td><td>0.7188</td><td>0.7151</td><td>0.6503</td><td>1</td></tr>
<tr><th colspan="5">KRCC</th></tr>
<tr><td></td><td>3DIQ</td><td>DQ</td><td>VC</td><td>3DQoE</td></tr>
<tr><td>3DIQ</td><td>1</td><td>0.5564</td><td>0.5431</td><td>0.6116</td></tr>
<tr><td>DQ</td><td>0.5564</td><td>1</td><td>0.5362</td><td>0.6076</td></tr>
<tr><td>VC</td><td>0.5431</td><td>0.5362</td><td>1</td><td>0.5457</td></tr>
<tr><td>3DQoE</td><td>0.6116</td><td>0.6076</td><td>0.5457</td><td>1</td></tr>
</table>

Table 3.23: Performance comparison of 2D-to-3D quality prediction models on Waterloo-IVC 3D Image Database

	Waterloo-IVC 3D Image Database Phase I					Waterloo-IVC 3D Image Database Phase II				
Group	PLCC	SRCC	KRCC	RMSE	MAE	PLCC	SRCC	KRCC	RMSE	MAE
All 3D	0.8835	0.8765	0.7161	7.3700	5.3293	0.8763	0.8820	0.7145	9.2201	6.7164
Symmetric	0.9801	0.9657	0.8482	3.8266	2.9160	0.9799	0.9696	0.8557	4.0525	3.1712
Asymmetric	0.8572	0.8471	0.6780	7.9895	5.8643	0.8418	0.8501	0.6693	10.4221	7.9037
Group 3D.1	0.9801	0.9657	0.8482	3.8266	2.9160	0.9799	0.9696	0.8557	4.0525	3.1712
Group 3D.2	0.6613	0.5433	0.4406	11.6394	9.0069	0.6121	0.5874	0.4524	14.2881	11.6204
Group 3D.3	0.9666	0.9164	0.7597	3.6078	2.7578	0.9471	0.8898	0.7176	5.7909	4.4262
Group 3D.4	0.9223	0.8271	0.6390	5.9710	4.5387	0.9225	0.8798	0.7047	7.0928	5.4776

3DIQ-MOS scores and the average 2DIQ-MOS scores, including the results for all stereoscopic images and for each test image group. PLCC, RMSE and MAE are adopted to evaluate prediction accuracy [46] and SRCC and KRCC are employed to assess prediction monotonicity [46]. Higher PLCC, SRCC and KRCC or lower RMSE and MAE values indicate better consistency with human opinions of quality. PLCC, RMSE and MAE are usually computed after a nonlinear mapping between the subjective and objective scores and the results may be sensitive to the choice of the mapping function. SRCC and KRCC are nonparametric rank order-based correlation metrics, independent of any monotonic nonlinear mapping between subjective and objective scores but do not explicitly estimate the accuracy of quality prediction.

From Table 3.23 and Figure 3.11 to Figure 3.15, it can be observed that the best prediction occurs in Group 3D.1, which is the category for symmetrically distorted 3D images (consistent with the literature [90, 23]). By contrast, the PLCC, SRCC, KRCC, RMSE and MAE values drop significantly in other test groups (corresponding to asymmetrical distortions) as well as in the all-image group. The drops of correlation coefficient values are also reflected in the scatter plots shown in Figure 3.11 to Figure 3.15, where this simple averaging prediction model generates substantial bias of many stereopairs. Most interestingly, this bias leans towards opposite directions, largely depending on the distortion types. In particular, for noise contamination and JPEG compression, average prediction overestimates 3D quality of many images (or 3D image quality is more affected by the poorer quality view), while for blur, average prediction often underestimates 3D image quality (or 3D image quality is more affected by the better quality view). Furthermore, Table 3.23 suggests that the worst performance occurs in Group 3D.2, where only one view image is



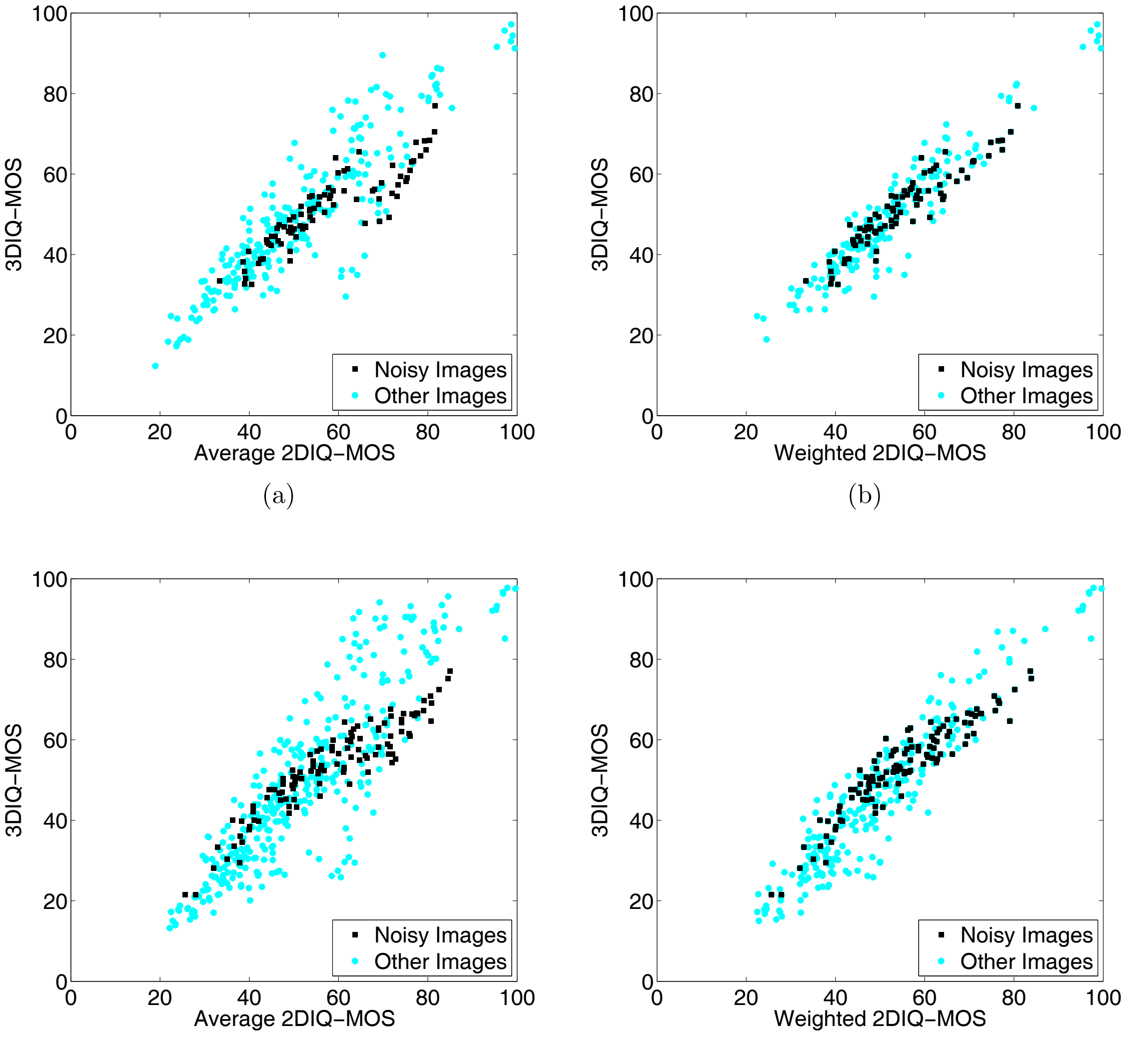
(a)

(b)

(c)

(d)

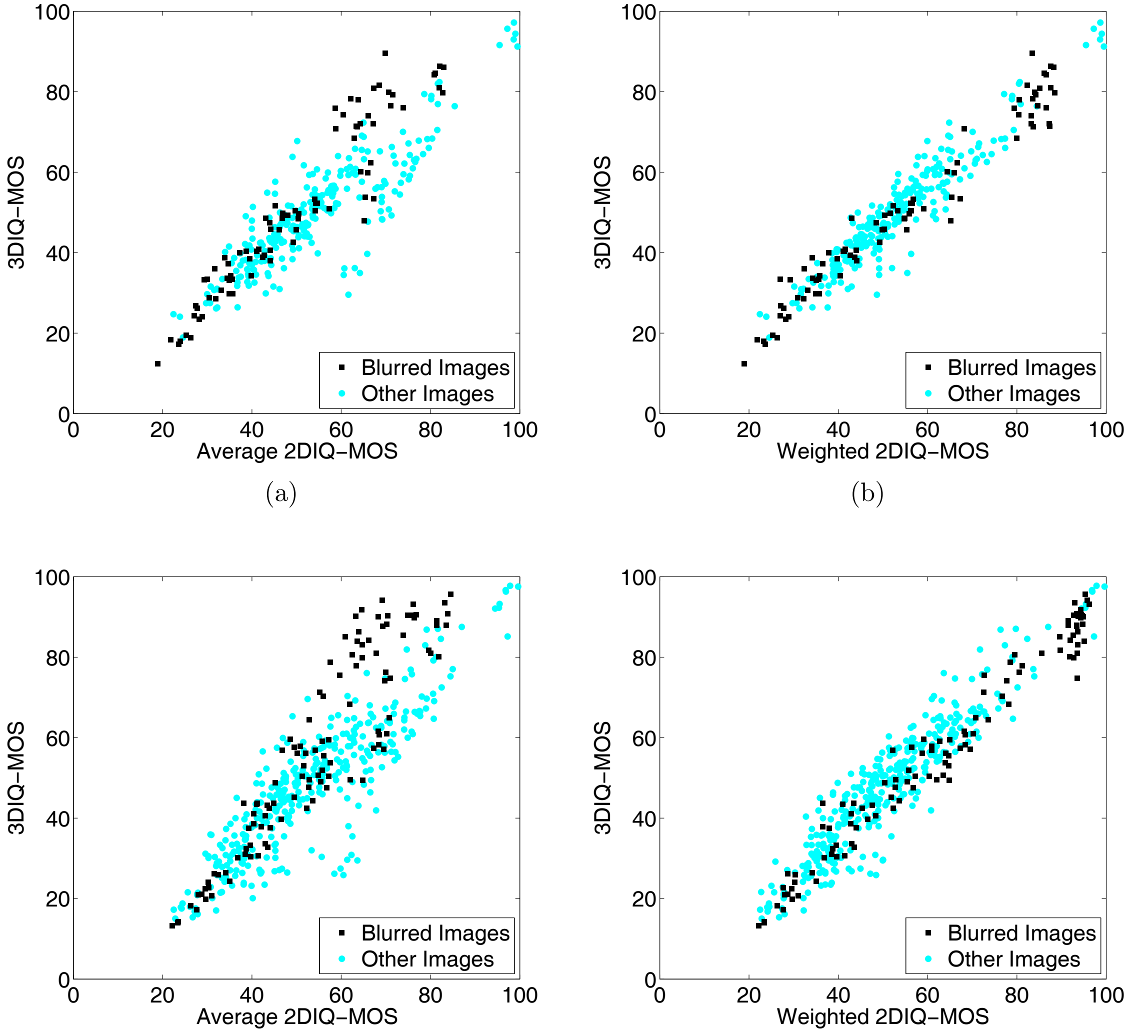
Figure 3.11: 3DIQ-MOS versus predictions from 2DIQ-MOS of 2D left- and right-views. (a) All Images, Average 2DIQ-MOS on Waterloo-IVC Phase I. (b) All Images, Weighted 2DIQ-MOS on Waterloo-IVC Phase I. (c) All Images, Average 2DIQ-MOS on Waterloo-IVC Phase II. (d) All Images, Weighted 2DIQ-MOS on Waterloo-IVC Phase II.



(a) (b)

(c) (d)

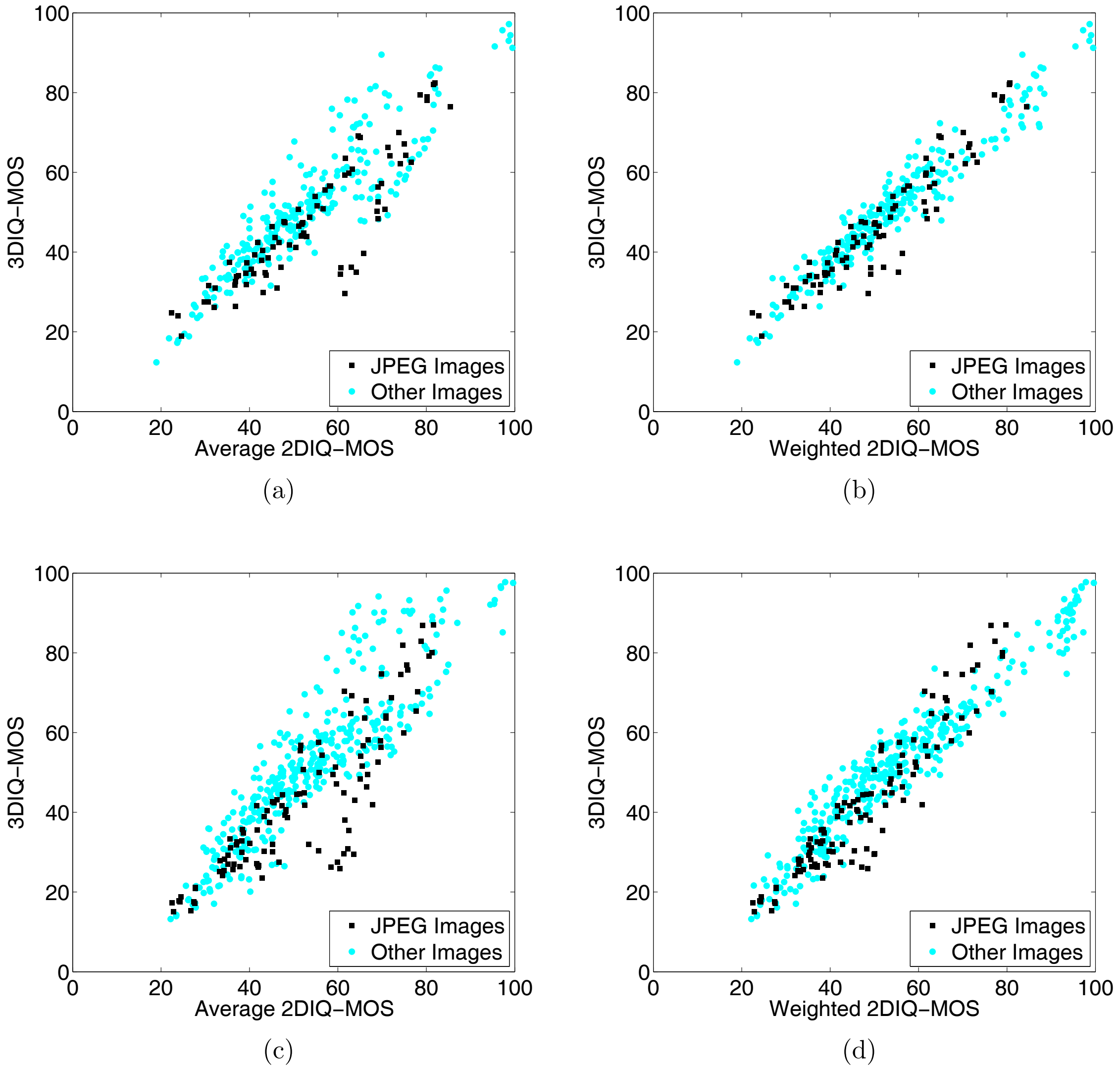
Figure 3.12: 3DIQ-MOS versus predictions from 2DIQ-MOS of 2D left- and right-views. (a) Noisy Images, Average 2DIQ-MOS on Waterloo-IVC Phase I. (b) Noisy Images, Weighted 2DIQ-MOS on Waterloo-IVC Phase I. (c) Noisy Images, Average 2DIQ-MOS on Waterloo-IVC Phase II. (d) Noisy Images, Weighted 2DIQ-MOS on Waterloo-IVC Phase II.



(a) (b)

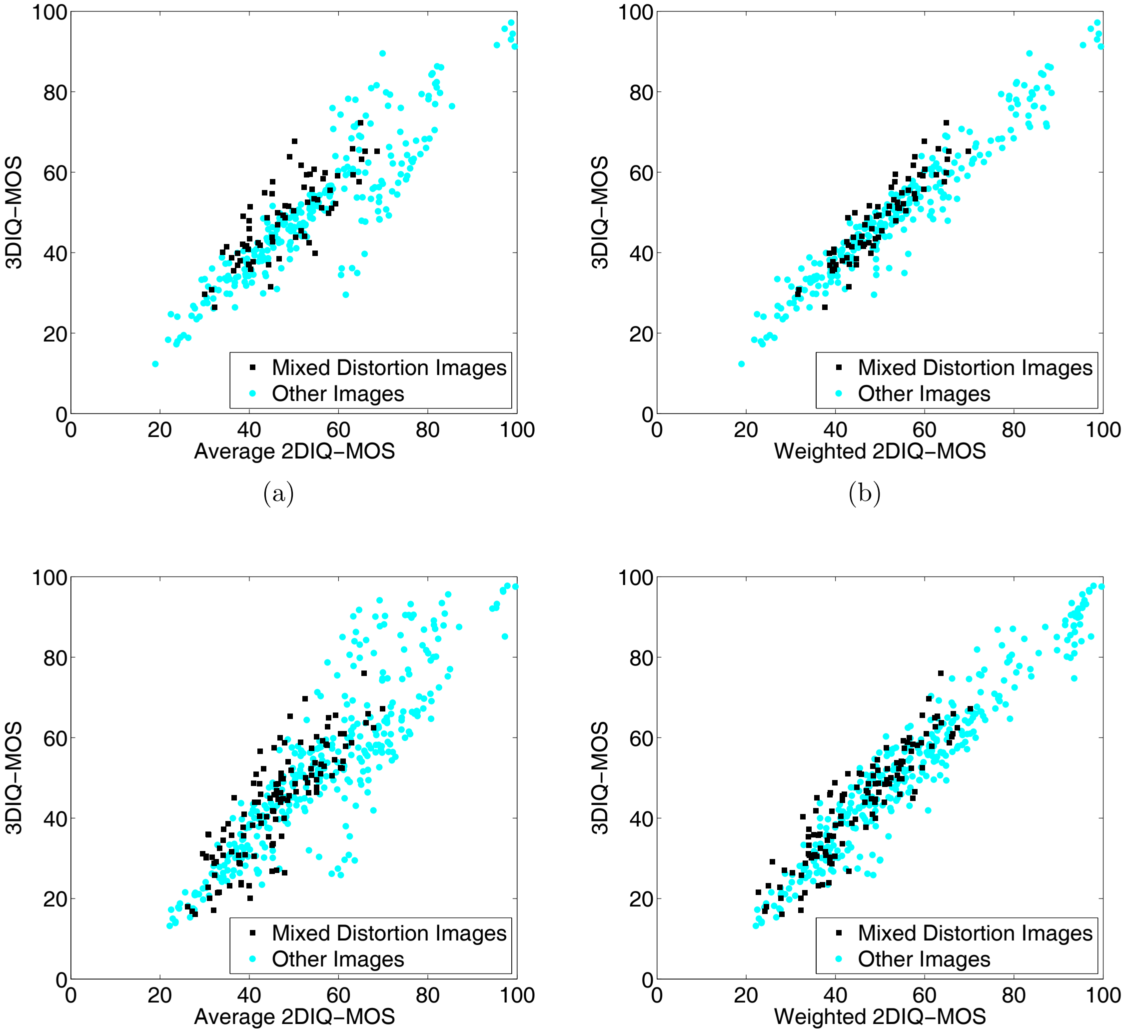
(c) (d)

Figure 3.13: 3DIQ-MOS versus predictions from 2DIQ-MOS of 2D left- and right-views. (a) Blurred Images, Average 2DIQ-MOS on Waterloo-IVC Phase I. (b) Blurred Images, Weighted 2DIQ-MOS on Waterloo-IVC Phase I. (c) Blurred Images, Average 2DIQ-MOS on Waterloo-IVC Phase II. (d) Blurred Images, Weighted 2DIQ-MOS on Waterloo-IVC Phase II.



(a) (b) (c) (d)

Figure 3.14: 3DIQ-MOS versus predictions from 2DIQ-MOS of 2D left- and right-views. (a) JPEG Images, Average 2DIQ-MOS on Waterloo-IVC Phase I. (b) JPEG Images, Weighted 2DIQ-MOS on Waterloo-IVC Phase I. (c) JPEG Images, Average 2DIQ-MOS on Waterloo-IVC Phase II. (d) JPEG Images, Weighted 2DIQ-MOS on Waterloo-IVC Phase II.



(a) (b) (c) (d)

Figure 3.15: 3DIQ-MOS versus predictions from 2DIQ-MOS of 2D left- and right-views. (a) Mixed Distortion, Average 2DIQ-MOS on Waterloo-IVC Phase I. (b) Mixed Distortion, Weighted 2DIQ-MOS on Waterloo-IVC Phase I. (c) Mixed Distortion, Average 2DIQ-MOS on Waterloo-IVC Phase II. (d) Mixed Distortion, Weighted 2DIQ-MOS on Waterloo-IVC Phase II.

distorted and thus the quality difference between two views is maximized.

It is interesting to compare our observations regarding distortion type dependency with those published in the literature. For image blur, it was reported in [148, 83] that 3D image quality is less affected by the view with lower quality, which is consistent with our result. For image blockiness from JPEG compression, in [120], it is claimed that 3D image quality is approximately the average of the higher quality and the lower quality but there is a tendency towards the lower quality view, which is consistent with our observations especially when one of view is highly compressed and the other keeps uncompressed. Meanwhile, in [148], no bias was discovered when lower levels of asymmetric JPEG compression were evaluated. These seemingly controversial results are well explained by the scatter plots shown in Figure 3.14 and the 2D-line plot shown in Figure 3.16, which shows 3DIQ-MOS versus average 2DIQ-MOS minus 3DIQ-MOS for the case of strong asymmetric compressions (Group 3D.2) on Waterloo-IVC Phase I and Phase II. From these figures, it can be observed that the bias of the averaging prediction model increases with the level of distortions, and thus the strength of the bias is pronounced depending on the quality range being investigated.

3.6 Objective Study I: 2D-to-3D Quality Prediction

We opt to use a two-stage approach in the design of an objective 3DIQ predictor. The first stage aims to evaluate the perceptual quality of single-view images, while in the second stage, a binocular rivalry inspired multi-scale model is developed to combine 2D image quality of both views into a quality estimation of 3D image quality.

3.6.1 Objective 2D Quality Assessment

In the literature, the SSIM index [153] as well as its derivatives MS-SSIM [158] and IW-SSIM [154] have demonstrated competitive performance in 2D objective IQA tests [154]. An advantage of the SSIM approach is that it provides a quality map that indicates the variations of image quality over space [153]. It was shown that spatial pooling built upon

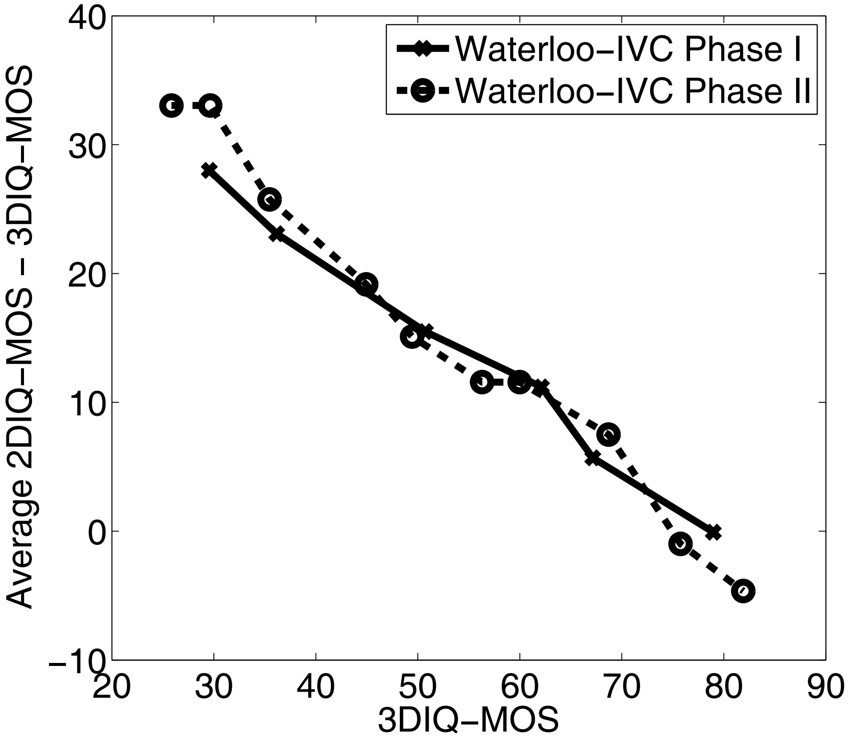


Figure 3.16: 3DIQ-MOS versus average 2DIQ-MOS - 3DIQ-MOS for Group 3D.2 on Waterloo-IVC 3D Image Database Phase I and Phase II.

the quality map based on information content weighting or distortion weighting further improves the performance [154]. Here we build our 2D-IQA model upon SSIM, but improve it further by incorporating an information content and divisive normalization based pooling scheme.

A general form of spatially weighted pooling is given by

$$Q^{2D} = \frac{\sum_{i=1}^{N} w_i q_i}{\sum_{i=1}^{N} w_i}, \tag{3.1}$$

where q_i and w_i are the local quality value (e.g., local SSIM value) and the weight assigned to the i-th spatial location (i-th pixel), respectively. The assumption behind information content weighted pooling is that the spatial locations that contain more information are more likely to attract visual attention, and thus should be given larger weights. Let $\mathbf{x}_i$ and $\mathbf{y}_i$ be the local image patches extracted around the i-th spatial location from the reference and the distorted images, respectively. Following the information content evaluation method in [156], we compute the weighting factor by

$$w_i^{ic} = \log\left[\left(1 + \frac{\sigma_{x_i}^2}{C}\right)\left(1 + \frac{\sigma_{y_i}^2}{C}\right)\right], \tag{3.2}$$

where σ_{x_i} and σ_{y_i} are the standard deviations of $\mathbf{x}_i$ and $\mathbf{y}_i$, respectively, and C is the noisy visual channel power.

Another useful pooling strategy is distortion weighted pooling, which is based on the intuitive idea that the spatial locations that contain more distortions are more likely to attract visual attention, and thus should be given more weights. Since the local quality has been gauged by q_i (e.g., the SSIM value at location i), it is straightforward to convert it to a local distortion measure, for example, let $d_i = 1 - \mathrm{SSIM}_i$. Divisive normalization has been recognized as a perceptually and statistically motivated non-linear transformation [140]. We apply divisive normalization to the local distortion map and define a normalized distortion based weighting factor by

$$w_i^d = \frac{d_i}{\sqrt{\sum_{j \in \mathcal{N}_i} d_j^2 + D_0}}, \tag{3.3}$$

where $\mathcal{N}_i$ denotes the set of neighboring pixels surrounding the i-th spatial location, and D_0 is a stability constant.

The final weighting factor is obtained by combining information content and divisive normalization-based distortion weighting factors

$$w_i = \max\left\{\left(w_i^{ic}\right)^2, \left(w_i^{d}\right)^2\right\} . \tag{3.4}$$

where the max operation is based on the strategy to choose either w^{ic} or w^d, depending on which one is more significant.

Applying this weighted pooling approach to the SSIM map, we obtain an Information Content and Distortion Weighted Structural Similarity Index (IDW-SSIM) measure. This has led to significant performance improvement when tested using the single-view images in our new Waterloo-IVC 3D Image Database Phase I and Phase II. Quantitative measures of PLCC, SRCC, KRCC, RMSE and MAE can be found in Table 3.24.

3.6.2 2D-to-3D Quality Prediction Model

The competition between binocular fusion and binocular rivalry [66] provides a potential theory to develop 2D-to-3D quality prediction models. When the left- and right-view images are consistent, they are fused in the visual system to a single percept of the scene, known as binocular fusion. On the other hand, when the images of the two views are inconsistent, instead of the two images being seen superimposed, one of them may dominate or two images may be seen alternately, known as binocular rivalry [66]. Although there is a rich literature on binocular fusion and rivalry in biological vision science [66] and [60] (where simple and ideal visual stimuli are often used), how to apply the principle to 3D-IQA remains an active research topic. Since in 3D-IQA we need to work on complicated scenes and distortions, simplifications are essential to create practical solutions.

As shown in Figure 3.17, our work is motivated by existing vision studies on binocular rivalry [72, 14, 37], and [35], where it was found that for simple ideal stimuli, an increasing contrast increases the predominance of one view against the other. Also note that in complicated scenes the contrast of a signal increases with its signal strength measured

Table 3.24: Performance comparison of 2D-IQA models on Waterloo-IVC 3D Image Database (Single-view Images) and other 2D-IQA databases

	Waterloo-IVC 3D Image Database Phase I					Waterloo-IVC 3D Image Database Phase II				
2D-IQA	PLCC	SRCC	KRCC	RMSE	MAE	PLCC	SRCC	KRCC	RMSE	MAE
PSNR	0.7862	0.4800	0.3552	11.8402	9.5212	0.7086	0.3885	0.2832	13.1298	11.1620
SSIM	0.8740	0.7726	0.5715	9.3099	7.7367	0.7643	0.6232	0.4441	11.9994	10.3491
MS-SSIM	0.8440	0.6402	0.4821	10.2769	7.8258	0.7429	0.5049	0.3590	12.4564	10.6135
IW-SSIM	0.8615	0.7696	0.5909	9.7299	8.3574	0.7683	0.6310	0.4640	11.9118	10.1677
IDW-SSIM	0.9572	0.9311	0.7605	5.5437	4.4359	0.9352	0.9327	0.7758	5.6271	4.3880
	Cornell A57 Database					CSIQ Database				
2D-IQA	PLCC	SRCC	KRCC	RMSE	MAE	PLCC	SRCC	KRCC	RMSE	MAE
PSNR	0.6346	0.6176	0.4301	0.1900	0.1607	0.7605	0.8053	0.6081	0.1705	0.1339
SSIM	0.8019	0.8067	0.6063	0.1469	0.1209	0.8569	0.8718	0.6858	0.1354	0.1007
MS-SSIM	0.8604	0.8415	0.6483	0.1253	0.1007	0.8903	0.9061	0.7298	0.1196	0.0897
IW-SSIM	0.9035	0.8713	0.6846	0.1054	0.0892	0.9017	0.9108	0.7390	0.1135	0.0841
IDW-SSIM	0.8842	0.8683	0.6832	0.1148	0.0961	0.8756	0.8884	0.7060	0.1268	0.0931
	IVC Database					LIVE Database				
2D-IQA	PLCC	SRCC	KRCC	RMSE	MAE	PLCC	SRCC	KRCC	RMSE	MAE
PSNR	0.6719	0.6884	0.5218	0.9023	0.7191	0.9302	0.9092	0.7484	8.4819	6.3731
SSIM	0.9119	0.9018	0.7223	0.4999	0.3777	0.9455	0.9496	0.8149	7.5252	6.0247
MS-SSIM	0.9108	0.8980	0.7203	0.5029	0.3813	0.9468	0.9512	0.8181	7.4379	5.9897
IW-SSIM	0.9231	0.9125	0.7339	0.4686	0.3694	0.9515	0.9604	0.8379	7.1116	5.6548
IDW-SSIM	0.9198	0.9081	0.7301	0.4780	0.3696	0.9576	0.9627	0.8381	6.6573	5.0979
	TID 2008 Database					TID 2013 Database				
2D-IQA	PLCC	SRCC	KRCC	RMSE	MAE	PLCC	SRCC	KRCC	RMSE	MAE
PSNR	0.5235	0.5531	0.4027	1.1434	0.8680	0.6775	0.6394	0.4699	0.9119	0.6801
SSIM	0.7732	0.7749	0.5768	0.8511	0.6546	0.7895	0.7417	0.5588	0.7608	0.5926
MS-SSIM	0.8451	0.5768	0.6568	0.7173	0.5578	0.8329	0.7859	0.6047	0.6861	0.5309
IW-SSIM	0.8579	0.8511	0.6636	0.6895	0.5276	0.8319	0.7779	0.5977	0.6880	0.5290
IDW-SSIM	0.7427	0.6546	0.5718	0.8985	0.6571	0.7569	0.7267	0.5479	0.8102	0.6069
	Toyama-MICT Database									
2D-IQA	PLCC	SRCC	KRCC	RMSE	MAE					
PSNR	0.7328	0.7221	0.5398	0.8983	0.6999					
SSIM	0.9125	0.9023	0.7250	0.5401	0.4030					
MS-SSIM	0.9154	0.9070	0.7316	0.5315	0.3995					
IW-SSIM	0.9401	0.9289	0.7693	0.4502	0.3437					
IDW-SSIM	0.9132	0.9030	0.7272	0.5381	0.4012					

using energy. This inspires us to hypothesize that the strength of view dominance in binocular rivalry of stereoscopic images is related to the relative energy of the two views.

The diagram of the proposed method is shown in Figure 3.18. Let ($I_{r,l}$, $I_{r,r}$) and ($I_{d,l}$, $I_{d,r}$) be the left- and right-view image pairs of the reference and distorted stereoscopic images, respectively. We first create their local energy maps by computing the local variances at each spatial location, i.e., the variances of local image patches extracted around each spatial location from the reference or the distorted images are computed, for which an 11×11 circular-symmetric Gaussian weighting function $\mathbf{w} = \{w_i | i = 1, 2, \cdots, N\}$ with standard deviation of 1.5 samples, normalized to unit sum ($\sum_{i=1}^{N} w_i = 1$), is employed. The resulting energy maps are denoted as $E_{r,l}$, $E_{r,r}$, $E_{d,l}$ and $E_{d,r}$, respectively.

The local energy maps E are displayed in Figure 3.19 using the reference or distorted images as the background, and the pixels with local energy larger than 50 are highlighted as black. In the reference energy map, the high-energy are concentrated at structural regions; in the noise energy map, the high-energy are across the whole image; in the blur energy map, the high-energy appears only in areas surrounding the structural regions; and in the JPEG energy map, the high-energy not only appears on structural regions but also on at regions as blocking artifacts. In general, compared with the reference image, blur images contain less energy while noise or JPEG images have more energy.

The local energy ratio maps $R = E_d/E_r$ are displayed in Figure 3.20 using the reference or distorted images as the back ground, and the pixels with local energy ratio larger than 1 are highlighted as black. It can be observed that for noisy images, the energy ratios are all larger than 1 while for JPEG images, the energy ratios are larger than 1 specifically on both structural regions and areas with blocking artifact. For blur images, larger energy ratios only appear in areas surrounding structural regions, which may be explained by the leakage of energy from highly structural regions to their surrounding areas due to low-pass filtering. Thus these local energy ratio maps are spatially similar with local energy maps while as the normalized signals, these local energy ratios are used in the pooling stage.

Examples are given in Figure 3.21 and Figure 3.22, where the reference or distorted images are used as the background, and the pixels with local energy larger than 50 are highlighted as black. In Figure 3.21, the left-view is original and the right-view is blurred.

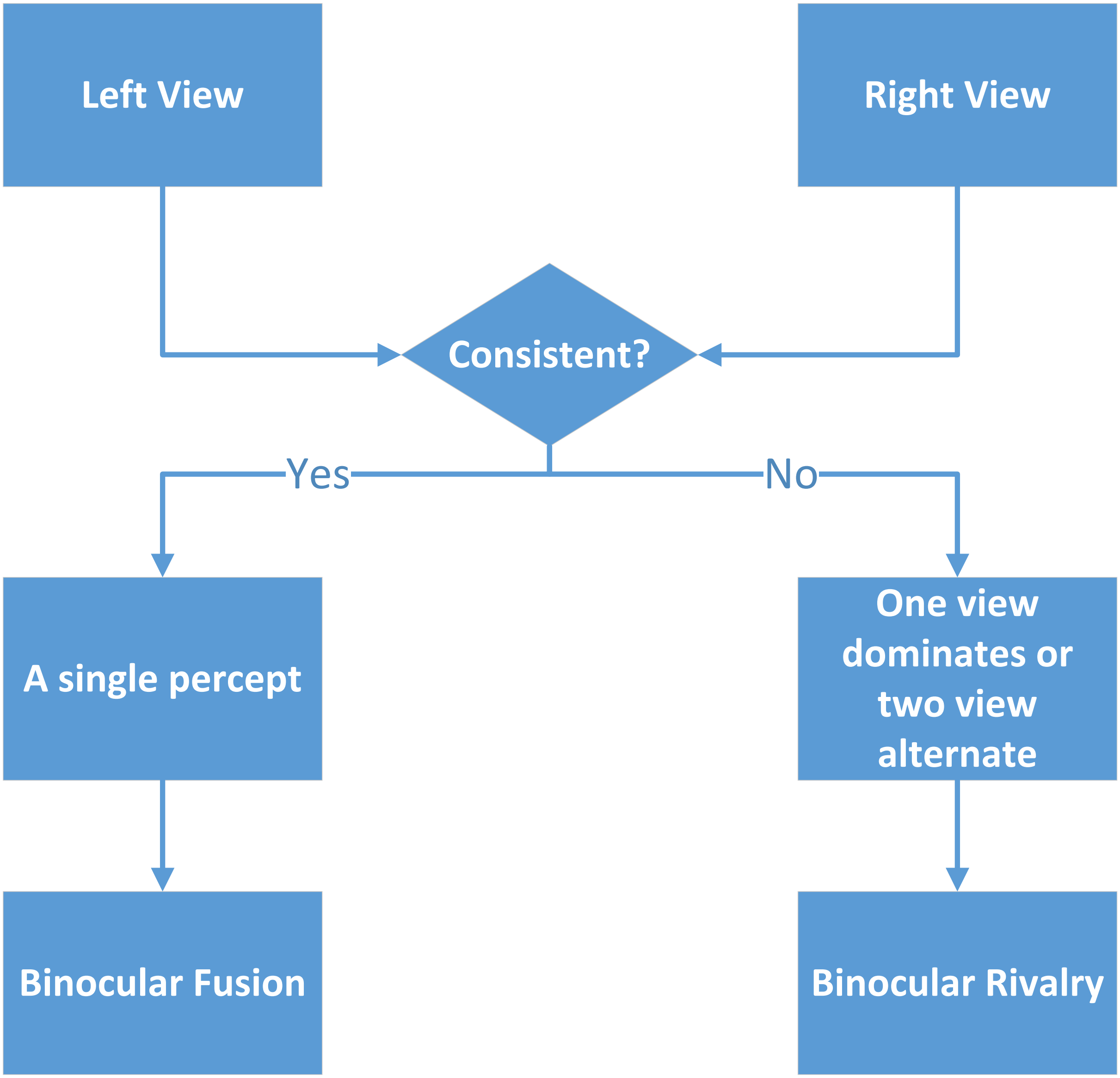


Figure 3.17: Binocular fusion versus Binocular Rivalry.

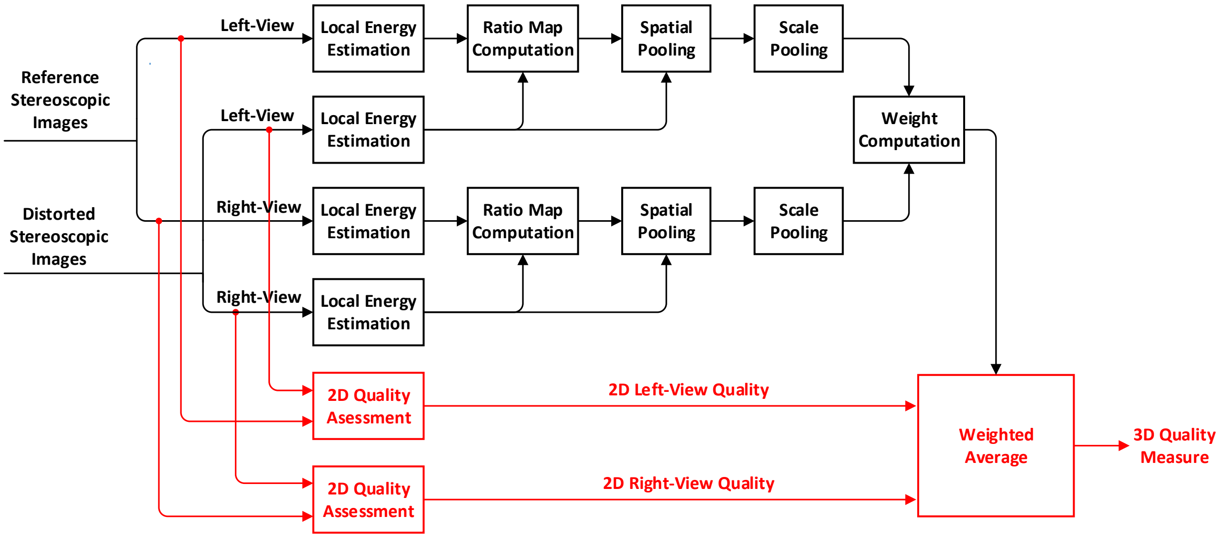


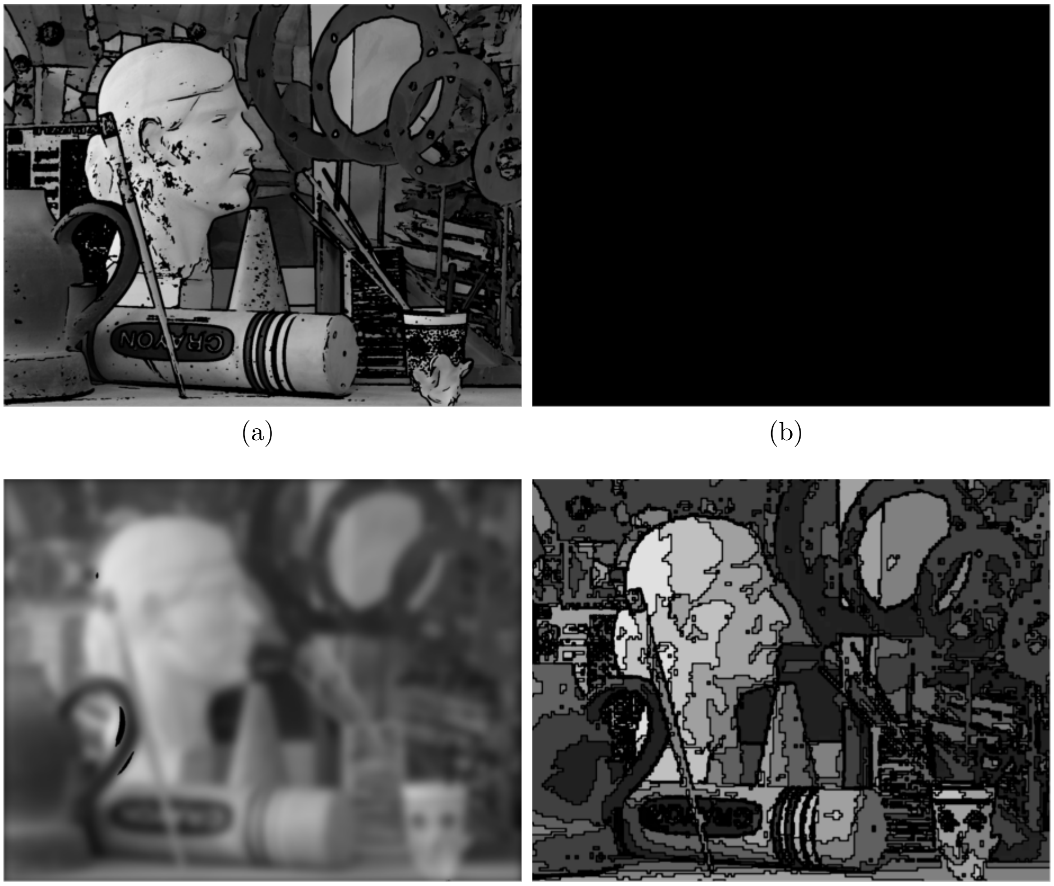
Figure 3.18: Diagram of the proposed 2D-to-3D quality prediction model.

The structural consistency between two-views is affected. The left-view may dominate the right-view at any time instance. In Figure 3.21, the left-view is original and the right-view is JPEG compressed. The structural consistency between two-views is affected. The right-view may dominate the left-view at any time instance.

Assume that the reference stereopair has perfect quality with strong 3D effect, where binocular fusion prevails. When at least one of the single-view images is distorted at some spatial locations, the distortion may affect the consistency between the image structures from the two views, and thus binocular rivalry prevails. As a result, one view may dominate the other at any time instance. Based on our hypothesis, we compute the local energy ratio maps in both views:

$$R_l = \frac{E_{d,l}}{E_{r,l}} \qquad \text{and} \qquad R_r = \frac{E_{d,r}}{E_{r,r}} \,. \tag{3.5}$$

The energy ratio maps provide useful local binocular rivalry information, which may be combined with the qualities of single-view images to predict 3D image quality. A pooling stage is necessary for this purpose. High-energy image regions are likely to contain more information. If the ultimate goal of visual perception is to efficiently extract useful information from the visual scene, then the high-energy regions are more likely to attract visual



(a) (b)

(c) (d)

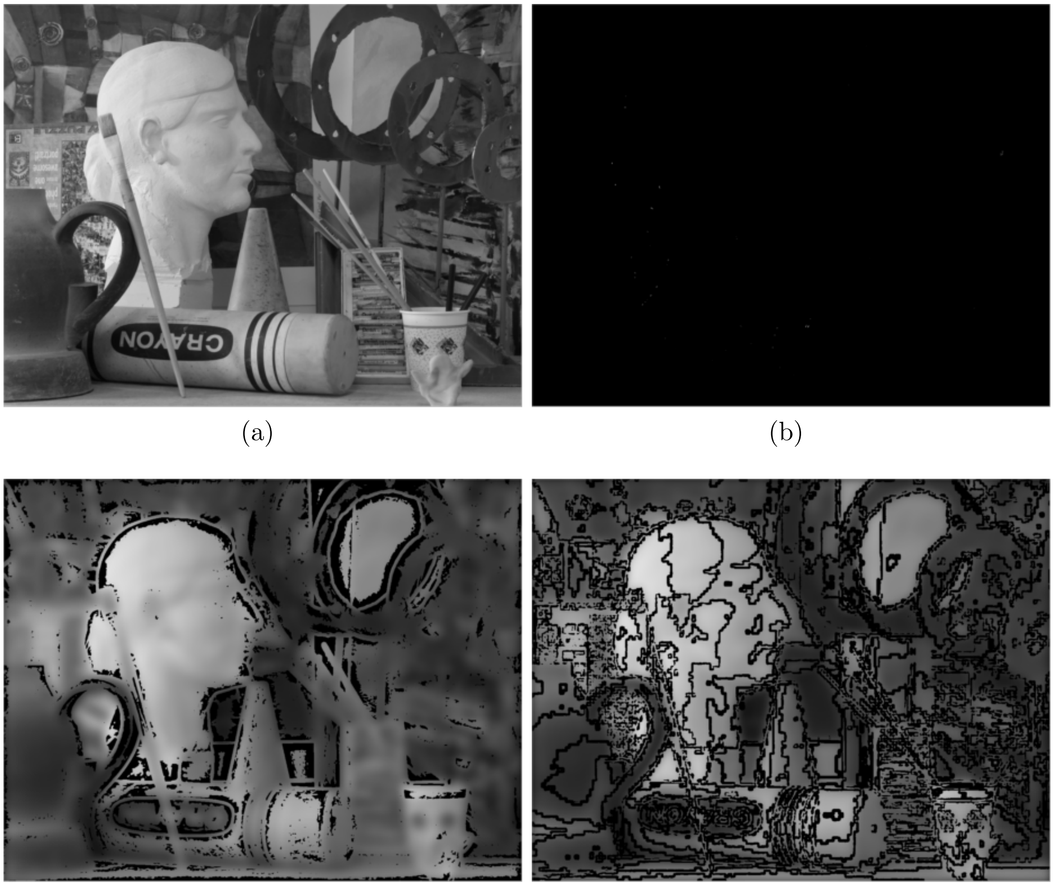
Figure 3.19: Local Energy Map E. (a) Reference. (b) Noise. (c) Blur. (d) JPEG.

attention, and thus should be given more importance. To emphasize on the importance of high-energy image regions in binocular rivalry, we adopt an energy weighted pooling method given by

$$g_l = \frac{\sum E_{d,l} R_l}{\sum E_{d,l}} \qquad \text{and} \qquad g_r = \frac{\sum E_{d,r} R_r}{\sum E_{d,r}}\,, \tag{3.6}$$

where the summations are over the full energy and ratio maps. Here g_l and g_r are estimations of the level of dominance of the left- and right-views, respectively.

Meanwhile, the study presented in [37] suggests that the dominance of one view over



(a) (b)

(c) (d)

Figure 3.20: Local Energy Ratio Map R. (a) Reference. (b) Noise. (c) Blur. (d) JPEG.

the other in binocular rivalry depends on the spatial frequency content of the stimuli used. Psychophysical experiments have shown that the spatial frequency sensitivity of human stereopsis behaves similar to the visual contrast sensitivity function (CSF) [165], which accounts for the visual contrast sensitivity as a function of spatial frequency [9]. We take advantage of this similarity and treat different spatial frequency subbands based on CSF. Specifically, we divide an image into multiple scales, by employing an iterative low-pass filtering and downsampling procedure, and subsequently calculate the level of view dominance for every subband of the left- and right-view images. We then combine the

(a) (b)

Figure 3.21: Binocular Rivalry Example 1. (a) Left-view: Reference. (b) Right-view: Blur

scale level dominance values of the left- and right-view images and the overall levels of dominance g_l and g_r are given by:

$$g_l = \sum_{i=1}^{N_s} \alpha_i g_{i,l} \qquad \text{and} \qquad g_r = \sum_{i=1}^{N_s} \alpha_i g_{i,r}\,, \tag{3.7}$$

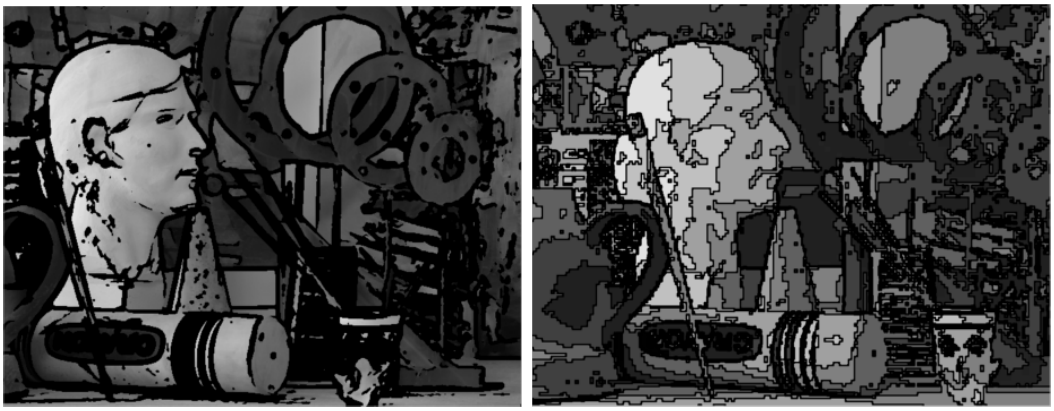
where $g_{i,l}$ and $g_{i,r}$ denote the levels of dominance of the i^{th} scale of the left- and right-view images, respectively. N_s is the number of scales and α_i denotes the perceptual importance of the i^{th} scale determined using the CSF formula given by [9]:

$$S(u) = \frac{5200 e^{(-0.0016u^2(1+100/L)^{0.08})}}{\sqrt{\left(1+\frac{144}{X_o^2}+0.64u^2\right)\left(\frac{63}{L^{0.83}}+\frac{1}{1-e^{(-0.02u^2)}}\right)}}\,, \tag{3.8}$$

where u, L, and X_o^2 denote spatial frequency in cycles/degree, luminance in cd/m^2, and angular object area in square degrees, respectively. α_i values are calculated using

$$\alpha_i = S(f_i^{\mathrm{c}})\,, \tag{3.9}$$

where f_i^{c} denotes the center spatial frequency of the i^{th} scale.



(a) (b)

Figure 3.22: Binocular Rivalry Example 2. (a) Left-view: Reference. (b) Right-view: JPEG.

Given the values of g_l and g_r, the weights assigned to the left- and right-view images are given by

$$w_l = \frac{g_l^2}{g_l^2 + g_r^2} \qquad \text{and} \qquad w_r = \frac{g_r^2}{g_l^2 + g_r^2}\,, \tag{3.10}$$

respectively.

Finally, the overall prediction of 3D image quality is calculated by a weighted average of the left- and right-view image quality:

$$Q^{3D} = w_l Q_l^{2D} + w_r Q_r^{2D}\,, \tag{3.11}$$

where Q_l^{2D} and Q_r^{2D} denote the 2D image quality of the left- and right-views, respectively.

In our previous work, preliminary results on simplified single-scale model of the proposed approach that ignores the variation of visual sensitivity across scales were reported in [145] and [144].

3.6.3 Validation

We use two 3D image quality databases to test the proposed algorithm, which are the new Waterloo-IVC 3D Image Database (Phase I and Phase II) and the LIVE 3D Image Database Phase II [23]. The latter is a recent database that contains both symmetrically and asymmetrically distorted images. Note that the parameters of the proposed 2D-to-3D quality prediction method are selected empirically when working with Waterloo-IVC 3D Image Database Phase I, but are completely independent of the Waterloo-IVC 3D Image Database Phase II and the LIVE 3D Image Database Phase II.

We test the proposed 2D-to-3D quality prediction model on all 3D images in Waterloo-IVC 3D Image Database Phase I and Phase II by applying it to the ground truth 2DIQ-MOS scores. The PLCC, SRCC, KRCC, RMSE and MAE values between 3DIQ-MOS and the predicted Q^{3D} value for all stereoscopic images and for each test image group are given in Table 3.25. The corresponding scatter plots are shown in Figure 3.11 to Figure 3.15. From Table 3.25 and Figure 3.11 to Figure 3.15, it can be observed that the proposed model outperforms the direct averaging method in almost all cases, and the improvement is most pronounced in the case of strong asymmetric distortions (Group 3D.2) or when all test images are put together (All 3D image case). By comparing different figures in Figure 3.11 to Figure 3.15, we observe how the proposed 2D-to-3D prediction model affects each image distortion type. For different distortion types, although the direct averaging method produces different levels of quality prediction biases towards different directions, the proposed method, which does not attempt to recognize the distortion types or give any specific treatment for any specific distortion type, removes or significantly reduces the prediction biases for all distortion types. Moreover, as mentioned earlier, the mixed distortion case provides the strongest test on the generalization ability of the model, for which the proposed method maintains consistent performance.

We also test the proposed 2D-to-3D quality prediction model by applying it to different base 2D-IQA approaches on both databases. Note that exactly the same 2D-to-3D quality prediction model obtained from 2DIQ-MOS and 3DIQ-MOS scores with Waterloo-IVC database Phase I is used and thus the model is completely independent of any tested objective 2D-IQA approaches including PSNR, SSIM and IDW-SSIM. The comparison

Table 3.25: Performance comparison of 2D-to-3D quality prediction models on Waterloo-IVC 3D Image Database: 2DIQ-MOS and IDW-SSIM

		Waterloo-IVC 3D Image Database Phase I					Waterloo-IVC 3D Image Database Phase II				
	2DIQ-MOS	PLCC	SRCC	KRCC	RMSE	MAE	PLCC	SRCC	KRCC	RMSE	MAE
All	Average	0.8835	0.8765	0.7161	7.3700	5.3293	0.8763	0.8820	0.7145	9.2201	6.7164
	Weighted	0.9561	0.9522	0.8162	4.6108	3.5645	0.9568	0.9477	0.8080	5.5655	4.3371
Sym.	Average	0.9801	0.9657	0.8482	3.8266	2.9160	0.9799	0.9696	0.8557	4.0525	3.1712
	Weighted	0.9801	0.9657	0.8482	3.8266	2.9160	0.9799	0.9696	0.8557	4.0525	3.1712
Asym.	Average	0.8572	0.8471	0.6780	7.9895	5.8643	0.8418	0.8501	0.6693	10.4221	7.9037
	Weighted	0.9522	0.9452	0.8026	4.7406	3.6791	0.9511	0.9424	0.7977	5.9620	4.6683
3D.1	Average	0.9801	0.9657	0.8482	3.8266	2.9160	0.9799	0.9696	0.8557	4.0525	3.1712
	Weighted	0.9801	0.9657	0.8482	3.8266	2.9160	0.9799	0.9696	0.8557	4.0525	3.1712
3D.2	Average	0.6613	0.5433	0.4406	11.6394	9.0069	0.6121	0.5874	0.4524	14.2881	11.6204
	Weighted	0.9286	0.9160	0.7556	5.7574	4.7098	0.9414	0.9497	0.8070	6.0932	4.8568
3D.3	Average	0.9666	0.9164	0.7597	3.6078	2.7578	0.9471	0.8898	0.7176	5.7909	4.4262
	Weighted	0.9714	0.9307	0.7789	3.3391	2.6247	0.9602	0.9318	0.7745	5.0392	4.0923
3D.4	Average	0.9223	0.8271	0.6390	5.9710	4.5387	0.9225	0.8798	0.7047	7.0928	5.4776
	Weighted	0.9656	0.9357	0.7822	4.0179	3.2243	0.9549	0.9320	0.7803	5.4557	4.2124
	IDW-SSIM	PLCC	SRCC	KRCC	RMSE	MAE	PLCC	SRCC	KRCC	RMSE	MAE
All	Average	0.7800	0.7572	0.5779	9.8471	7.5579	0.7790	0.7700	0.5858	12.0013	9.2888
	Weighted	0.9273	0.9156	0.7494	5.8899	4.7668	0.8896	0.8645	0.6852	8.7435	7.0266
Sym.	Average	0.9579	0.9371	0.7725	5.5328	4.5967	0.9360	0.9017	0.7242	7.1484	5.6536
	Weighted	0.9580	0.9375	0.7732	5.5320	4.5966	0.9360	0.9017	0.7242	7.1488	5.6529
Asym.	Average	0.7475	0.7111	0.5333	10.3042	7.8824	0.7480	0.7410	0.5509	12.8163	9.9539
	Weighted	0.9278	0.9112	0.7435	5.7887	4.6891	0.8903	0.8648	0.6826	8.7937	7.1850
3D.1	Average	0.9579	0.9371	0.7725	5.5328	4.5967	0.9360	0.9017	0.7242	7.1484	5.6536
	Weighted	0.9580	0.9375	0.7732	5.5320	4.5966	0.9360	0.9017	0.7242	7.1488	5.6529
3D.2	Average	0.5496	0.3465	0.2818	12.9639	10.9663	0.5406	0.4850	0.3626	15.2005	12.1859
	Weighted	0.9170	0.8969	0.7298	6.1906	5.0713	0.8729	0.8860	0.6943	8.8172	7.0825
3D.3	Average	0.9065	0.7384	0.5577	5.9426	4.7459	0.8673	0.6200	0.4520	8.9802	7.2966
	Weighted	0.9383	0.8393	0.6550	4.8674	3.9636	0.9048	0.7820	0.5798	7.6819	6.3187
3D.4	Average	0.8702	0.6984	0.5248	7.6114	5.8683	0.8578	0.7469	0.5726	9.4437	7.3285
	Weighted	0.9308	0.8657	0.6804	5.6461	4.5552	0.9021	0.8343	0.6610	7.9272	6.3546

results with the direct averaging method are shown in Table 3.26 to Table 3.28, where it can be seen that the proposed method significantly improves most base 2D-IQA methods. The only exception is PSNR, which might be due to its poor performance in 2D image quality assessment, and thus merely changing 2D to 3D prediction method would not lead to any meaningful result. The scatter plots of 3DIQ-MOS scores versus predictions by averaging IDW-SSIM and weighting IDW-SSIM are shown in Figure 3.23 to Figure 3.27, the (a) and (b) are the corresponding scatter plots for Waterloo-IVC 3D image database Phase I while the (c) and (d) are the scatter plots for Waterloo-IVC 3D image database Phase II. From Table 3.26 to Table 3.28 and Figure 3.23 to Figure 3.27, it can be observed that the proposed 2D-to-3D model produces the most significant performance improvement from symmetric to asymmetric distortions in the case of using IDW-SSIM as the base 2D-IQA approach.

Table 3.26: Performance comparison of 2D-to-3D prediction models on Waterloo-IVC 3D Image Database: All Images

	Waterloo-IVC 3D Image Database Phase I					Waterloo-IVC 3D Image Database Phase II				
2D-IQA	PLCC	SRCC	KRCC	RMSE	MAE	PLCC	SRCC	KRCC	RMSE	MAE
Average 2DIQ-MOS	0.8835	0.8765	0.7161	7.3700	5.3293	0.8763	0.8820	0.7145	9.2201	6.7164
Weighted 2DIQ-MOS	0.9561	0.9522	0.8162	4.6108	3.5645	0.9568	0.9477	0.8080	5.5655	4.3371
Average PSNR	0.6914	0.5209	0.3775	11.3689	9.1715	0.6304	0.4964	0.3521	14.8577	12.2310
Weighted PSNR	0.7419	0.5192	0.3830	10.5514	8.7829	0.7186	0.4948	0.3605	13.3107	11.2477
Average SSIM	0.6771	0.5963	0.4354	11.5794	9.3504	0.5197	0.4684	0.3316	16.3519	13.3223
Weighted SSIM	0.8035	0.6712	0.5047	9.3675	7.6946	0.7257	0.5640	0.4090	13.1691	11.2493
Average MS-SSIM	0.6629	0.5034	0.3614	11.7821	9.6847	0.5178	0.4182	0.2948	16.3736	13.6070
Weighted MS-SSIM	0.7723	0.5380	0.4128	9.9964	7.8308	0.7066	0.4728	0.3412	13.5430	11.4356
Average IW-SSIM	0.6595	0.5900	0.4433	11.8286	9.4377	0.5311	0.4845	0.3471	16.2168	12.9443
Weighted IW-SSIM	0.8025	0.6841	0.5245	9.3890	7.8411	0.7456	0.5906	0.4354	12.7550	10.7761
Average IDW-SSIM	0.7800	0.7572	0.5779	9.8471	7.5579	0.7790	0.7700	0.5858	12.0013	9.2888
Weighted IDW-SSIM	0.9273	0.9156	0.7494	5.8899	4.7668	0.8896	0.8645	0.6852	8.7435	7.0266
Benoit [10]	0.6812	0.5961	0.4293	11.5208	9.0513	0.5344	0.4492	0.3138	16.1770	12.8308
Chen [25]	0.7327	0.6831	0.5223	10.7095	7.7474	0.6130	0.5781	0.4165	15.1222	11.7693
Yang [164]	0.7061	0.6107	0.4429	11.1428	9.1004	0.6742	0.5873	0.4181	14.1355	11.5621
You [168]	0.6720	0.5969	0.4352	11.6526	9.3838	0.6383	0.5874	0.4136	14.7335	12.1816

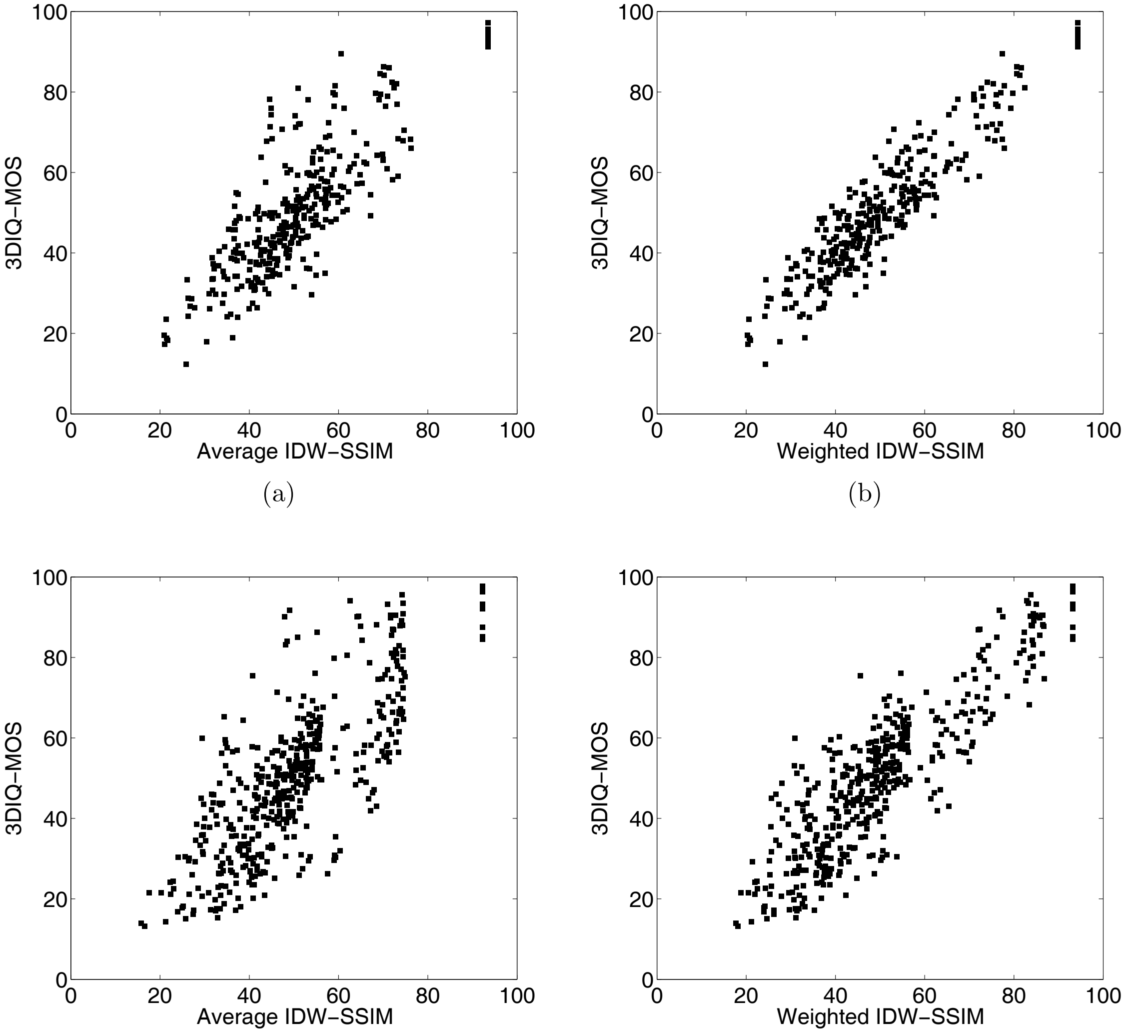
We have also compared the proposed method with state-of-the-art 3D-IQA approaches [23, 10, 164, 168], and [25] using both databases, and the results are shown in Table 3.29 and Table 3.30, respectively. The proposed method achieves the best performance in both databases among all objective IQA methods. The highly competitive performance in the Waterloo-IVC 3D Image Database Phase II and LIVE 3D Image Database Phase II is a more convincing result because no parameter has been determined using the Waterloo-IVC

Table 3.27: Performance comparison of 2D-to-3D prediction models on Waterloo-IVC 3D Image Database Phase I: Symmetric vs. Asymmetric

	PLCC		SRCC		KRCC		RMSE		MAE	
2D-IQA	Sym.	Asym.	Sym.	Asym.	Sym.	Asym.	Sym.	Asym.	Sym.	Asym.
Average 2DIQ-MOS	0.9801	0.8572	0.9657	0.8471	0.8482	0.6780	3.8266	7.9895	2.9160	5.8643
Weighted 2DIQ-MOS	0.9801	0.9522	0.9657	0.9452	0.8482	0.8026	3.8266	4.7406	2.9160	3.6791
Average PSNR	0.7877	0.7223	0.4565	0.5561	0.3492	0.4037	11.8784	10.7284	9.4310	8.5327
Weighted PSNR	0.7877	0.7876	0.4572	0.5593	0.3505	0.4093	11.8786	9.5591	9.4311	7.9879
Average SSIM	0.8611	0.6843	0.7517	0.5694	0.5615	0.4171	9.8040	11.3126	7.7319	8.9344
Weighted SSIM	0.8611	0.8338	0.7517	0.6631	0.5615	0.4985	9.8041	8.5642	7.7317	7.0285
Average MS-SSIM	0.8245	0.6834	0.6048	0.5043	0.4580	0.3608	10.9108	11.3254	8.1563	9.2084
Weighted MS-SSIM	0.8245	0.8082	0.6048	0.5503	0.4580	0.4182	10.9107	9.1361	8.1557	7.2159
Average IW-SSIM	0.8611	0.6508	0.7582	0.5501	0.5729	0.4153	9.8034	11.7784	8.0576	9.1672
Weighted IW-SSIM	0.8611	0.8278	0.7584	0.6688	0.5735	0.5119	9.8032	8.7042	8.0573	7.2423
Average IDW-SSIM	0.9579	0.7475	0.9371	0.7111	0.7725	0.5333	8.0576	10.3042	4.5967	7.8824
Weighted IDW-SSIM	0.9580	0.9278	0.9375	0.9112	0.7732	0.7435	8.0573	5.7887	4.5966	4.6891
Benoit [10]	0.8563	0.6986	0.7427	0.5880	0.5622	0.4218	9.9599	11.0993	8.6247	8.5279
Chen [25]	0.9521	0.7306	0.9250	0.6441	0.7558	0.4860	5.8938	10.5920	4.9568	7.6630
Yang [164]	0.8347	0.7149	0.6670	0.6106	0.4827	0.4434	10.6175	10.8465	8.8134	8.6119
You [168]	0.8572	0.6818	0.7517	0.5706	0.5615	0.4175	9.9290	11.3482	7.7502	8.9733

Table 3.28: Performance comparison of 2D-to-3D prediction models on Waterloo-IVC 3D Image Database Phase II: Symmetric vs. Asymmetric

	PLCC		SRCC		KRCC		RMSE		MAE	
2D-IQA	Sym.	Asym.	Sym.	Asym.	Sym.	Asym.	Sym.	Asym.	Sym.	Asym.
Average 2DIQ-MOS	0.9799	0.8418	0.9696	0.8501	0.8557	0.6693	4.0525	10.4221	3.1712	7.9037
Weighted 2DIQ-MOS	0.9799	0.9511	0.9696	0.9424	0.8557	0.7977	4.0525	5.9620	3.1712	4.6683
Average PSNR	0.6878	0.6576	0.3430	0.5573	0.2593	0.3976	14.7395	14.5480	12.5625	11.6053
Weighted PSNR	0.6878	0.7669	0.3429	0.5746	0.2597	0.4192	14.7393	12.3924	12.5623	10.3401
Average SSIM	0.7463	0.5273	0.5933	0.4541	0.4250	0.3212	13.5161	16.4072	11.6680	13.0249
Weighted SSIM	0.7463	0.7668	0.5933	0.5822	0.4249	0.4208	13.5154	12.3952	11.6674	10.3845
Average MS-SSIM	0.7165	0.5412	0.4456	0.4475	0.3190	0.3192	14.1639	16.2377	12.0754	13.2690
Weighted MS-SSIM	0.7165	0.7527	0.4456	0.5236	0.3191	0.3764	14.1633	12.7136	12.0748	10.5823
Average IW-SSIM	0.7680	0.5208	0.6345	0.4521	0.4668	0.3195	13.0045	16.4850	11.0739	12.8515
Weighted IW-SSIM	0.7680	0.7849	0.6348	0.5941	0.4669	0.4326	13.0036	11.9658	11.0728	9.8922
Average IDW-SSIM	0.9360	0.7480	0.9017	0.7410	0.7242	0.5509	7.1484	12.8163	5.6536	9.9539
Weighted IDW-SSIM	0.9360	0.8903	0.9017	0.8648	0.7242	0.6826	7.1488	8.7937	5.6529	7.1850
Benoit [10]	0.7488	0.5480	0.5589	0.4466	0.3925	0.3136	13.4575	16.1520	11.2175	12.4789
Chen [25]	0.8202	0.6317	0.7581	0.5627	0.5635	0.4055	11.6165	14.9701	9.6565	11.6210
Yang [164]	0.7495	0.6850	0.5602	0.5997	0.4006	0.4225	13.4422	14.0686	11.0352	11.3139
You [168]	0.7917	0.6404	0.6631	0.5944	0.4850	0.4154	12.4031	14.8307	10.8100	11.9458



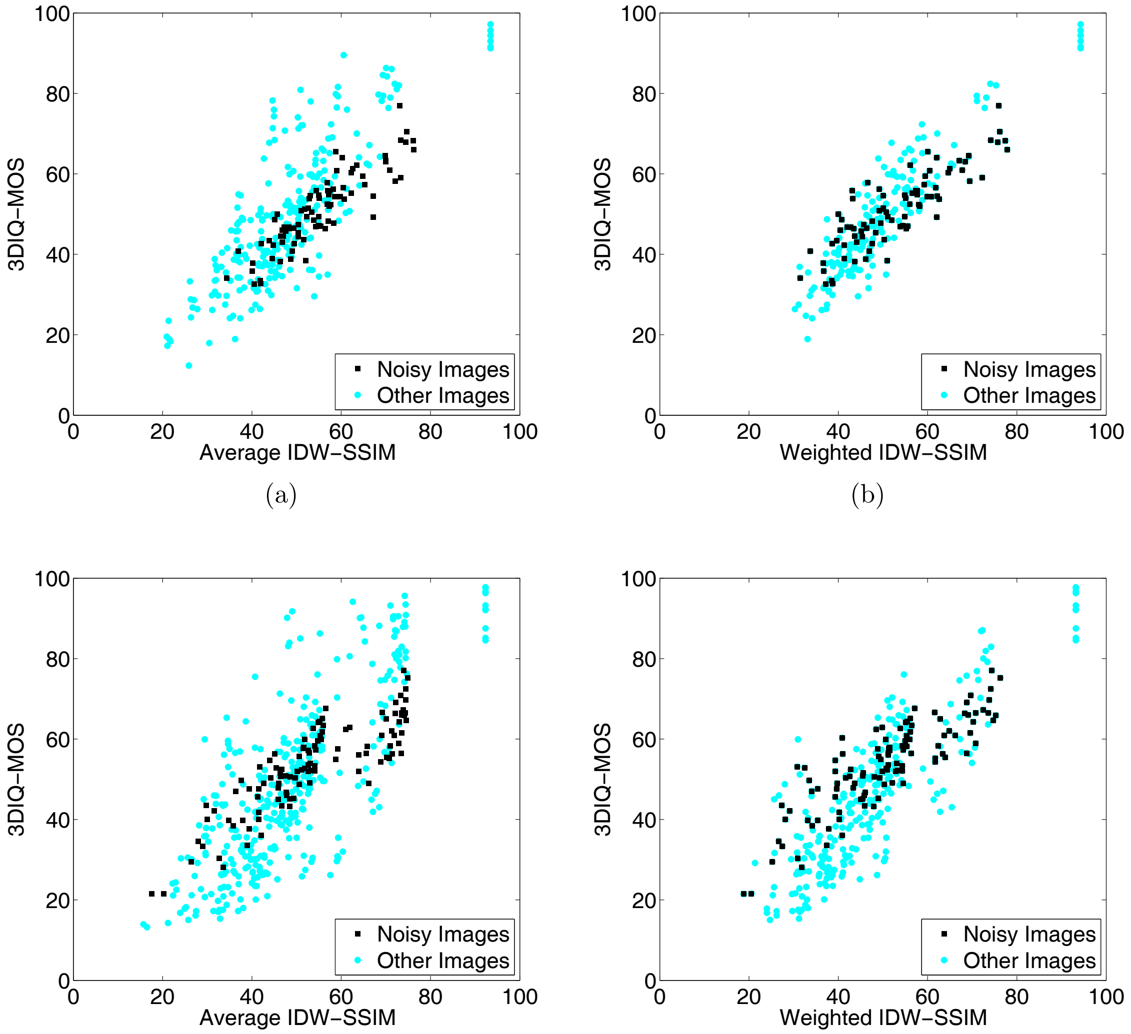
(a)

(b)

(c)

(d)

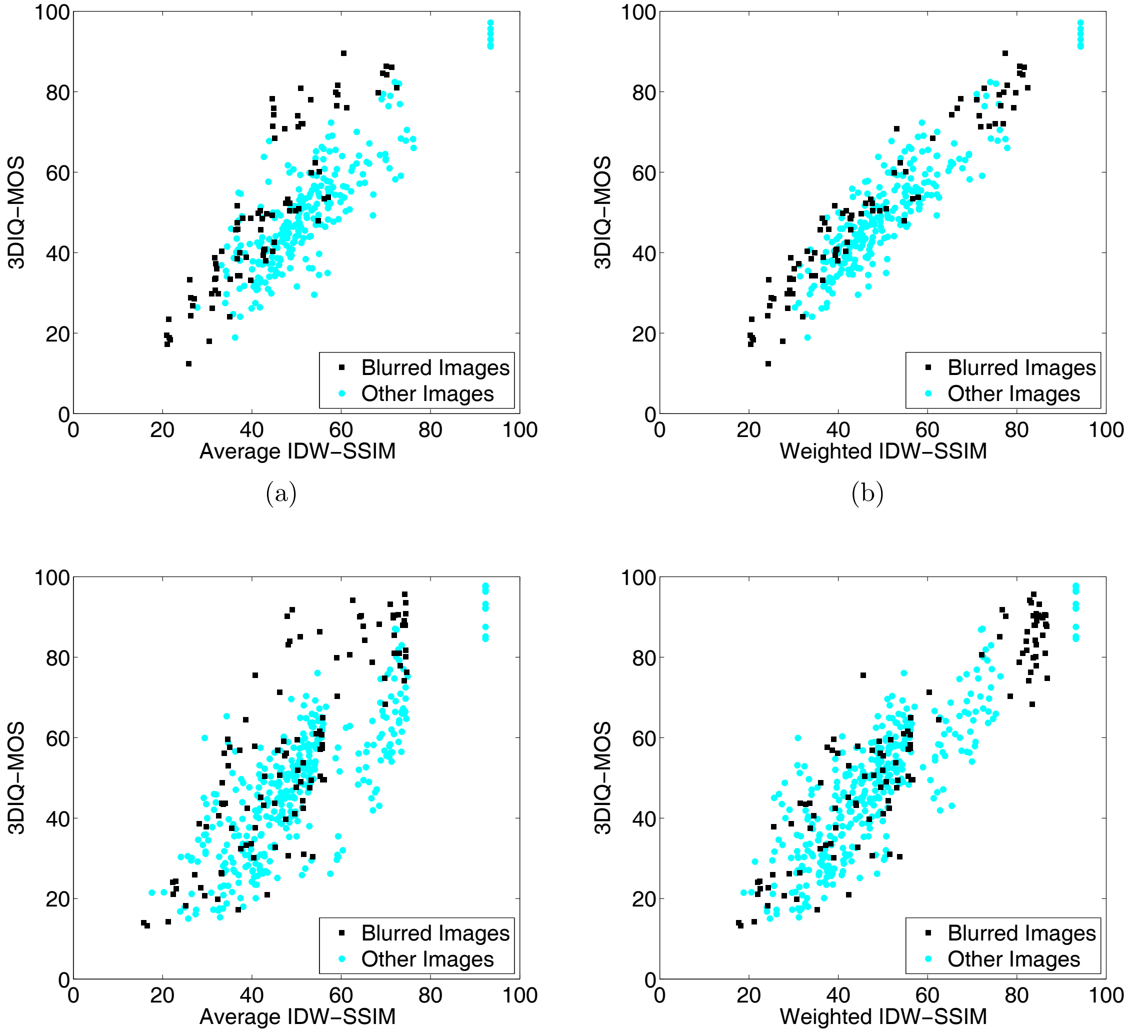
Figure 3.23: 3DIQ-MOS versus predictions from IDW-SSIM of 2D left- and right-views. (a) All Images, Average IDW-SSIM on Waterloo-IVC Phase I. (b) All Images, Weighted IDW-SSIM on Waterloo-IVC Phase I. (c) All Images, Average IDW-SSIM on Waterloo-IVC Phase II. (d) All Images, Weighted IDW-SSIM on Waterloo-IVC Phase II.



(a) (b)

(c) (d)

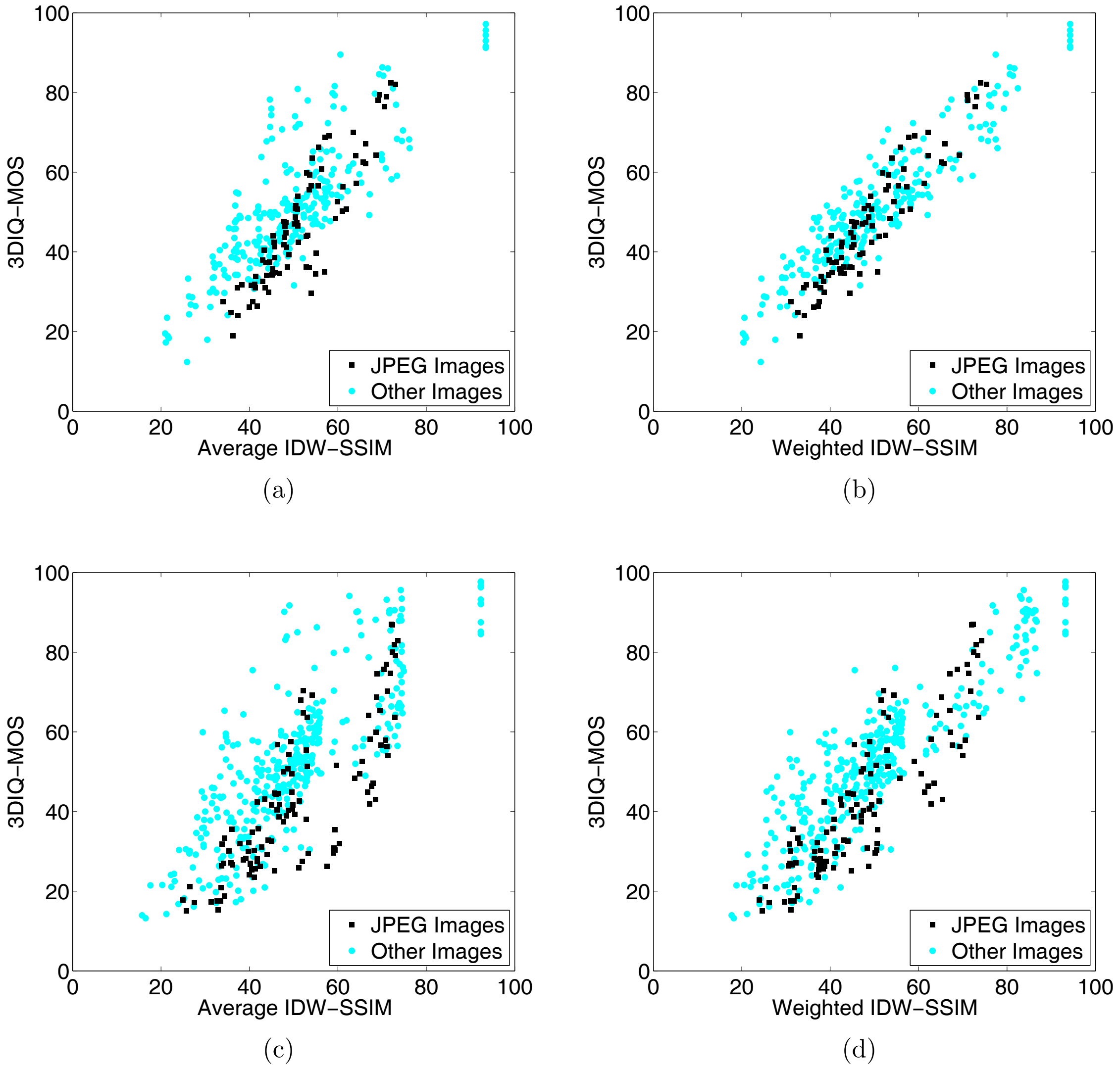
Figure 3.24: 3DIQ-MOS versus predictions from IDW-SSIM of 2D left- and right-views. (a) Noisy Images, Average IDW-SSIM on Waterloo-IVC Phase I. (b) Noisy Images, Weighted IDW-SSIM on Waterloo-IVC Phase I. (c) Noisy Images, Average IDW-SSIM on Waterloo-IVC Phase II. (d) Noisy Images, Weighted IDW-SSIM on Waterloo-IVC Phase II.



(a) (b)

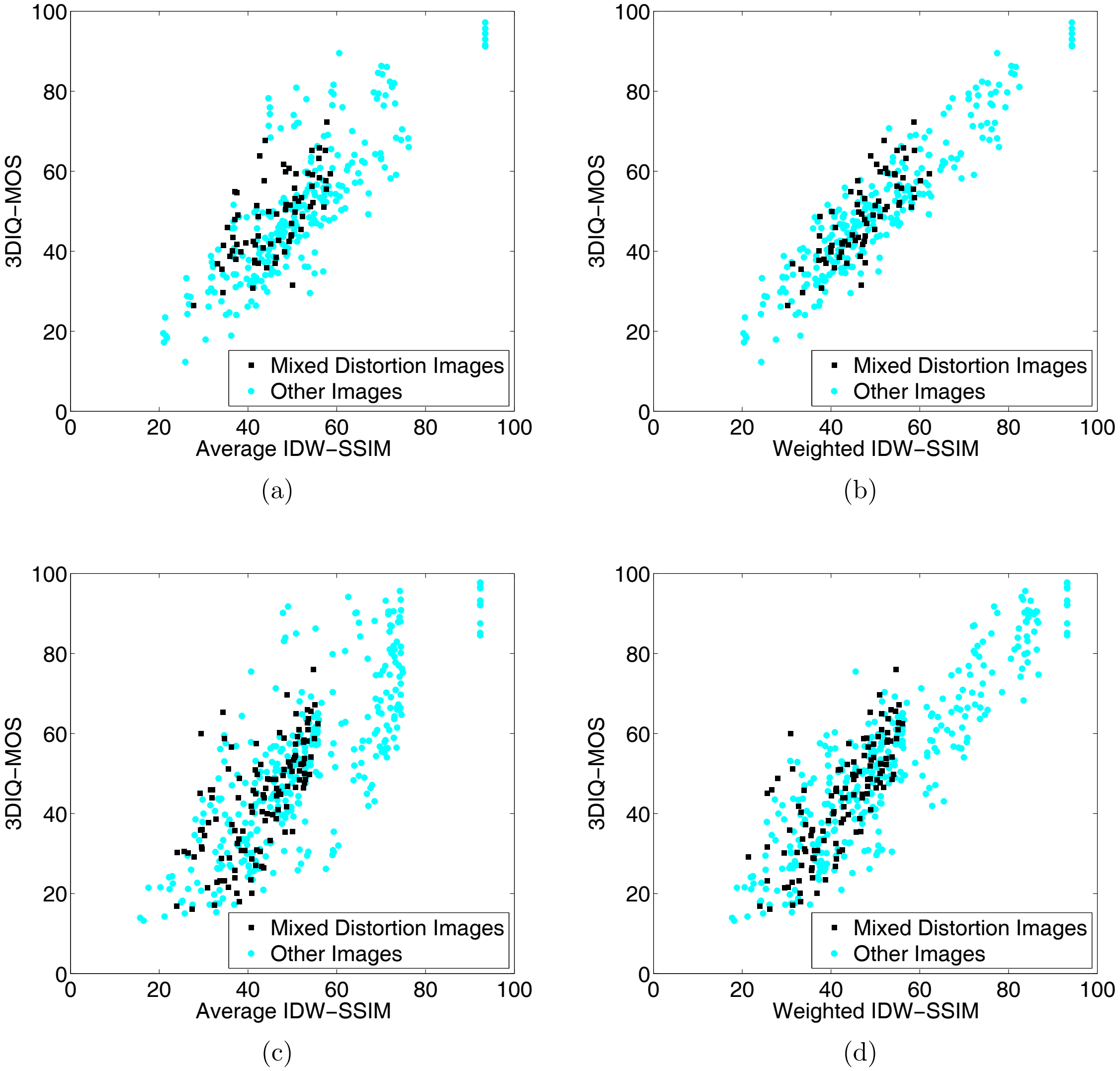
(c) (d)

Figure 3.25: 3DIQ-MOS versus predictions from IDW-SSIM of 2D left- and right-views. (a) Blurred Images, Average IDW-SSIM on Waterloo-IVC Phase I. (b) Blurred Images, Weighted IDW-SSIM on Waterloo-IVC Phase I. (c) Blurred Images, Average IDW-SSIM on Waterloo-IVC Phase II. (d) Blurred Images, Weighted IDW-SSIM on Waterloo-IVC Phase II.



(a) (b) (c) (d)

Figure 3.26: 3DIQ-MOS versus predictions from IDW-SSIM of 2D left- and right-views. (a) JPEG Images, Average IDW-SSIM on Waterloo-IVC Phase I. (b) JPEG Images, Weighted IDW-SSIM on Waterloo-IVC Phase I. (c) JPEG Images, Average IDW-SSIM on Waterloo-IVC Phase II. (d) JPEG Images, Weighted IDW-SSIM on Waterloo-IVC Phase II.



(a) (b) (c) (d)

Figure 3.27: 3DIQ-MOS versus predictions from IDW-SSIM of 2D left- and right-views. (a) Mixed Distortion, Average IDW-SSIM on Waterloo-IVC Phase I. (b) Mixed Distortion, Weighted IDW-SSIM on Waterloo-IVC Phase I. (c) Mixed Distortion, Average IDW-SSIM on Waterloo-IVC Phase II. (d) Mixed Distortion, Weighted IDW-SSIM on Waterloo-IVC Phase II.

Phase II and the LIVE database. Another important observation is that there is a large performance drop in all other objective methods from symmetric to asymmetric distortions, whereas the drop is much smaller in the proposed method.

Table 3.29: Performance comparison of 2D-to-3D prediction models on LIVE 3D Image Database

	LIVE 3D Image Database Phase I					LIVE 3D Image Database Phase II				
2D-IQA	PLCC	SRCC	KRCC	RMSE	MAE	PLCC	SRCC	KRCC	RMSE	MAE
Average PSNR	0.8370	0.8341	0.6298	8.9738	7.0181	0.7546	0.7303	0.5372	7.4068	5.8369
Weighted PSNR	0.8352	0.8328	0.6288	9.0185	7.0448	0.7764	0.7528	0.5531	7.1132	5.6647
Average SSIM	0.8766	0.8765	0.6791	7.8913	6.0833	0.8024	0.7925	0.6016	6.7366	5.1798
Weighted SSIM	0.8742	0.8740	0.6761	7.9634	6.1388	0.8433	0.8408	0.6444	6.0671	4.7670
Average MS-SSIM	0.9301	0.9239	0.7487	6.0227	4.7112	0.7947	0.7774	0.6067	6.8509	5.1541
Weighted MS-SSIM	0.9317	0.9250	0.7498	5.9550	4.6816	0.9079	0.9096	0.7342	4.7322	3.6490
Average IW-SSIM	0.9417	0.9335	0.7667	5.5190	4.3693	0.7842	0.7491	0.5920	7.0040	5.1988
Weighted IW-SSIM	0.9450	0.9359	0.7709	5.3617	4.2582	0.9123	0.9132	0.7417	4.6225	3.5126
Average IDW-SSIM	0.9281	0.9237	0.7471	6.1053	4.7562	0.8177	0.7987	0.6191	6.4979	4.9159
Weighted IDW-SSIM	0.9295	0.9244	0.7485	6.0477	4.7298	0.9152	0.9178	0.7425	4.5493	3.5763
Benoit [10]	0.8836	0.8835	0.6861	7.6793	6.0346	0.7554	0.7409	0.5674	7.3967	5.6058
Chen [25]	0.9241	0.9157	0.7368	6.2681	4.8262	0.9073	0.9013	0.7307	4.7464	3.5824
Yang [164]	0.8131	0.8143	0.6099	9.5452	7.4921	0.7344	0.7210	0.5327	7.6604	6.0953
You [168]	0.8677	0.8765	0.6792	8.1517	6.2893	0.7974	0.7924	0.6015	6.8108	5.2472

Table 3.30: Performance comparison of 2D-to-3D prediction models on LIVE 3D Image Database Phase II: Symmetric vs. Asymmetric

	PLCC		SRCC		KRCC		RMSE		MAE	
2D-IQA	Sym.	Asym.	Sym.	Asym.	Sym.	Asym.	Sym.	Asym.	Sym.	Asym.
Average PSNR	0.7839	0.7237	0.7764	0.6974	0.5814	0.5106	7.7512	6.9944	5.9384	5.5623
Weighted PSNR	0.7841	0.7654	0.7770	0.7149	0.5823	0.5223	7.7485	6.5228	5.9352	5.2701
Average SSIM	0.8517	0.7665	0.8256	0.7360	0.6327	0.5516	6.5430	6.5098	5.0729	5.0304
Weighted SSIM	0.8519	0.8394	0.8256	0.8318	0.6324	0.6364	6.5394	5.5082	5.0723	4.2911
Average MS-SSIM	0.9266	0.7187	0.9121	0.6842	0.7425	0.5219	4.6939	7.0471	3.6498	5.4580
Weighted MS-SSIM	0.9271	0.8912	0.9124	0.8950	0.7430	0.7156	4.6809	4.5970	3.6387	3.5482
Average IW-SSIM	0.9458	0.6986	0.9314	0.6397	0.7694	0.4951	4.0528	7.2515	3.1837	5.6560
Weighted IW-SSIM	0.9463	0.8852	0.9324	0.8876	0.7705	0.7091	4.0357	4.7151	3.1755	3.6293
Average IDW-SSIM	0.9363	0.7383	0.9227	0.6902	0.7545	0.5216	4.3849	6.8363	3.4405	5.2556
Weighted IDW-SSIM	0.9367	0.8974	0.9231	0.9011	0.7559	0.7225	4.3706	4.4720	3.4311	3.5202
Benoit [10]	0.8812	0.7026	0.8547	0.6747	0.6593	0.5124	5.9011	7.2116	4.7997	5.6079
Chen [25]	0.9384	0.8753	0.9252	0.8538	0.7599	0.6783	4.3129	4.9020	3.3895	3.6553
Yang [164]	0.7481	0.7029	0.7282	0.6787	0.5386	0.4942	8.2854	7.2087	6.4751	5.7345
You [168]	0.8411	0.7556	0.8256	0.7358	0.6327	0.5513	6.7537	6.6383	5.4177	5.0773

Moreover, we have also tested the proposed method with state-of-the-art 3D-IQA approaches on Ningbo 3D Image Database Phase I (Asymmetric distortion) and Phase II (Symmetric distortion), IVC 3D Image Database (Symmetric distortion) and MCL-3D Image Database (Symmetric distortion), and the results are shown in Table 3.31, Table 3.32 and Table 3.33, respectively.

Table 3.31: Performance comparison of 2D-to-3D prediction models on Ningbo 3D Image Database

	Symmetric					Asymmetric				
2D-IQA	PLCC	SRCC	KRCC	RMSE	MAE	PLCC	SRCC	KRCC	RMSE	MAE
Average PSNR	0.8867	0.9040	0.7189	7.9429	5.7697	0.7982	0.8364	0.6288	7.5908	5.1413
Weighted PSNR	0.8864	0.9033	0.7181	7.9542	5.7861	0.7439	0.5463	0.3739	8.4211	6.4544
Average SSIM	0.8887	0.9013	0.7203	7.8781	5.6587	0.8032	0.8431	0.6338	7.5073	5.2828
Weighted SSIM	0.8886	0.9013	0.7204	7.8794	5.6614	0.9345	0.8982	0.7227	4.4856	3.1651
Average MS-SSIM	0.9216	0.9241	0.7475	6.6672	4.9693	0.8373	0.8676	0.6645	6.8896	4.5837
Weighted MS-SSIM	0.9216	0.9240	0.7472	6.6690	4.9717	0.9034	0.8794	0.6887	5.4041	3.8208
Average IW-SSIM	0.9432	0.9344	0.7692	5.7060	4.2819	0.7856	0.8240	0.6114	7.7960	5.4199
Weighted IW-SSIM	0.9432	0.9344	0.7691	5.7073	4.2851	0.9375	0.9002	0.7243	4.3863	3.1087
Average IDW-SSIM	0.9379	0.9305	0.7630	5.9591	4.5166	0.6999	0.7726	0.5503	9.0001	6.8424
Weighted IDW-SSIM	0.9379	0.9306	0.7631	5.9583	4.5173	0.8704	0.8559	0.6559	6.2040	4.6081
Benoit [10]	0.7298	0.7253	0.5275	11.7456	8.8430	0.7542	0.7891	0.5682	8.2745	6.1042
Chen [25]	0.9109	0.9077	0.7208	7.0879	5.3972	0.8600	0.8377	0.6266	6.4303	4.7547
Yang [164]	0.8835	0.9044	0.7179	8.0486	5.8571	0.7815	0.8276	0.6163	7.8621	5.3697
You [168]	0.8884	0.9010	0.7196	7.8869	5.6766	0.8042	0.8432	0.6338	7.4891	5.2717

3.7 Objective Study II: Blind 2D-to-3D Quality Prediction

3.7.1 Introduction

NR or blind image quality assessment (BIQA) predict perceived quality of a test image without referring to an original image that is assumed to have pristine quality [152]. BIQA is highly challenging not only because of the difficulty in accurately estimating human behaviors in evaluating image quality across different visual content, distortion types and distortion levels, but also because real-world applications such as online quality monitoring often require the image and video streams to be evaluated at high speed, ideally in real-time. Therefore, 3D-BIQA is an even more challenging problem, especially when the distortions in the left- and right-views are asymmetric.

Table 3.34 to Table 3.38 report PLCC, SRCC, KRCC, RMSE and MAE between 3DIQ-MOS and averaging some state-of-the-art 2D-BIQA estimations of both views on

Table 3.32: Performance comparison of 2D-to-3D prediction models on IVC 3D Image Database

2D-IQA	PLCC	SRCC	KRCC	RMSE	MAE
Average PSNR	0.5936	0.5552	0.3978	17.7548	14.5179
Weighted PSNR	0.5936	0.5552	0.3973	17.7550	14.5219
Average SSIM	0.7674	0.6899	0.5416	14.1455	10.4948
Weighted SSIM	0.7674	0.6899	0.5416	14.1452	10.4945
Average MS-SSIM	0.7701	0.6858	0.5416	14.0746	10.3848
Weighted MS-SSIM	0.7701	0.6859	0.5421	14.0737	10.3842
Average IW-SSIM	0.8096	0.6979	0.5730	12.9488	9.0842
Weighted IW-SSIM	0.8097	0.6979	0.5730	12.9475	9.0831
Average IDW-SSIM	0.8120	0.7171	0.5935	12.8756	8.9459
Weighted IDW-SSIM	0.8120	0.7174	0.5940	12.8759	8.9471
Benoit [10]	0.5051	0.4940	0.3603	19.0406	15.6000
Chen [25]	0.6152	0.6043	0.4482	17.3939	13.7941
Yang [164]	0.5976	0.5432	0.3928	17.6896	14.2696
You [168]	0.7669	0.6902	0.5421	14.1590	10.5039

Table 3.33: Performance comparison of 2D-to-3D prediction models on MCL-3D Image Database

2D-IQA	PLCC	SRCC	KRCC	RMSE	MAE
Average PSNR	0.8320	0.8405	0.6406	1.4435	1.1383
Weighted PSNR	0.8323	0.8407	0.6409	1.4420	1.1371
Average SSIM	0.8935	0.9034	0.7201	1.1682	0.9214
Weighted SSIM	0.8938	0.9036	0.7203	1.1670	0.9206
Average MS-SSIM	0.8644	0.8750	0.6837	1.3081	1.0197
Weighted MS-SSIM	0.8647	0.8752	0.6839	1.3068	1.0188
Average IW-SSIM	0.9197	0.9260	0.7564	1.0213	0.8079
Weighted IW-SSIM	0.9198	0.9261	0.7564	1.0206	0.8071
Average IDW-SSIM	0.8682	0.8792	0.6907	1.2910	1.0314
Weighted IDW-SSIM	0.8681	0.8792	0.6907	1.2913	1.0307
Benoit [10]	0.6330	0.6359	0.4562	2.0142	1.6316
Chen [25]	0.8528	0.8587	0.6627	1.3586	1.0726
Yang [164]	0.8346	0.8445	0.6455	1.4331	1.1255
You [168]	0.8922	0.9037	0.7205	1.1750	0.9271

LIVE 3D Image Database Phase II, Waterloo-IVC 3D Image Database Phase I and Phase II. The ten tested state-of-the-art 2D-BIQA metrics include Blind Image Quality Index (BIQI) [88], BLind Image Integrity Notator using DCT-Statistics II (BLIINDS-II) [109], Blind/Referenceless Image Spatial QUality Evaluator (BRISQUE) [86], Codebook Representation for No-Reference Image Assessment (CORNIA) [167], Distortion Identification-based Image Verity and INtegrity Evalutation (DIIVINE) [89], Local Pattern Statistics Index (LPSI) [160], M_3 [161], Naturalness Image Quality Evaluator (NIQE) [87], Quality-Aware Clustering (QAC) [162] and Distortion Type Classification and Label Transfer (TCLT) [159]. Among them, BIQI, BLIINDS-II, BRISQUE, CORNIA, DIIVINE, M_3 TCLT are opinion-aware (OA) BIQA methods that require subject-rated images for training, and are trained using all images from LIVE Image Quality Assessment Database Release 2 [126]. LPSI, NIQE and QACS are opinion-free (OF) BIQA methods, and are tested directly with their default parameters.

Table 3.34: Performance comparison of 2D-BIQA models on Waterloo-IVC 3D Image Database

	Waterloo-IVC 3D Image Database Phase I					Waterloo-IVC 3D Image Database Phase II				
2D-IQA/BIQA	PLCC	SRCC	KRCC	RMSE	MAE	PLCC	SRCC	KRCC	RMSE	MAE
Average 2DIQ-MOS	0.8835	0.8765	0.7161	7.3700	5.3293	0.8763	0.8820	0.7145	9.2201	6.7164
Weighted 2DIQ-MOS	0.9509	0.9413	0.8045	4.8718	3.6164	0.9507	0.9382	0.8012	5.9343	4.5083
Average IDW-SSIM	0.7800	0.7572	0.5779	9.8471	7.5579	0.7790	0.7700	0.5858	12.0013	9.2888
Weighted IDW-SSIM	0.9259	0.9071	0.7372	5.9437	4.7332	0.8864	0.8637	0.6860	8.8609	7.1558
Average BIQI	0.7824	0.7708	0.5752	9.8001	7.4893	0.6129	0.5861	0.4192	15.1227	11.9554
Weighted BIQI	0.8664	0.8432	0.6586	7.8584	6.1899	0.7409	0.6756	0.5054	12.8539	10.4482
Average BLIINDS-II	0.7314	0.6494	0.4773	10.7309	8.3156	0.7425	0.7031	0.5202	12.8207	10.2602
Weighted BLIINDS-II	0.7705	0.6614	0.4926	10.0308	7.5943	0.7812	0.7213	0.5406	11.9483	9.5227
Average BRISQUE	0.8216	0.8062	0.6331	8.9700	6.9088	0.8135	0.8184	0.6410	11.1299	8.6111
Weighted BRISQUE	0.8856	0.8459	0.6846	7.3088	5.7013	0.8809	0.8641	0.6973	9.0570	7.0220
Average CORNIA	0.8295	0.8216	0.6384	8.7894	6.8598	0.8223	0.8200	0.6364	10.8927	8.4097
Weighted CORNIA	0.8741	0.8446	0.6638	7.6428	6.0783	0.8844	0.8636	0.6882	8.9323	7.0463
Average DIIVINE	0.4623	0.3663	0.2736	13.9531	10.7008	0.4699	0.4424	0.3211	16.8952	13.0883
Weighted DIIVINE	0.5353	0.4516	0.3433	13.3895	10.3664	0.5848	0.5214	0.3856	15.5261	12.1795
Average LPSI	0.5777	0.5882	0.4236	12.8446	10.0398	0.6029	0.6295	0.4465	15.2703	12.0742
Weighted LPSI	0.6156	0.6395	0.4606	12.4005	9.8048	0.7015	0.7194	0.5204	13.6410	11.4144
Average M_3	0.8437	0.8326	0.6531	8.4461	6.4844	0.7934	0.7834	0.5970	11.6509	9.0754
Weighted M_3	0.9122	0.8968	0.7317	6.4482	5.0033	0.8692	0.8416	0.6643	9.4625	7.5488
Average NIQE	0.7635	0.6461	0.4714	10.1628	8.3894	0.6278	0.5461	0.3851	14.8974	12.3827
Weighted NIQE	0.7975	0.6429	0.4806	9.4946	7.9265	0.7300	0.5677	0.4108	13.0805	10.9013
Average QACS	0.7360	0.6241	0.4460	10.6522	8.6039	0.6995	0.6146	0.4375	13.6778	11.4087
Weighted QACS	0.7909	0.6203	0.4510	9.6292	7.8129	0.7677	0.6225	0.4519	12.2637	10.2884
Average TCLT	0.6809	0.6425	0.4733	11.5245	9.2046	0.6262	0.6032	0.4348	14.9220	11.9369
Weighted TCLT	0.8236	0.7459	0.5782	8.9248	7.0218	0.7525	0.7036	0.5360	12.6055	10.1367

Table 3.35: Performance comparison of 2D-BIQA models on LIVE 3D Image Database

	LIVE 3D Image Database Phase I					LIVE 3D Image Database Phase II				
2D-IQA/BIQA	PLCC	SRCC	KRCC	RMSE	MAE	PLCC	SRCC	KRCC	RMSE	MAE
Average IDW-SSIM	0.9281	0.9237	0.7471	6.1053	4.7562	0.8177	0.7987	0.6191	6.4979	4.9159
Weighted IDW-SSIM	0.9263	0.9216	0.7443	6.1800	4.7894	0.8984	0.8982	0.7166	4.9582	3.8889
Average BIQI	0.8715	0.8652	0.6764	8.0415	6.1198	0.7541	0.7304	0.5378	7.4128	5.9455
Weighted BIQI	0.8480	0.8354	0.6480	8.6901	6.4033	0.7962	0.7721	0.5771	6.8289	5.3989
Average BLIINDS-II	0.7754	0.7749	0.5692	10.3554	8.1994	0.6359	0.6213	0.4371	8.7113	7.2177
Weighted BLIINDS-II	0.7657	0.7633	0.5564	10.5476	8.2967	0.6610	0.6549	0.4625	8.4703	6.9385
Average BRISQUE	0.8879	0.9010	0.7144	7.5440	6.0948	0.7890	0.7698	0.5698	6.9345	5.5640
Weighted BRISQUE	0.8819	0.8959	0.7064	7.7311	6.2119	0.8062	0.8002	0.5989	6.6780	5.2926
Average CORNIA	0.8636	0.8827	0.6862	8.2669	6.6521	0.8002	0.7766	0.5815	6.7693	5.3646
Weighted CORNIA	0.8559	0.8754	0.6754	8.7508	7.0928	0.8249	0.8054	0.6089	6.3801	5.0358
Average DIIVINE	0.8337	0.8577	0.6544	9.0556	7.2598	0.6891	0.6594	0.4748	8.1793	6.6609
Weighted DIIVINE	0.8310	0.8539	0.6479	9.1222	7.3540	0.7406	0.7011	0.5105	7.5847	6.0300
Average LPSI	0.8191	0.8058	0.5877	9.4057	7.3925	0.5997	0.5532	0.4053	9.0327	7.3449
Weighted LPSI	0.8207	0.8055	0.5878	9.3684	7.4189	0.6914	0.6202	0.4633	8.1547	6.5399
Average M_3	0.8814	0.8894	0.7018	7.7443	6.0212	0.7782	0.7831	0.5865	7.0883	5.6512
Weighted M_3	0.8756	0.8834	0.6931	7.9195	6.1613	0.8058	0.8116	0.6153	6.6844	5.2443
Average NIQE	0.8624	0.8544	0.6535	8.3005	6.5169	0.7681	0.7553	0.5549	7.2270	5.8560
Weighted NIQE	0.8557	0.8475	0.6468	8.4841	6.5726	0.7900	0.7810	0.5819	6.9202	5.4231
Average QACS	0.8917	0.8966	0.7069	7.4207	5.7564	0.8054	0.7858	0.5896	6.6898	5.2544
Weighted QACS	0.8851	0.8898	0.7005	7.6320	5.8229	0.8522	0.8376	0.6437	5.9061	4.6317
Average TCLT	0.8577	0.8577	0.6590	8.4310	6.5665	0.7371	0.7292	0.5384	7.6278	6.1020
Weighted TCLT	0.8491	0.8490	0.6489	8.6611	6.6977	0.7355	0.7386	0.5429	7.6471	6.2115

It can be observed from Table 3.34 to Table 3.38 that for most of the tested 2D-BIQA methods, simply averaging 2D-BIQA measures of both views provides reasonably accurate image quality predictions of symmetrically distorted stereoscopic images but there is a significant drop in the performance for asymmetrically distorted stereoscopic images on all tested 3D databases, which is consistent with the trend we have observed with the FR 2D-IQA methods.

A more straightforward way to examine the relationship between the perceptual quality of stereoscopic images and that of its single-view images is to perform subjective test on both 2D and 3D images. In Section 3.5.2, it was found that for symmetrically distorted stereoscopic images, directly averaging the 2DIQ-MOS of both views provides excellent 3D image quality predictions, while for asymmetrically distorted stereoscopic images, a similar performance drop as of those in objective methods is observed. The performance drop is largely due to the significant prediction bias that could lean towards opposite directions (either overestimate or underestimate image quality), depending on the distortion types and levels. In Section 3.6, a binocular rivalry inspired multi-scale model to predict the

Table 3.36: Performance comparison of 2D-BIQA models on Waterloo-IVC 3D Image Database Phase I

	PLCC		SRCC		KRCC		RMSE		MAE	
2D-IQA/BIQA	Sym.	Asym.	Sym.	Asym.	Sym.	Asym.	Sym.	Asym.	Sym.	Asym.
Average 2DIQ-MOS	0.9801	0.8572	0.9657	0.8471	0.8482	0.6780	3.8266	7.9895	2.9160	5.8643
Weighted 2DIQ-MOS	0.9801	0.9460	0.9657	0.9324	0.8482	0.7890	3.8266	5.0298	2.9160	3.7254
Average IDW-SSIM	0.9579	0.7475	0.9371	0.7111	0.7725	0.5333	5.5328	10.3042	4.5967	7.8824
Weighted IDW-SSIM	0.9580	0.9254	0.9371	0.9002	0.7728	0.7273	5.5313	5.8810	4.5953	4.6420
Average BIQI	0.8674	0.7820	0.7965	0.7582	0.5951	0.5624	9.5962	9.6679	7.4527	7.4405
Weighted BIQI	0.8671	0.8864	0.7957	0.8528	0.5937	0.6692	9.6037	7.1801	7.4604	5.7658
Average BLIINDS-II	0.7583	0.5545	0.6766	0.5850	0.4983	0.4366	10.1131	16.0459	7.9037	12.2592
Weighted BLIINDS-II	0.8061	0.7925	0.6913	0.5854	0.5161	0.4359	9.1798	11.7597	6.9777	8.9039
Average BRISQUE	0.9371	0.8012	0.9102	0.7800	0.7536	0.6020	6.7321	9.2833	5.0274	7.2133
Weighted BRISQUE	0.9371	0.8873	0.9096	0.8339	0.7529	0.6705	6.7324	7.1541	5.0278	5.6135
Average CORNIA	0.9097	0.8252	0.8638	0.8150	0.6737	0.6309	8.0060	8.7630	6.3190	6.7914
Weighted CORNIA	0.9096	0.8847	0.8635	0.8470	0.6730	0.6675	8.0103	7.2326	6.3209	5.6995
Average DIIVINE	0.5872	0.3974	0.5753	0.3261	0.4399	0.2439	15.6071	14.2353	11.5002	10.8804
Weighted DIIVINE	0.5872	0.4937	0.5753	0.4269	0.4399	0.3219	15.6072	13.4909	11.5002	10.3680
Average LPSI	0.7642	0.5512	0.7165	0.5891	0.5431	0.4171	12.4367	12.9439	9.6287	10.2910
Weighted LPSI	0.7642	0.6118	0.7165	0.6353	0.5431	0.4542	12.4369	12.2705	9.6291	9.9735
Average M_3	0.9418	0.8265	0.9147	0.8114	0.7602	0.6279	6.4801	8.7335	4.8622	6.7558
Weighted M_3	0.9418	0.9154	0.9147	0.8942	0.7602	0.7274	6.4812	6.2436	4.8638	4.8895
Average NIQE	0.8276	0.7871	0.6395	0.6617	0.4752	0.4810	10.8217	9.5685	8.5048	7.8259
Weighted NIQE	0.8276	0.8311	0.6395	0.6596	0.4752	0.4920	10.8217	8.6277	8.5047	7.2452
Average QACS	0.8077	0.7627	0.5620	0.6448	0.4039	0.4635	11.3686	10.0337	8.9479	8.1177
Weighted QACS	0.8077	0.8291	0.5620	0.6456	0.4039	0.4720	11.3689	8.6731	8.9481	7.1790
Average TCLT	0.8609	0.6624	0.7632	0.6245	0.6004	0.4524	9.8091	11.6216	7.9132	9.1888
Weighted TCLT	0.8609	0.8438	0.7620	0.7612	0.5997	0.5868	9.8103	8.3259	7.9145	6.5492

Table 3.37: Performance comparison of 2D-BIQA models on Waterloo-IVC 3D Image Database Phase II

	PLCC		SRCC		KRCC		RMSE		MAE	
2D-IQA/BIQA	Sym.	Asym.	Sym.	Asym.	Sym.	Asym.	Sym.	Asym.	Sym.	Asym.
Average 2DIQ-MOS	0.9799	0.8418	0.9696	0.8501	0.8557	0.6693	4.0525	10.4221	3.1712	7.9037
Weighted 2DIQ-MOS	0.9799	0.9433	0.9696	0.9302	0.8557	0.7874	4.0525	6.4106	3.1712	4.9383
Average IDW-SSIM	0.9360	0.7480	0.9017	0.7410	0.7242	0.5509	7.1484	12.8163	5.6536	9.9539
Weighted IDW-SSIM	0.9360	0.8844	0.9017	0.8630	0.7244	0.6833	7.1483	9.0137	5.6541	7.3627
Average BIQI	0.7343	0.6089	0.6239	0.5795	0.4650	0.4062	13.7833	15.3174	11.4632	11.9922
Weighted BIQI	0.7344	0.7695	0.6239	0.7071	0.4650	0.5279	13.7813	12.3321	11.4618	9.9003
Average BLIINDS-II	0.7774	0.4747	0.7586	0.5962	0.5673	0.4364	12.1457	17.8708	9.7039	14.1579
Weighted BLIINDS-II	0.8223	0.7389	0.7829	0.5951	0.5970	0.4355	10.9899	13.6824	8.7019	11.3054
Average BRISQUE	0.9293	0.6577	0.9022	0.7957	0.7360	0.6119	7.4978	14.5464	5.7534	11.8813
Weighted BRISQUE	0.9293	0.8780	0.9023	0.8627	0.7362	0.6957	7.5007	9.2444	5.7546	7.1089
Average CORNIA	0.9370	0.6490	0.9160	0.7877	0.7417	0.5993	7.0948	14.6912	5.9472	12.0882
Weighted CORNIA	0.9369	0.8762	0.9164	0.8535	0.7422	0.6751	7.0994	9.3076	5.9524	7.2591
Average DIIVINE	0.6512	0.4449	0.6234	0.3920	0.4879	0.2768	15.4101	17.2936	11.9414	13.2858
Weighted DIIVINE	0.6513	0.5905	0.6238	0.4910	0.4884	0.3533	15.4068	15.5844	11.9394	12.0956
Average LPSI	0.7526	0.5859	0.7809	0.6076	0.6052	0.4224	13.3701	15.6488	10.9118	12.9794
Weighted LPSI	0.7526	0.7092	0.7809	0.7207	0.6052	0.5165	13.3692	13.6140	10.9111	11.5828
Average M_3	0.9114	0.7618	0.8713	0.7553	0.6978	0.5638	8.3555	12.5095	6.7502	9.7012
Weighted M_3	0.9114	0.8644	0.8717	0.8399	0.6985	0.6594	8.3541	9.7096	6.7487	7.6974
Average NIQE	0.7429	0.6468	0.5282	0.5758	0.3853	0.4104	13.5916	14.7272	11.3547	12.0440
Weighted NIQE	0.7429	0.7656	0.5284	0.6103	0.3856	0.4433	13.5920	12.4224	11.3547	10.2187
Average QACS	0.7715	0.7125	0.5305	0.6567	0.3803	0.4721	12.9173	13.5501	10.9738	11.1365
Weighted QACS	0.7716	0.7994	0.5303	0.6757	0.3801	0.4958	12.9168	11.6028	10.9734	9.6357
Average TCLT	0.8135	0.6075	0.7099	0.5735	0.5470	0.3984	11.8072	15.3378	9.4792	12.3072
Weighted TCLT	0.8136	0.7558	0.7098	0.7177	0.5468	0.5430	11.8051	12.6452	9.4774	10.0603

Table 3.38: Performance comparison of 2D-BIQA models on LIVE 3D Image Database Phase II

	PLCC		SRCC		KRCC		RMSE		MAE	
2D-IQA/BIQA	Sym.	Asym.	Sym.	Asym.	Sym.	Asym.	Sym.	Asym.	Sym.	Asym.
Average IDW-SSIM	0.9363	0.7383	0.9227	0.6902	0.7545	0.5216	4.3849	6.8363	3.4405	5.2556
Weighted IDW-SSIM	0.9365	0.8601	0.9231	0.8539	0.7554	0.6652	4.3796	5.1696	3.4354	4.0833
Average BIQI	0.8350	0.6794	0.8278	0.6247	0.6293	0.4514	6.8700	7.4372	5.2966	6.0588
Weighted BIQI	0.8351	0.7455	0.8276	0.6971	0.6296	0.5119	6.8671	6.7552	5.2924	5.3710
Average BLIINDS-II	0.6649	0.5520	0.6243	0.5334	0.4471	0.3725	9.3253	8.4510	7.9703	6.9464
Weighted BLIINDS-II	0.6649	0.6146	0.6237	0.6011	0.4453	0.4218	9.3257	7.9951	7.9667	6.5302
Average BRISQUE	0.8688	0.6993	0.8491	0.6670	0.6506	0.4829	6.1821	7.2445	4.9924	5.8667
Weighted BRISQUE	0.8575	0.7513	0.8493	0.7365	0.6512	0.5406	6.4229	6.6890	5.0676	5.3923
Average CORNIA	0.8748	0.7026	0.8748	0.6778	0.6852	0.4940	6.0485	7.2123	4.7592	5.7550
Weighted CORNIA	0.8749	0.7567	0.8752	0.7398	0.6863	0.5471	6.0461	6.6254	4.7531	5.2791
Average DIIVINE	0.7835	0.6035	0.7808	0.5536	0.5800	0.3914	7.7664	8.0815	6.1737	6.6866
Weighted DIIVINE	0.7697	0.6764	0.7806	0.6196	0.5797	0.4428	7.9702	7.4646	6.2219	6.0382
Average LPSI	0.7611	0.4898	0.7483	0.4051	0.5612	0.2955	8.0976	8.8361	6.3656	7.2424
Weighted LPSI	0.7611	0.6056	0.7488	0.4979	0.5621	0.3718	8.0979	8.0653	6.3685	6.5092
Average M_3	0.8512	0.7136	0.8480	0.7126	0.6607	0.5184	6.5532	7.1006	5.0660	5.6436
Weighted M_3	0.8329	0.7695	0.8476	0.7668	0.6593	0.5682	6.9095	6.4727	5.4987	5.0552
Average NIQE	0.7801	0.7383	0.7592	0.7120	0.5719	0.5161	7.8110	6.8359	6.1195	5.5881
Weighted NIQE	0.7801	0.7751	0.7585	0.7557	0.5710	0.5605	7.8107	6.4037	6.1182	5.0662
Average QACS	0.8821	0.7388	0.8727	0.6960	0.6741	0.5101	5.8806	6.8305	4.7616	5.3262
Weighted QACS	0.8822	0.8169	0.8728	0.7858	0.6741	0.5963	5.8791	5.8453	4.7609	4.4927
Average TCLT	0.8252	0.6961	0.7972	0.6564	0.6044	0.4797	7.0526	7.2769	5.5719	5.8691
Weighted TCLT	0.8253	0.7093	0.7981	0.6845	0.6055	0.4974	7.0503	7.1447	5.5708	5.7172

quality of stereoscopic images from that of the single-view images was applied to 2DIQ-MOS scores and different base 2D-IQA measures. The experimental results showed that the quality prediction performance is significantly improved for most base 2D-IQA methods as well as with 2DIQ-MOS scores. Unfortunately, the model proposed in Section 3.6 requires access to the pristine reference stereopairs, which are not available in the case of 3D-BIQA.

In this section, we aim to develop an objective 3D-BIQA predictor. We take advantage of the previous findings on the relationship between the perceptual quality of stereoscopic images and that of its single-view images. We assume that existing successful 2D-BIQA methods are reliable for evaluating single-view images, and what is missing is an effective blind 2D-to-3D prediction model to combine single-vew quality scores, so as to eliminate the prediction bias for asymmetric distortions. Therefore, we opt to use a two-stage approach. The first stage builds a binocular rivalry inspired multi-scale 2D-to-3D quality prediction model without referring to the original images. In the second stage, this quality prediction model is applied to combine state-of-the-art 2D-BIQA estimations of both views, resulting in a 3D image quality estimation.

3.7.2 Blind 2D-to-3D Quality Prediction

In Section 3.6, motivated by existing vision studies on binocular rivalry, where it was found that for simple ideal stimuli, an increase in contrast enhances the predominance of one view against the other, we showed that the strength of view dominance in binocular rivalry of stereoscopic images is related to the relative energy of the two views. However, the computation of the relative energy involves the original left- and right-view images, which are not available in the case of 3D-BIQA. In this section, to overcome the problem, we apply a divisive normalization transform (DNT) [48, 131, 41] to the distorted left- and right-view images, and then estimate the strength of view dominance from DNT domain representations.

A DNT is typically built upon a linear image decomposition, followed by a divisive normalization stage [73]. The linear transformations may be discrete cosine transform or wavelet-type of transforms. Here, we assume a wavelet image decomposition, which provides a convenient framework for localized representation of images simultaneously in

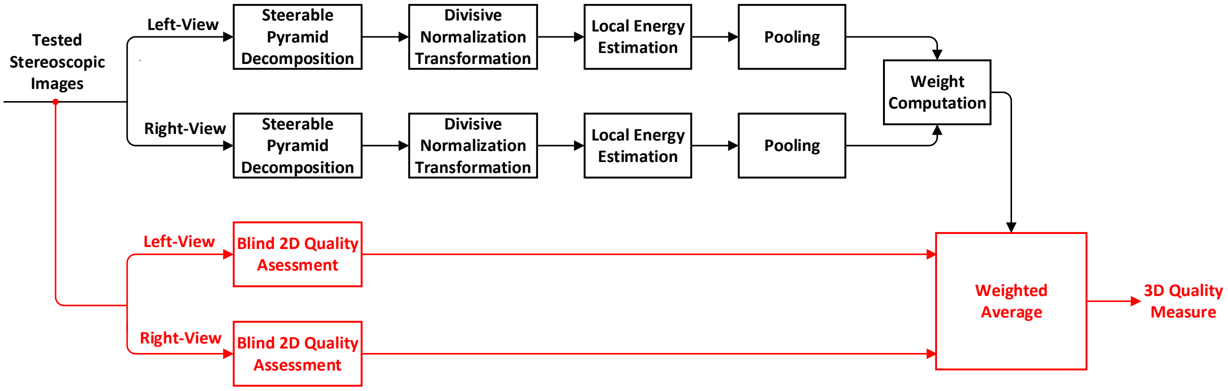


Figure 3.28: Diagram of the proposed blind 2D-to-3D quality prediction model.

space, frequency (scale) and orientation. The DNT image representation of the image is then calculated by dividing each wavelet coefficient by a local energy measure based on its neighboring coefficients. The DNT image representation is not only an effective way to reduce the statistical redundancies between wavelet coefficients [73], but is also highly relevant to biological vision [20]. In [110], a divisive normalization framework was applied to develop a computational model that trades off between binocular rivalry and suppression, and the predictions were confirmed with psychophysical tests. This binocular perceptual relevance of divisive normalization representation leads us to design a multi-scale and multi-orientation DNT domain 2D-to-3D prediction model without referring to the original left- and right-view images.

The diagram of the proposed method is shown in Figure 3.28. Let $(I_{d,l}, I_{d,r})$ be the left- and right-view image pairs of the distorted stereoscopic images. To compute the DNT representation of $I_{d,l}$ and $I_{d,r}$, we first apply a 3-scale, 2-orientation steerable pyramid wavelet transform [130] to decompose $I_{d,l}$ and $I_{d,r}$ into 6 oriented subbands (2 for each scale) and a highpass and a lowpass residual subbands, respectively.

At the i-th oriented subband, for each center coefficient y_c, we define a DNT neighboring vector Y that contains 11×11 coefficients from the same subband (including the center coefficient itself). As such, the corresponding DNT center coefficient $\tilde{y}_c$ at the i-th normalized subband is computed as

$$\tilde{y}_c = \frac{y_c}{\sum |Y| + c}\,, \tag{3.12}$$

where $c = 1$.

For each divisive normalized subband, we estimate its energy by computing the local variances at each spatial location, i.e., the variances of local patches extracted around each spatial location from the DNT coefficients are computed, for which an 11×11 circular-symmetric Gaussian weighting function $\mathbf{w} = \{w_i | i = 1, 2, \cdots, N\}$ with standard deviation of 1.5 samples, normalized to unit sum ($\sum_{i=1}^{N} w_i = 1$), is employed. The resulting mean energies for the i-th normalized subband in the $I_{d,l}$ and $I_{d,r}$ are denoted as $E_{i,l}$ and $E_{i,r}$, respectively. The overall energy estimations in both views are computed as the sum of the energies of all divisive normalized subbands

$$g_l = \sum_{i=1}^{6} E_{i,l} \qquad \text{and} \qquad g_r = \sum_{i=1}^{6} E_{i,r}\,, \tag{3.13}$$

Here g_l and g_r are estimations of the level of dominance of the left- and right-views, respectively. Given the values of g_l and g_r, the weights assigned to the left- and right-view images are given by

$$w_l = \frac{g_l^2}{g_l^2 + g_r^2} \qquad \text{and} \qquad w_r = \frac{g_r^2}{g_l^2 + g_r^2}\,, \tag{3.14}$$

respectively.

Finally, the overall prediction of 3D image quality is calculated by a weighted average of the left- and right-view image quality:

$$Q^{3D} = w_l Q_l^{2D} + w_r Q_r^{2D}\,, \tag{3.15}$$

where Q_l^{2D} and Q_r^{2D} denote the 2D image quality of the left- and right-views, respectively.

3.7.3 Validation

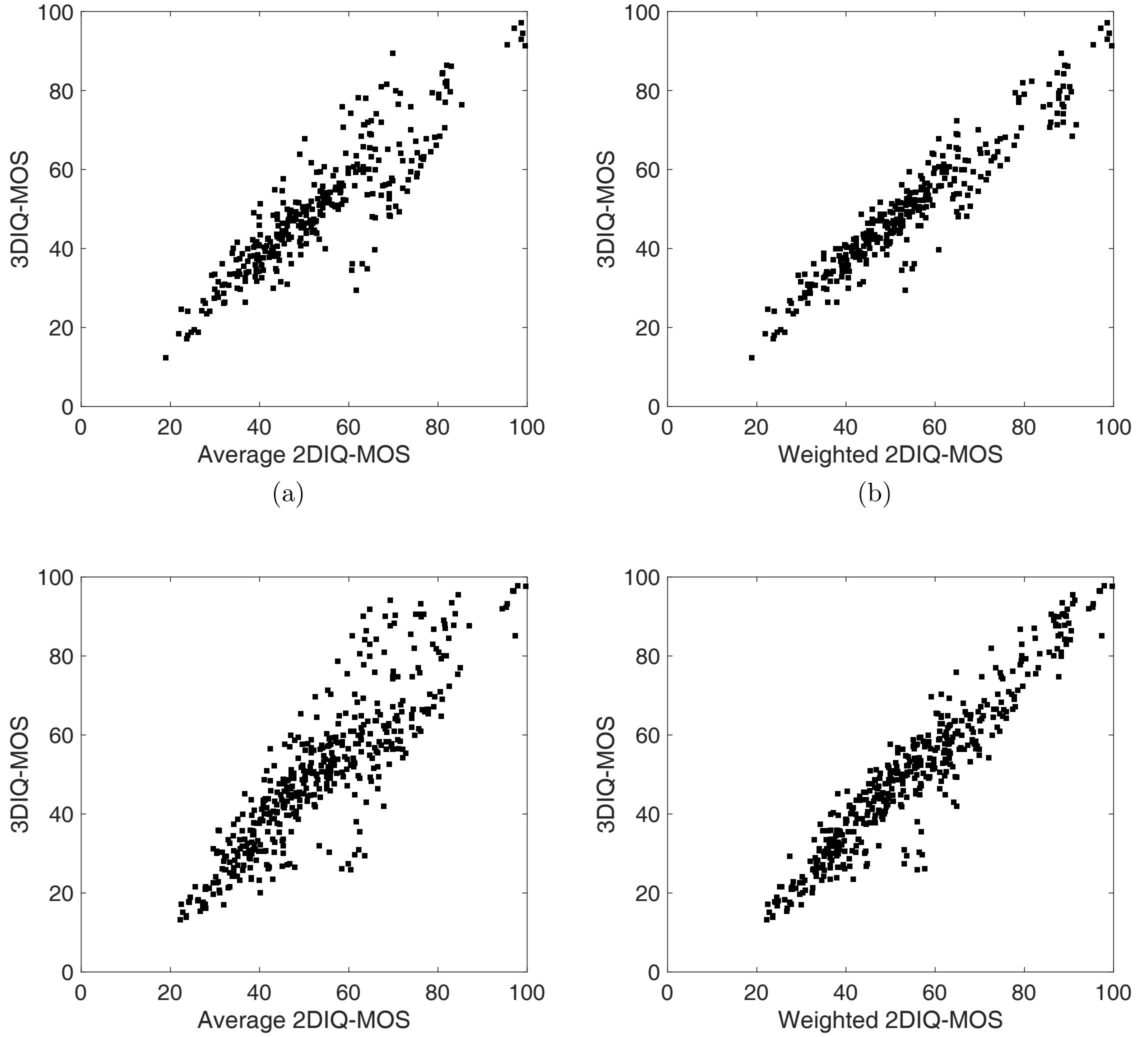
We use three 3D image quality databases to test the proposed algorithm, which are the Waterloo-IVC 3D Image Databases Phase I and Phase II and the LIVE 3D Image Database Phase II. All these databases contain both symmetrically and asymmetrically distorted stereoscopic images. The parameters of the proposed blind 2D-to-3D quality prediction method are selected empirically when working with Waterloo-IVC database Phase I, but are completely independent of Waterloo-IVC database Phase II and the LIVE database Phase II.

Blind 2D-to-3D quality prediction with 2D-BIQA

We first test the proposed blind 2D-to-3D quality prediction model on all 3D images in Waterloo-IVC database by applying it to the ground truth 2DIQ-MOS scores. The PLCC, SRCC, KRCC, RMSE and MAE values between 3DIQ-MOS and the predicted Q^{3D} value for all stereoscopic images and for each test image group are given in Table 3.39 (Please refer to Section 3.2 for categories of Waterloo-IVC database). The comparison results with our FR 2D-to-3D quality prediction model (Section 3.6.2) are also given in Table 3.39. The corresponding scatter plots are shown in Figure 3.29 to Figure 3.33. From Table 3.39, it can be observed that the proposed blind model performs as well as the FR 2D-to-3D quality prediction model. The proposed model outperforms the direct averaging method in almost all cases, and the improvement is most pronounced in the case of strong asymmetric distortions (Group 3D.2) or when all test images are put together (All 3D image case). By comparing different sub-figures of Figure 3.29 to Figure 3.33, we observe the impact of the proposed blind 2D-to-3D prediction model on each image distortion type. For different distortion types, although the direct averaging method produces different distortions and levels of quality prediction biases, the proposed method, which does not attempt to recognize the distortion types or give any specific treatment for any specific distortion type, removes or significantly reduces the prediction biases for all distortion types. Moreover, for the mixed distortion case that provides the strongest test on the generalization ability of the model, the proposed method maintains consistent performance.

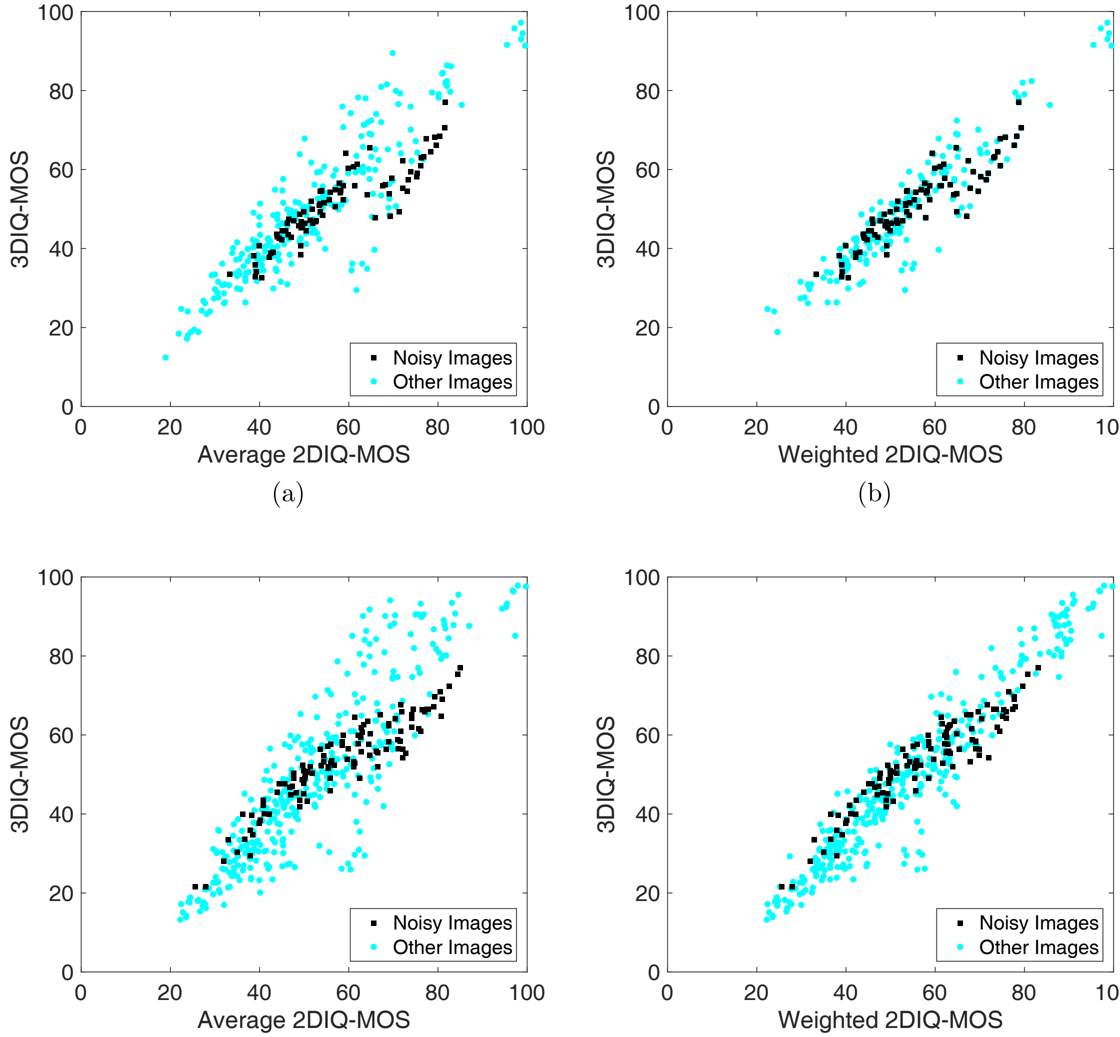
Table 3.39: Performance comparison of 2D-to-3D quality prediction models (direct average, FR (Section 3.6.2), and the proposed blind prediction model), where the single-view quality is given by ground truth 2DIQ-MOS

		Waterloo-IVC 3D Image Database Phase I					Waterloo-IVC 3D Image Database Phase II				
Group	Method	PLCC	SRCC	KRCC	RMSE	MAE	PLCC	SRCC	KRCC	RMSE	MAE
All	Average	0.8835	0.8765	0.7161	7.3700	5.3293	0.8763	0.8820	0.7145	9.2201	6.7164
	FR	0.9561	0.9522	0.8162	4.6108	3.5645	0.9568	0.9477	0.8080	5.5655	4.3371
	Proposed	0.9509	0.9413	0.8045	4.8718	3.6164	0.9507	0.9382	0.8012	5.9343	4.5083
Sym.	Average	0.9801	0.9657	0.8482	3.8266	2.9160	0.9799	0.9696	0.8557	4.0525	3.1712
	FR	0.9801	0.9657	0.8482	3.8266	2.9160	0.9799	0.9696	0.8557	4.0525	3.1712
	Proposed	0.9801	0.9657	0.8482	3.8266	2.9160	0.9799	0.9696	0.8557	4.0525	3.1712
Asym.	Average	0.8572	0.8471	0.6780	7.9895	5.8643	0.8418	0.8501	0.6693	10.4221	7.9037
	FR	0.9522	0.9452	0.8026	4.7406	3.6791	0.9511	0.9424	0.7977	5.9620	4.6683
	Proposed	0.9460	0.9324	0.7890	5.0298	3.7254	0.9433	0.9302	0.7874	6.4106	4.9383
3D.1	Average	0.9801	0.9657	0.8482	3.8266	2.9160	0.9799	0.9696	0.8557	4.0525	3.1712
	FR	0.9801	0.9657	0.8482	3.8266	2.9160	0.9799	0.9696	0.8557	4.0525	3.1712
	Proposed	0.9801	0.9657	0.8482	3.8266	2.9160	0.9799	0.9696	0.8557	4.0525	3.1712
3D.2	Average	0.6613	0.5433	0.4406	11.6394	9.0069	0.6121	0.5874	0.4524	14.2881	11.6204
	FR	0.9286	0.9160	0.7556	5.7574	4.7098	0.9414	0.9497	0.8070	6.0932	4.8568
	Proposed	0.9370	0.9141	0.7542	5.4196	4.4380	0.9459	0.9482	0.8001	5.8624	4.7605
3D.3	Average	0.9666	0.9164	0.7597	3.6078	2.7578	0.9471	0.8898	0.7176	5.7909	4.4262
	FR	0.9714	0.9307	0.7789	3.3391	2.6247	0.9602	0.9318	0.7745	5.0392	4.0923
	Proposed	0.9772	0.9427	0.7969	2.9870	2.3547	0.9665	0.9380	0.7867	4.6340	3.6830
3D.4	Average	0.9223	0.8271	0.6390	5.9710	4.5387	0.9225	0.8798	0.7047	7.0928	5.4776
	FR	0.9656	0.9357	0.7822	4.0179	3.2243	0.9549	0.9320	0.7803	5.4557	4.2124
	Proposed	0.9664	0.9366	0.7862	3.9740	3.1426	0.9652	0.9496	0.8101	4.8031	3.6934



(a) (b) (c) (d)

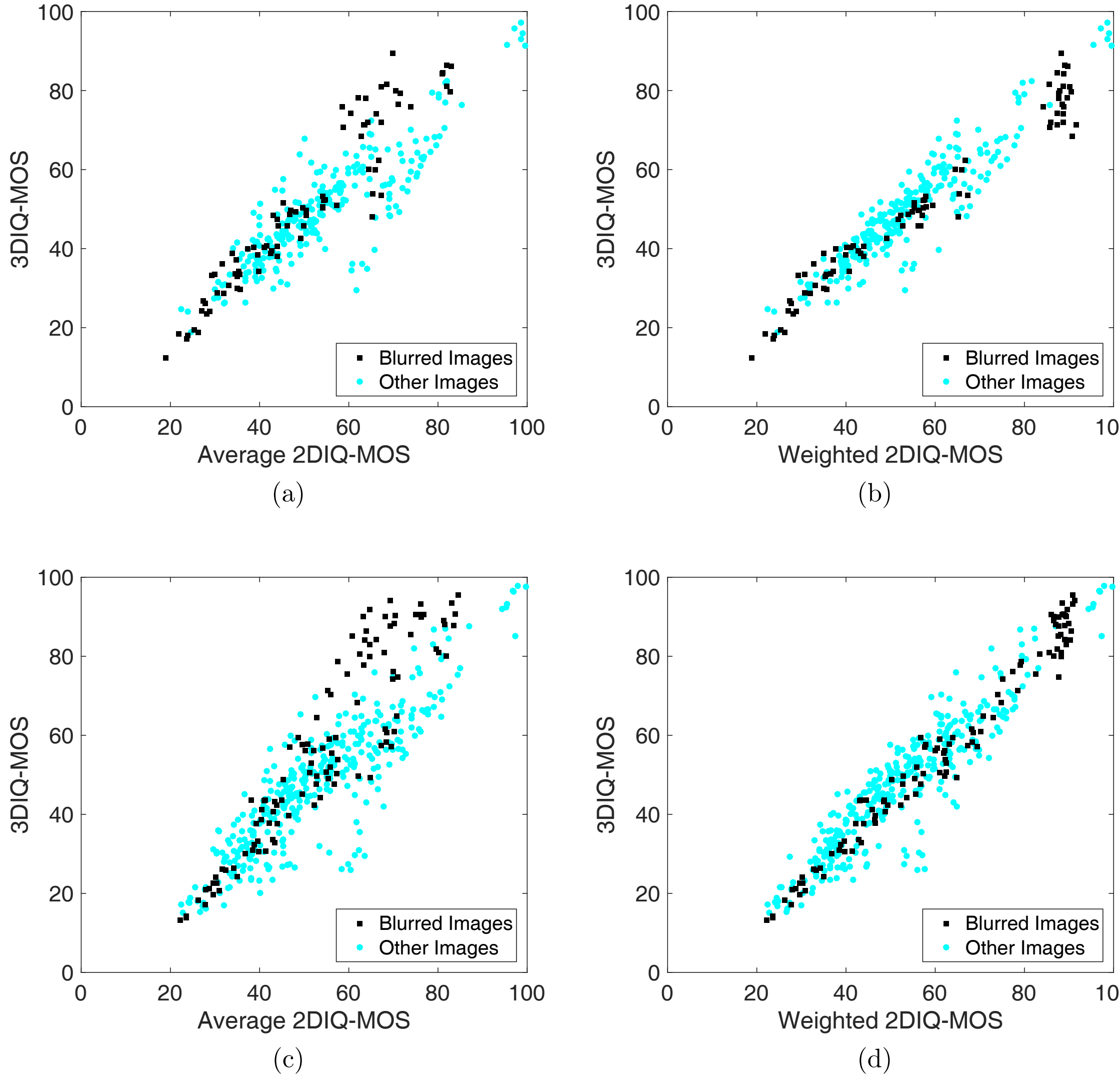
Figure 3.29: 3DIQ-MOS versus predictions from 2DIQ-MOS of 2D left- and right-views. (a) All Images, predictions by direct averaging the 2DIQ-MOS scores of both views on Waterloo-IVC Phase I. (b) All Images, predictions by the proposed model on Waterloo-IVC Phase I. (c) All Images, predictions by direct averaging 2DIQ-MOS scores of both views on Waterloo-IVC Phase II. (d) All Images, predictions by the proposed model on Waterloo-IVC Phase II.



(a) (b)

(c) (d)

Figure 3.30: 3DIQ-MOS versus predictions from 2DIQ-MOS of 2D left- and right-views. (a) Noisy Images, predictions by direct averaging the 2DIQ-MOS scores of both views on Waterloo-IVC Phase I. (b) Noisy Images, predictions by the proposed model on Waterloo-IVC Phase I. (c) Noisy Images, predictions by direct averaging 2DIQ-MOS scores of both views on Waterloo-IVC Phase II. (d) Noisy Images, predictions by the proposed model on Waterloo-IVC Phase II.



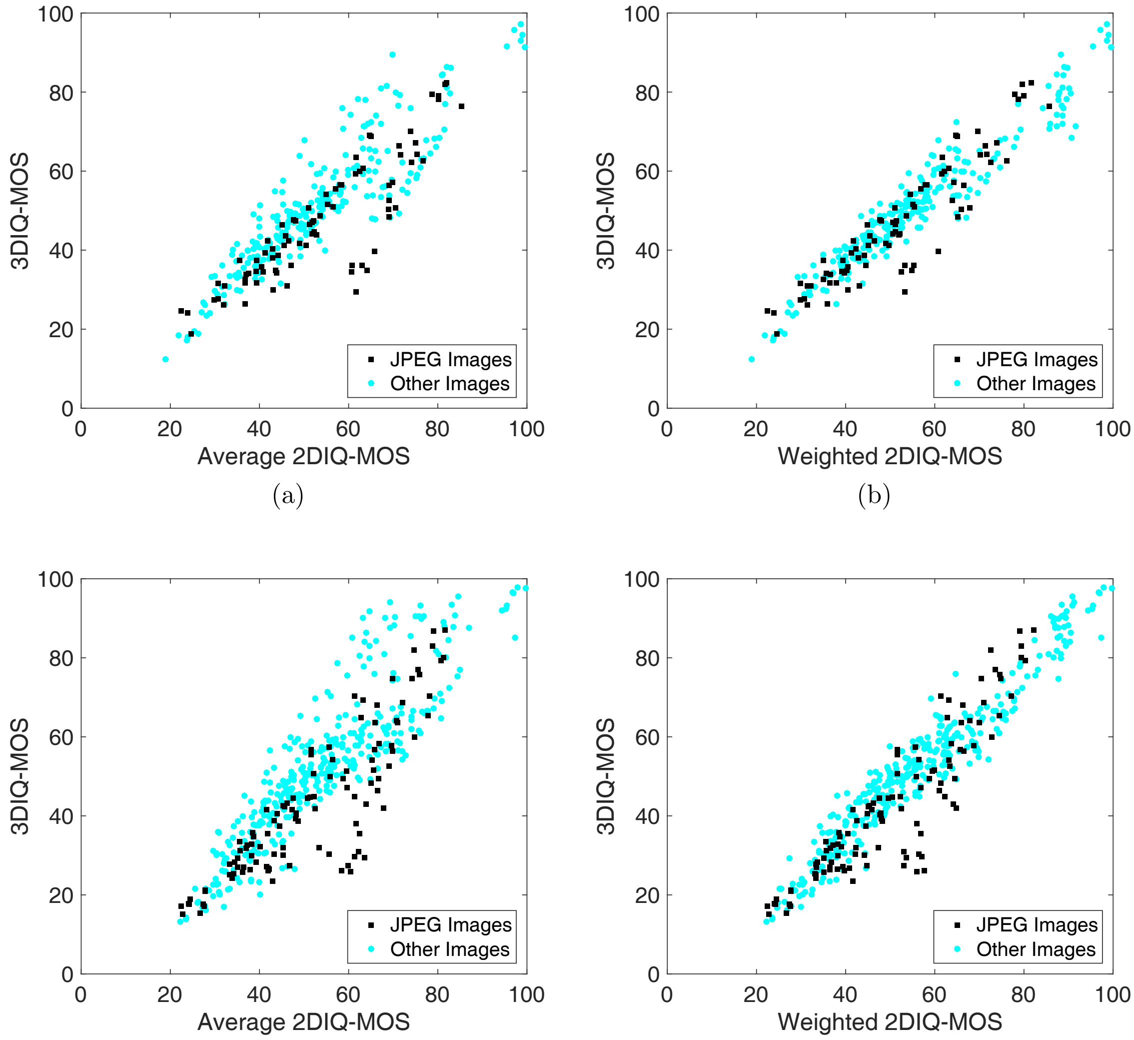
(a)

(b)

(c)

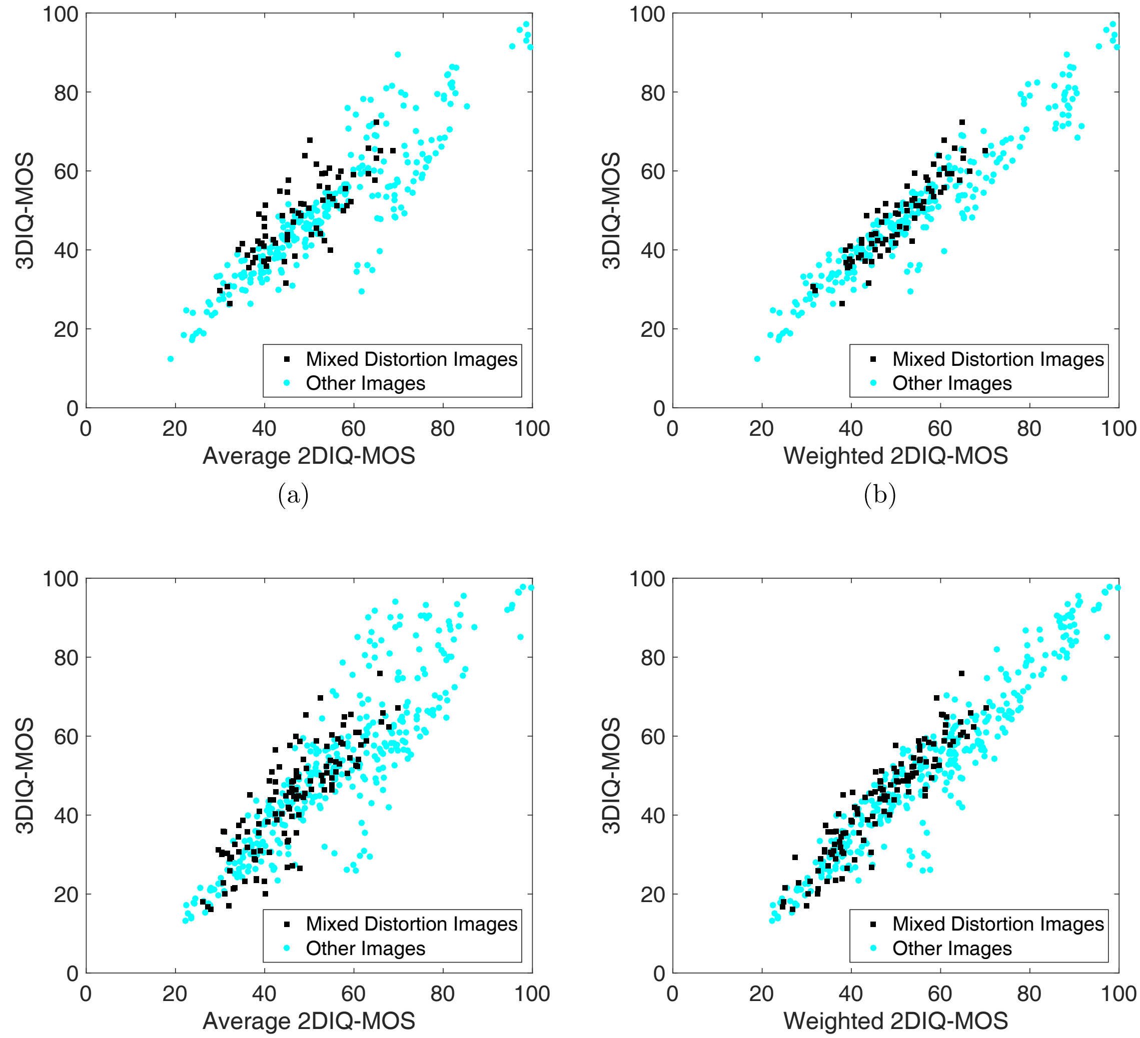
(d)

Figure 3.31: 3DIQ-MOS versus predictions from 2DIQ-MOS of 2D left- and right-views. (a) Blurred Images, predictions by direct averaging the 2DIQ-MOS scores of both views on Waterloo-IVC Phase I. (b) Blurred Images, predictions by the proposed model on Waterloo-IVC Phase I. (c) Blurred Images, predictions by direct averaging 2DIQ-MOS scores of both views on Waterloo-IVC Phase II. (d) Blurred Images, predictions by the proposed model on Waterloo-IVC Phase II.



(a) (b) (c) (d)

Figure 3.32: 3DIQ-MOS versus predictions from 2DIQ-MOS of 2D left- and right-views. (a) JPEG Images, predictions by direct averaging the 2DIQ-MOS scores of both views on Waterloo-IVC Phase I. (b) JPEG Images, predictions by the proposed model on Waterloo-IVC Phase I. (c) JPEG Images, predictions by direct averaging 2DIQ-MOS scores of both views on Waterloo-IVC Phase II. (d) JPEG Images, predictions by the proposed model on Waterloo-IVC Phase II.



(a) (b)

(c) (d)

Figure 3.33: 3DIQ-MOS versus predictions from 2DIQ-MOS of 2D left- and right-views. (a) Mixed Distortion, predictions by direct averaging the 2DIQ-MOS scores of both views on Waterloo-IVC Phase I. (b) Mixed Distortion, predictions by the proposed model on Waterloo-IVC Phase I. (c) Mixed Distortion, predictions by direct averaging 2DIQ-MOS scores of both views on Waterloo-IVC Phase II. (d) Mixed Distortion, predictions by the proposed model on Waterloo-IVC Phase II.

Blind 2D-to-3D quality prediction with 2D-BIQA

Before applying the proposed blind 2D-to-3D quality prediction model on the base 2D-BIQA methods, we first examine these 2D-BIQA methods' abilities in predicting 2DIQ-MOS scores of single-view images in Waterloo-IVC 3D Image Database. Table 3.40 reports PLCC, SRCC, KRCC, RMSE and MAE values between 2DIQ-MOS scores and 2D-BIQA estimations. From Table 3.40, it can be seen that BRISQUE, CORNIA and M_3 archive the highest correlations with subjective data among all tested 2D-BIQA methods.

Table 3.40: Performance comparison of 2D-BIQA models on Waterloo-IVC 3D Image Database (Single-view Images)

	Waterloo-IVC 3D Image Database Phase I					Waterloo-IVC 3D Image Database Phase II				
	PLCC	SRCC	KRCC	RMSE	MAE	PLCC	SRCC	KRCC	RMSE	MAE
BIQI	0.8694	0.7790	0.5911	9.4664	7.5939	0.7098	0.5798	0.4182	13.1089	11.0148
BLIINDS-II	0.8169	0.5977	0.4397	11.0524	8.5587	0.5436	0.6752	0.5067	15.6194	12.2840
BRISQUE	0.9223	0.8912	0.7136	7.4047	5.6882	0.9238	0.8892	0.7207	7.1251	5.4293
CORNIA	0.9260	0.8861	0.7096	7.2333	5.9434	0.9306	0.8934	0.7250	6.8119	5.3488
DIIVINE	0.5943	0.5804	0.4406	15.4111	10.9213	0.5660	0.5631	0.4247	15.3409	11.1501
LPSI	0.7537	0.6679	0.5052	12.5940	9.8846	0.7478	0.6962	0.5237	12.3556	10.0321
M_3	0.9353	0.9040	0.7329	6.7794	5.0253	0.9038	0.8534	0.6661	7.9632	6.4404
NIQE	0.8423	0.6685	0.5038	10.3296	8.5242	0.7384	0.5616	0.4016	12.5496	10.8120
QACS	0.7920	0.5543	0.3813	11.6987	9.3612	0.7692	0.5452	0.3865	11.8906	9.9032
TCLT	0.8471	0.6977	0.5591	10.1840	7.6408	0.7900	0.6345	0.4793	11.4086	9.0854

We then test the proposed blind 2D-to-3D quality prediction model by applying it to different base 2D-BIQA approaches on all three databases. Note that exactly the same blind 2D-to-3D quality prediction model obtained from 2DIQ-MOS and 3DIQ-MOS scores with Waterloo-IVC 3D image database Phase I is used and thus the model is completely independent of any base objective 2D-BIQA approaches. Table 3.34 to Table 3.38 report PLCC, SRCC, KRCC, RMSE and MAE values between 3DIQ-MOS and the predicted Q^{3D} value with the direct averaging method and the proposed blind 2D-to-3D quality prediction model. Note that the cases of using 2DIQ-MOS and IDW-SSIM are also included for comparison.

From Table 3.34 to Table 3.38, it can be seen that the proposed method significantly improves most base 2D-BIQA methods on both databases. On the Waterloo-IVC 3D Image Database, BRISQUE, CORNIA and M_3 perform better than all competing 2D-BIQA methods with both the direct averaging and the proposed prediction model, which is consistent with their performance on single-view images. Interestingly, the performance of these 2D-BIQA methods approximates that of IDW-SSIM, which gives the most accurate prediction among FR 2D-IQA methods on both Waterloo-IVC 3D and LIVE 3D database (See Section 3.6.3). This suggests that a good 2D-BIQA method can predict symmetrically distorted stereoscopic images with good accuracy, and when properly combined with a 2D-to-3D quality prediction model, can also well predict asymmetrically distorted stereoscopic images.

On the LIVE 3D Image Database Phase II, the proposed method achieves the best performance in the case of using QAC and also pronounces competitive performance with BRISQUE, CORNIA and M_3. However, there is a large gap when compared with the FR IDW-SSIM's prediction performance. This suggests that there is still potential to further improve 2D-BIQA methods in terms of robustness and generalizability.

We have also compared the proposed method with state-of-the-art 3D-BIQA approaches [23, 122, 4, 47] on the LIVE 3D image database Phase II. The PLCC, SRCC, KRCC, RMSE and MAE values are reported in Table 3.41. From Table 3.35 and Table 3.41, it can be seen that the proposed method, when combined with QAC, performs better than [122, 4, 47] but not as good as [23], which is a training-based method and the results reported here are median performance of 1000 trails, each uses 80% of the data for training and 20% for testing.

Table 3.41: Performance comparison of 3D-BIQA models on LIVE 3D image database Phase II

3D-BIQA	PLCC		SRCC	
Akhter [4]	N/A	N/A	0.4200	0.5170
Chen [23]	N/A	N/A	0.9180	0.8340
Gu [47]	0.0994	0.2271	0.1760	0.1141
Shao [122]	0.9119	0.5651	0.8966	0.5244

Computational complexity analysis

Speed is another important performance factor in evaluating a BIQA method. We use program running time in the test stage of all competing methods as an estimate of computational complexity. The average processing time for a single-view image and for a stereopair on Waterloo-IVC database Phase II and LIVE database Phase II is summarized in Table 3.42. The system platform is Intel(R) Core(TM) i7-3770 @3.40GHz, 16.0 GB RAM and Windows 7 64-bit version. All methods are tested with the MATLAB R2015a software. Note that the resolutions of the single-view images are 1920 × 1080 and 640 × 360 for Waterloo-IVC database Phase II and LIVE database Phase II, respectively. Also, the average running time for the proposed 2D-to-3D quality prediction model is 1.8079s for Waterloo-IVC database Phase II and 0.2175s for LIVE database Phase II. The total processing time for a stereopair should be computed as twice of the processing time for a single-view image plus the time for 2D-to-3D prediction. From Table 3.42, it can be seen that BLIINDS-II and DIIVINE are the slowest, while LPSI and M_3 are the fastest. Generally speaking, BRISQUE and M_3 achieve excellent tradeoffs between accuracy and complexity.

Table 3.42: Complexity comparison of 2D-BIQA models on 3D image databases based on average running time (seconds)

2D-BIQA	Waterloo 2D	Waterloo 3D	LIVE 2D	LIVE 3D
BIQI	0.7211	3.2501	0.4539	1.1253
BLIINDS-II	435.1705	872.1489	48.4902	97.1979
BRISQUE	**0.5219**	**2.8517**	**0.1510**	**0.5195**
CORNIA	3.9315	9.6709	2.4399	5.0973
DIIVINE	86.1006	174.0091	11.8495	23.9165
LPSI	**0.1220**	**2.0519**	**0.0130**	**0.2435**
$\mathbf{M_3}$	**0.3095**	**2.4269**	**0.0347**	**0.2869**
NIQE	1.3471	4.5021	0.1181	0.4537
QACS	**0.4539**	**2.7157**	**0.0483**	**0.3141**
TCLT	8.7635	19.3349	1.1605	2.5385

3.8 Summary

The major contributions of the current chapter are as follows: First, we create a new subjective 3D-IQA database that has two unique features – the inclusion of both 2D and 3D images, and the inclusion of mixed distortion types. Second, we observe strong distortion type dependent bias when using the direct average of 2D image quality of both views to predict 3D image quality. Third, we observe that eye dominance does not have strong impact on visual quality evaluations of asymmetrically distorted stereoscopic images. Fourth, we develop an information content and divisive normalization based pooling scheme that improves upon SSIM in estimating the quality of single-view images. Fifth, we propose a binocular rivalry inspired multi-scale model to predict the quality of stereoscopic images from that of its single-view 2D images. Our results show that the proposed model, without explicitly identifying image distortion types, successfully eliminates the prediction bias, leading to significantly improved quality prediction of stereoscopic 3D images. The performance gain is most pronounced in the case of asymmetric distortions. Sixth, we propose a binocular rivalry inspired multi-scale model to predict the quality of stereoscopic images from that of the single-view images without referring to the original left- and right-view images. We apply the proposed blind 2D-to-3D quality prediction model to ten state-of-the-art base 2D-BIQA measures for 3D-BIQA. Among all the base 2D-BIQA methods, BRISQUE and M_3 achieve excellent tradeoffs between accuracy and complexity.

Chapter 4

Quantifying Perceptual Depth Quality in Distorted Stereoscopic Images

Subjective and objective measurement of the perceptual quality of depth information in symmetrically and asymmetrically distorted stereoscopic images is a fundamentally important issue in stereoscopic 3D imaging that has not been deeply investigated. Here we first carry out a subjective test following the traditional absolute category rating protocol widely used in general image quality assessment research. We find this approach problematic because monocular cues and the spatial quality of images have strong impact on the depth quality scores given by subjects, making it difficult to single out the actual contributions of stereoscopic cues in depth perception. To overcome this problem, we carry out a novel subjective study where depth effect is synthesized at different depth levels before various types and levels of symmetric and asymmetric distortions are applied. Instead of following the traditional approach, we ask subjects to identify and label depth polarizations, and a notion of Depth Perception Difficulty Index (DPDI) is proposed based on the percentage of correct and incorrect subject judgements. We find this approach highly effective at quantifying depth perception induced by stereo cues and observe a number of interesting effects regarding image content dependency, distortion type dependency, and the impact of symmetric versus asymmetric distortions. Furthermore, we propose a novel computational model for DPDI prediction. Our results show that the proposed model, without explicitly identifying image distortion types, leads to highly promising DPDI prediction performance. We believe these are useful steps towards building a comprehensive understanding on 3D QoE of stereoscopic images.

4.1 Introduction

Depth quality is an essential aspect of human QoE when viewing stereoscopic 3D images. Existing studies on the topic appear to be unconclusive, limited, and sometimes conflicting. In [53], it was reported that the perceived depth performance cannot always be predicted from displaying image geometry alone, while other system factors, such as software drivers, electronic interfaces, and individual participant differences, may also play significant roles. In [119, 26], subjective studies suggested that increasing the degree of binocular depth does increase the perceived depth quantity. In [119, 64], it was suggested that depth quality may need to be considered independently from perceived 3D image quality. The results in [119] showed that increased JPEG coding has no effect on depth perception however a negative effect on image quality. In [26, 137], subjective studies suggested that 3D image quality is not sensitive to variations in the degree of binocular depth.

Other studies pointed out perceptual depth quality as an important component in the holistic 3D QoE. In [177], a blurring filter, where the level of blur depends on the depth of the area where it is applied, is used to enhance the viewing experience. In [116], subjective studies revealed that humans tend to prefer DCT compressed stereopairs over the monoscopic single-views even though the blocking artifacts are annoying. In [69], depth naturalness is shown to be a useful ingredient in the assessment of 3D video QoE. Similarly, in [56], the added value of depth naturalness has been verified for pristine and blurred stereoscopic images. In [70], stimuli with various stereo depth and image quality were evaluated subjectively in terms of naturalness, viewing experience, image quality, and depth perception, and the experimental results suggested that the overall 3D QoE is approximately 75% determined by image quality and 25% by perceived depth. In [24], Chen *et al.* showed that subjective evaluation of depth quality has a low correlation with that of 3D image quality and verified that the overall 3D QoE can be predicted using a single linear model from 3D image quality and depth quality.

Meanwhile, several studies have been proposed to objectively predict perceived depth quality and subsequently to predict 3D quality by combining depth quality and 2D image quality. In [166], PSNR, SSIM [153] and VQM [102] were employed to predict perceived depth quality, and PSNR and SSIM appear to have slightly better performance. In [10, 11], disparity maps between left- and right-views were estimated, followed by 2D quality assessment of disparity quality using SSIM and C4 [21], which was subsequently combined with 2D image quality to produce an overall 3D image quality score. The results suggested that C4 outperforms SSIM on evaluating stereoscopic image pairs and disparity maps on IRCCyN/IVC 3D Image Database [10] and also showed that the 3D-IQA performance of SSIM can be improved when adding depth quality. You *et al.* [168] evaluated stereopairs as well as disparity maps with respect to ten well-known 2D-IQA metrics, i.e., PSNR, SSIM, MS-SSIM [158], UQI [149], VIF [123], Visual Signal-to-noise Ratio (VSNR) [22], etc. The results suggested that an improved performance can be achieved when stereo image quality and depth quality are combined appropriately. Similarly, Yang *et al.* [164, 163] proposed a 3D-IQA algorithm based on the average PSNR of left- and right-views and the absolute difference with respect to disparity map. In [176], Zhu *et al.* proposed a 3D-VQA model by considering depth perception, and the experimental results showed that the proposed HVS based model performs better than PSNR.

Nevertheless, in [90, 23, 121], comparative studies show that none of these 3D-IQA/VQA models, with depth information involved, perform better than or in most cases, even as good as, direct averaging 2D-IQA measures of both views. In particular, in [90], it was shown that averaging PSNR, SSIM, MS-SSIM, UQI and VIF measurements of left- and right-views performs equally well or better than the advanced 3D-IQA models [10, 168, 164, 176] on LIVE 3D Image Quality Database Phase I. Similar results were also observed in [23], where averaging SSIM and MS-SSIM measurements of both views outperformed advanced 3D-IQA models [10, 168] on LIVE 3D Image Quality Database Phase II. In [121], it was reported that directly averaging MS-SSIM outperformed 3D-IQA models [10, 168] on Ningbo University 3D Image Quality Assessment Database. All these observations suggest that the progress on how to automatically predict depth quality and how to combine 3D image quality and depth quality remains limited. This lack of successful objective QoE methods for 3D visual experience has limited their applications in the development of 3D imaging applications and services.

In this chapter, we carry out two subjective experiments on depth quality. The first one adopts a traditional ACR [146] protocol widely used in general IQA research. We find this approach problematic in this scenario because monocular cues and the spatial quality of images have strong impact on the depth quality scores given by subjects, making it difficult to single out the actual contributions of stereoscopic cues in depth perception. To overcome this problem, we conduct the second subjective study where depth effect is synthesized at different depth levels before various types and levels of symmetric and asymmetric distortions are applied. Instead of following the traditional approach, we ask subjects to identify and label depth polarizations, and a DPDI is developed based on the percentage of correct and incorrect subject judgements. We find the second approach highly effective at quantifying depth perception induced by stereo cues. We then carry out a series of analysis to investigate the impact of image content, distortion type, and distortion symmetricity on perceived depth quality. Furthermore, we propose a novel computational model for DPDI prediction. Our results show that the proposed model, without explicitly identifying image distortion types, leads to highly promising DPDI prediction performance.

4.2 Subjective Study I

4.2.1 Image Database

The Waterloo-IVC 3D Image Quality Database Phase I was created from 6 pristine stereoscopic image pairs and their corresponding single-view images. Each single-view image was altered by three types of distortions: additive white Gaussian noise contamination, Gaussian blur, and JPEG compression, and each distortion type had four distortion levels. The single-view images are employed to generate distorted stereopairs, either symmetrically or asymmetrically. There are totally 78 single-view images and 330 stereoscopic images in the database. More comprehensive descriptions are in Section 3.2. Here we focus on the depth perception part, where the definition of depth quality is the amount, naturalness and clearness of depth perception experience.

4.2.2 Subjective Test

The subjective test was conducted in the Lab for Image and Vision Computing at University of Waterloo. The test environment has no reflecting ceiling walls and floor, and was not insulated by any external audible and visual pollution. An ASUS 27” VG278H 3D LED monitor with NVIDIA 3D Vision^{TM}2 active shutter glasses is used for the test. The default viewing distance was 3.5 times the screen height. In the actual experiment, some subjects did not feel comfortable with the default viewing distance and were allowed to adjust the actual viewing distance around it. The details of viewing conditions are given in Table 3.17. Twenty-four naïve subjects, 14 males and 10 females aged from 22 to 45, participated in the study. A 3D vision test was conducted first to verify their ability to view stereoscopic 3D content. Three of them (1 male, 2 females) failed the vision test and did not continue with the subsequent experiment. As a result, a total of twenty-one subjects proceeded to the formal test.

We followed the ACR protocol and the subjects were asked to rate the depth quality of each image between 0 and 10 pts. A self-training process was employed to help the subjects establishing their own rating strategies with the help of the depth comparison test (stimuli with the same source image similar to what are used in the formal test but different depth levels were presented to help the subjects establish the concept on the amount of depth), and subjects were introduced to build their own rating strategies.

The motivation of introducing a depth comparison test is to help human subjects understand the amount of depth perception for each pristine stereopairs from their own preference and thus let them focus on evaluating the depth quality degraded by different distortions in the following depth quality test. The six pristine stereopairs from Waterloo-IVC 3D image database were utilized in this test. For each pristine stereopair, a single-view image (view 1) was firstly displayed to help the subjects get familiar with image content and then six different stereopairs with an increasing amount of depth were presented, which are S1

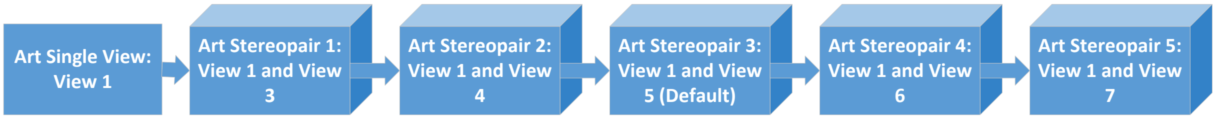


Figure 4.1: Procedure of the depth comparison test: Art Stereopairs.

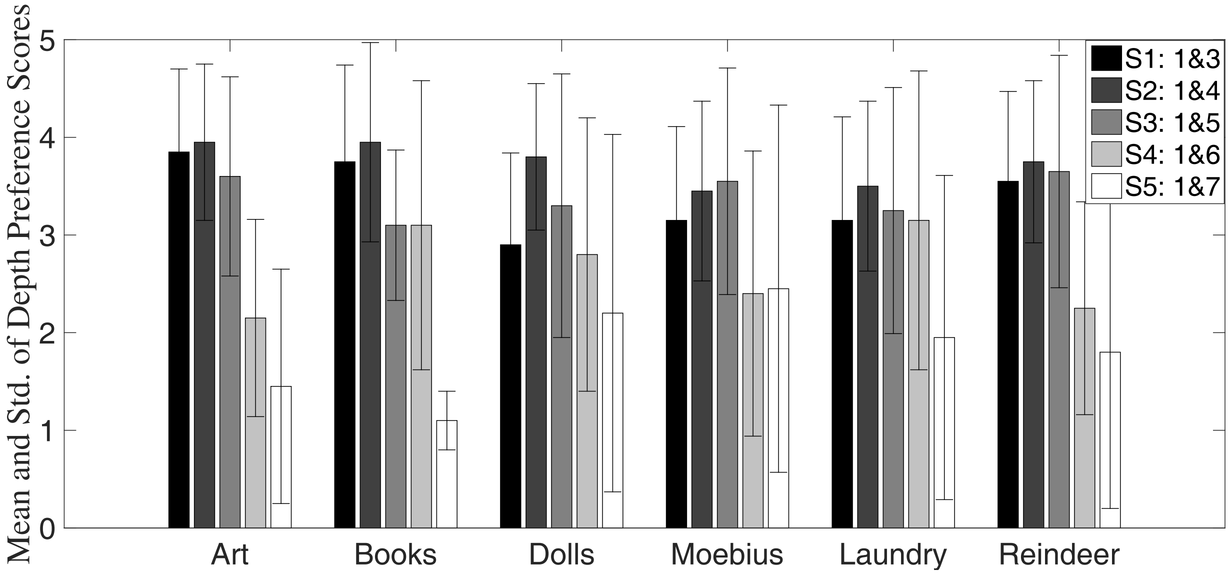


Figure 4.2: Means and standard deviations of depth preference scores in depth comparison test.

(view 1 and view 3), S2 (view 1 and view 4), S3 (view 1 and view 5), S4 (view 1 and view 6) and S5 (view 1 and view 7) as shown in Figure 4.1. Subjects were allowed to compare these six stereopairs back and forth and then to rank them based on their own preference for depth perception. Some subjects favored S5 with the largest amount of depth while others preferred the mid-level S3 as they felt the 3D objects presented in S5 come too close to their faces. The depth preference score is assigned from 1 to 5 pts, for which 1 represents the least preferred and 5 the most preferred. The means and standard deviations of depth preference scores are shown in Figure 4.2, where we observe high variations between subject scores, suggesting diverse subject opinions in depth preference.

Previous works reported that the perception of depth quality are both highly content and texture dependent [120] and subject dependent [26, 24]. Therefore, it is not desirable to over-educate the subjects to use the same given rating strategy. Thus after the depth comparison test, the 3D pristine stereopairs were first presented and the subjects were instructed to give high scores (close to 10 pts) to such images, and the 2D pristine images (with no depth from stereo cues) were presented and the subjects were instructed to give low scores (close to 0 pts). Next, stereopairs of different types/levels of distortions were presented and the subjects were asked to practice by giving their ratings on depth quality between 0 and 10 pts. During this process, the instructor also repeated the definition of depth quality and emphasized that there is not necessarily any correlation between depth quality and the type/level of distortions.

In the formal test, all stimuli were shown once. However, there were 12 repetitions, which means that for each subject, her/his first 12 stereopairs were shown twice. These repeated stimuli are to check whether the subjects perform consistently at the beginning and the end of the experiments, an indicator that reflects the strength of fatigue effect in the experiment. The order of stimuli was randomized and the consecutive testing stereopairs were from different source images. 342 testing stereopairs with 12 repetitions were partitioned into two sessions and each single session (171 stereopairs) was finished in 15 to 20 minutes. Sufficient relaxation periods (5 minutes or more) were given between sessions. Moreover, we found that repeatedly switching between viewing 3D images and grading on a piece of paper or a computer screen is a tiring experience. To overcome this problem, we asked the subject to speak out a score, and a customized graphical user interface on another computer screen was used by the instructor to record the score. All these efforts were intended to reduce visual fatigue and discomfort of the subjects.

4.2.3 Observations and Discussions

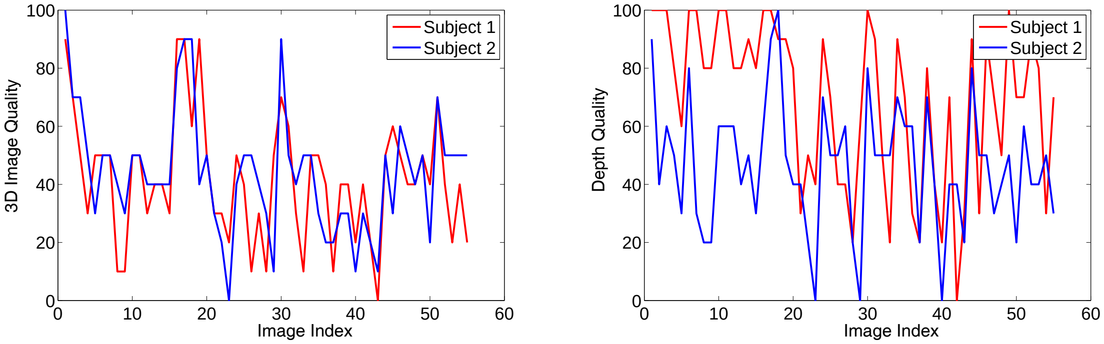
In Section 3.3, the 2D image quality (2DIQ) and 3D image quality (3DIQ) tests on Waterloo-IVC 3D Image Quality Database were introduced. The raw 2DIQ and 3DIQ scores given by each subject were converted to Z-scores, respectively. Then the entire data sets were rescaled to fill the range from 1 to 100 and the mean opinion scores (MOS) for each 2D and 3D image was computed. The detailed observations and analysis of the relationship between MOS 2DIQ and MOS 3DIQ and how to predict the image content quality of a stereoscopic 3D image from that of the 2D single-view images can be found in Section 3.5.2 and Section 3.6.2.

In this work, the raw depth quality (DQ) scores given by each subject were converted to Z-scores. Then the entire data set was rescaled to fill the range from 1 to 100 and the MOS DQ for each image was computed. For each stereopair, the standard deviation of Z-scores represents the degree of variation and the means of these standard deviations are 12.00 for 3DIQ scores and 20.01 for DQ scores, respectively, indicating large variations in DQ scores. Table 4.1 reports PLCC, SRCC and KRCC between individual 3DIQ/DQ scores and MOS 3DIQ/DQ scores, which reflect the degree of agreement of 3DIQ/DQ scores among the subjects. PLCC is adopted to evaluate prediction accuracy [46] and SRCC and KRCC are employed to assess prediction monotonicity [46]. Higher PLCC, SRCC and KRCC indicate better consistency with the average human opinions of quality. PLCC is usually computed after a nonlinear mapping between the subjective and objective scores and the results may be sensitive to the choice of the mapping function. SRCC and KRCC are nonparametric rank order-based correlation metrics, independent of any monotonic nonlinear mapping between subjective and objective scores but do not explicitly estimate the accuracy of quality prediction. From Table 4.1, it can be observed that DQ scores show less correlation with MOS compared with 3DIQ scores. To further understand this, Figure 4.3 shows a comparison of two subjects' 3DIQ and DQ scores on the Art stereopairs. It can be observed that these two subjects exhibit general agreement on 3DIQ scores but behave drastically differently in giving DQ scores.

Table 4.1: Mean and standard deviation of correlations between individual scores and MOS

	PLCC		SRCC		KRCC	
Criterion	Mean	Std.	Mean	Std.	Mean	Std.
3DIQ	0.8045	0.0658	0.7762	0.0726	0.6315	0.0695
DQ	0.7414	0.1410	0.7298	0.1443	0.5838	0.1259

Thus our preliminary analysis shows that there is a large variation between subjects on depth quality scores as different people may have very different perception and/or opinions about perceptual depth quality. The rest of this section will focus on the relationship



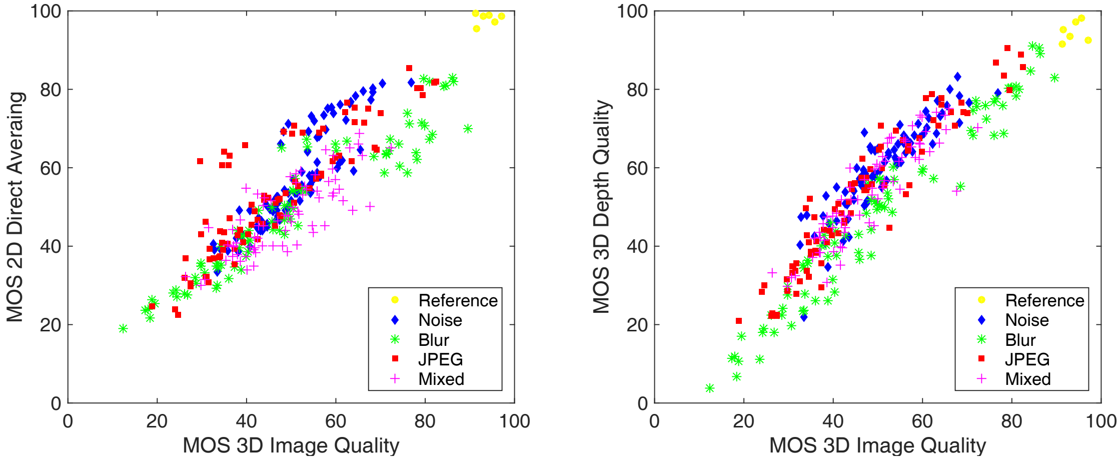
(a) 3D Image Quality (b) Depth Quality

Figure 4.3: Comparison of two subjects' 3DIQ and DQ scores on the Art stereopairs.

between DQ scores and the 3DIQ scores.

Figure 4.4 shows the scatter plots of MOS 3DIQ vs. averaging MOS 2DIQ of left- and right-views and MOS 3DIQ vs. MOS DQ. Figure 4.4 (a) suggests that there exists a strong distortion type dependent prediction bias when predicting quality of asymmetrically distorted stereoscopic images from single-views [144, 145]. Specifically, for noise contamination and JPEG compression, average prediction overestimates 3DIQ (or 3DIQ is more affected by the poorer quality view), while for blur, average prediction often underestimates 3DIQ (or 3DIQ is more affected by the better quality view).

From Figure 4.4 (b), it can be observed that human opinions on 3DIQ and 3D DQ are highly correlated. This is somewhat surprising because 3DIQ and DQ are two different perceptual attributes and the stimuli were generated to cover all combinations between picture qualities and stereo depths. Through more careful observations of the data and discussions with the subjects who did the experiment, we found two explanations. First, psychologically humans have the tendency to give high DQ scores whenever the 3DIQ is good and vice versa, and the strength of such a tendency varies between subjects. Second, humans interpret depth information using many physiological and psychological cues [85], including not only binocular cues such as stereopsis, but also monocular cues such as retinal image size, linear perspective, texture gradient, overlapping, aerial perspective, and shadowing and shading [82, 95]. In the real world, humans automatically use all available



(a) MOS 3DIQ vs. MOS 2DIQ

(b) MOS 3DIQ vs. MOS 3D DQ

Figure 4.4: Relationships between (3DIQ and 2DIQ) and (3DIQ and 3D DQ) in Subjective Study I.

depth cues to determine distances between objects but most often rely on psychological monocular cues. Therefore, the DQ scores obtained in the current study are a combined result from many monocular and binocular cues, and it becomes difficult to gauge the role of stereopsis.

However, what we are interested in the current study is to measure how much stereo information can help with depth perception. Based on the explanations above, in traditional ways of subjective testing like the current one, many depth cues are mixed together and the results are further altered by the spatial quality of the image, making it difficult to quantify the real contributions of using stereoscopic images in depth perception. This inspires us to design a novel depth perception test, which will be presented in the next section.

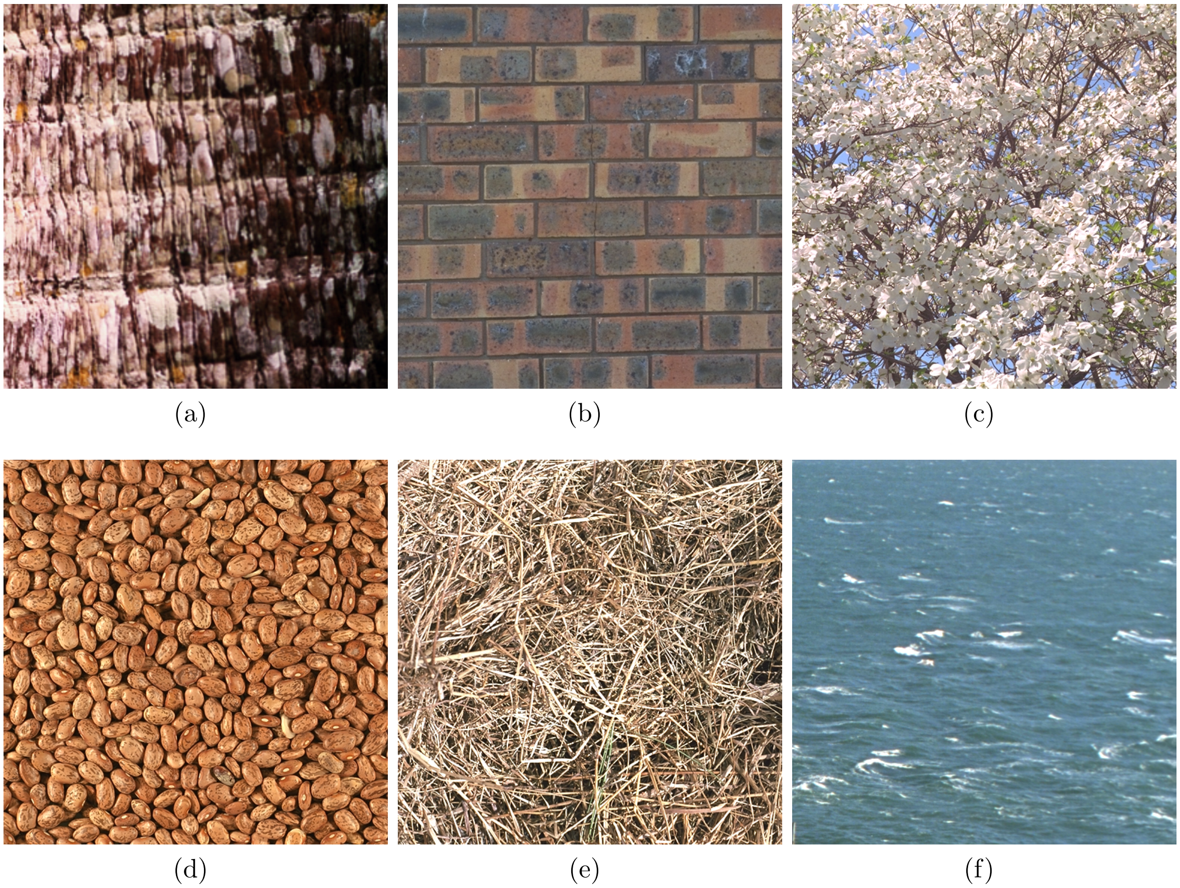
4.3 Subjective Study II

4.3.1 Image Database

We created a new Waterloo-IVC 3D Depth Quality Database from 6 pristine texture images (Bark, Brick, Flowers, Food, Grass and Water) as shown in Figure 4.5. All images were collected from the VisTex Database at MIT Media Laboratory [2]. A stereogram can be build by duplicating the image, selecting a region in one image, and shifting this region horizontally by a small amount in the other one. The region seems to virtually fly in front of the screen, or be behind the screen if the two views are swapped. In our experiment, six different levels of Gaussian surfaces (with different heights and different widths) were obtained by translating and scaling Gaussian profiles, where the 6 depth levels (where Depth 1 and Depth 6 denote the lowest and highest depths, respectively) were selected to ensure a good perceptual separation. Thus each texture image was used to generate 6 stereopairs with different depth levels. By switching left- and right-views, the hidden depth could be perceived towards inside or outside and we denote them as inner stereopairs and outer stereopairs, respectively. As such, for each texture image, we have 12 pristine stereopairs with different depth polarizations and depth levels. In addition, one flat stereopair without any hidden depth information is also included.

Each pristine stereopair (inner, outer and flat) was altered by three types of distortions: additive white Gaussian noise contamination, Gaussian blur, and JPEG compression. Each distortion type had four distortion levels as reported in Table 4.2 and Table 4.3, where the distortion control parameters were decided to ensure a good perceptual separation. PSNR and IW-SSIM evaluations of the simulated distorted single-view 2D images are reported in Table 4.4 and Table 4.5, respectively.

The distortions were simulated either symmetrically or asymmetrically. Symmetrically distorted stereopairs have the same distortion type and level on both views while asymmetrically distorted ones have the distortion on one view only. Altogether, there are 72 pristine stereoscopic images and 1728 distorted stereoscopic images (864 symmetrical and 864 asymmetrical distortions) in the database. In terms of the depth polarity, there are



(a) (b) (c)

(d) (e) (f)

Figure 4.5: The texture images used in Subjective Study II. (a) Bark. (b) Brick. (c) Flower. (d) Food. (e) Grass. (f) Water.

Table 4.2: Value ranges of control parameters to generate image distortions for Waterloo-IVC 3D Depth Quality Database

Distortion	Control Parameter	Range
White Noise	Variance of Gaussian	[0.11 1.12]
Gaussian Blur	Variance of Gaussian	[1.50 11.00]
JPEG Compression	Quality parameter	[1 10]

Table 4.3: Values of control parameters to generate image distortions for Waterloo-IVC 3D Depth Quality Database

Bark						
White Noise	σ	**Gaussian Blur**	σ	**Width**	**JPEG**	**Quality**
W1	0.32	G1	2.20	15	J1	10
W2	0.46	G2	3.20	21	J2	6
W3	0.63	G3	4.20	27	J3	3
W4	0.83	G4	5.70	33	J4	1
Brick						
White Noise	σ	**Gaussian Blur**	σ	**Width**	**JPEG**	**Quality**
W1	0.14	G1	2.40	15	J1	10
W2	0.20	G2	3.80	25	J2	6
W3	0.29	G3	5.60	35	J3	3
W4	0.39	G4	8.60	47	J4	1
Flower						
White Noise	σ	**Gaussian Blur**	σ	**Width**	**JPEG**	**Quality**
W1	0.41	G1	1.60	11	J1	10
W2	0.55	G2	2.30	15	J2	6
W3	0.71	G3	3.20	21	J3	3
W4	0.91	G4	4.30	27	J4	1
Food						
White Noise	σ	**Gaussian Blur**	σ	**Width**	**JPEG**	**Quality**
W1	0.41	G1	1.80	13	J1	10
W2	0.56	G2	2.70	17	J2	6
W3	0.74	G3	3.70	25	J3	3
W4	0.95	G4	4.90	31	J4	1
Grass						
White Noise	σ	**Gaussian Blur**	σ	**Width**	**JPEG**	**Quality**
W1	0.52	G1	1.50	11	J1	10
W2	0.69	G2	2.10	13	J2	6
W3	0.87	G3	2.90	21	J3	3
W4	1.12	G4	3.70	25	J4	1
Water						
White Noise	σ	**Gaussian Blur**	σ	**Width**	**JPEG**	**Quality**
W1	0.11	G1	2.70	19	J1	10
W2	0.16	G2	4.20	27	J2	6
W3	0.22	G3	6.40	41	J3	3
W4	0.30	G4	11.00	61	J4	1

Table 4.4: PSNR evaluations of distorted 2D images on Waterloo-IVC 3D Depth Quality Database

White Noise				
LEVEL	**W1**	**W2**	**W3**	**W4**
Bark	14.95	12.72	11.22	10.21
Brick	20.60	17.71	15.10	13.46
Flower	13.27	11.84	10.85	10.12
Food	13.37	11.92	10.94	10.28
Grass	11.98	10.70	9.90	9.21
Water	22.68	19.46	16.91	14.81
Gaussian Blur				
LEVEL	**G1**	**G2**	**G3**	**G4**
Bark	22.58	20.18	18.83	17.61
Brick	28.72	26.73	25.47	24.41
Flower	18.26	17.08	16.31	15.80
Food	19.34	17.69	16.57	15.71
Grass	14.05	12.99	12.32	11.96
Water	28.66	26.54	25.08	23.80
JPEG Compression				
LEVEL	**J1**	**J2**	**J3**	**J4**
Bark	28.31	25.80	22.35	21.39
Brick	30.50	28.48	25.47	25.13
Flower	21.08	19.58	17.59	17.13
Food	22.52	20.99	18.93	18.38
Grass	18.17	16.90	15.29	14.94
Water	32.07	29.58	27.07	26.78

Table 4.5: IW-SSIM evaluations of distorted 2D images on Waterloo-IVC 3D Depth Quality Database

White Noise				
LEVEL	**W1**	**W2**	**W3**	**W4**
Bark	0.8022	0.7060	0.6016	0.5015
Brick	0.8073	0.7132	0.6010	0.5083
Flower	0.8003	0.7060	0.6088	0.5113
Food	0.8048	0.7077	0.6054	0.5095
Grass	0.8047	0.7011	0.6062	0.5022
Water	0.8108	0.7047	0.6013	0.5027
Gaussian Blur				
LEVEL	**G1**	**G2**	**G3**	**G4**
Bark	0.8540	0.6994	0.5596	0.4064
Brick	0.8557	0.7038	0.5575	0.4079
Flower	0.8499	0.7135	0.5536	0.4064
Food	0.8522	0.7082	0.5523	0.4017
Grass	0.8458	0.7099	0.5404	0.4081
Water	0.8479	0.7041	0.5534	0.3993
JPEG Compression				
LEVEL	**J1**	**J2**	**J3**	**J4**
Bark	0.9722	0.9478	0.8781	0.8451
Brick	0.9171	0.8504	0.6935	0.6543
Flower	0.9648	0.9354	0.8474	0.8115
Food	0.9655	0.9389	0.8718	0.8469
Grass	0.9779	0.9613	0.9170	0.9018
Water	0.8974	0.8080	0.6647	0.6441

684 inner stereopairs, 684 outer stereopairs and 432 flat stereopairs. An example of the procedure of generating a symmetrically blurred stereopair is shown in Figure 4.6.

For each image, we provide the subjects with four available choices to respond, i.e., inner, outer, flat and unable to decide. The motivation of introducing the last choice is that for some distorted stereopairs, the subjects can perceive the existence of depth information but feel difficult to make confident judgements on depth polarity.

There are three important features of the current database that distinguish it from others. First, the depth information embedded in each stereopair is independent of its 2D scene contents, such that subjects can only make use of stereo cues to identify depth change and judge the polarity of depth. Second, the database contains distorted stereopairs from various distortion types, allowing us to compare the impacts of different distortions on depth perception. Third, the current database contains both symmetrically and asymmetrically distorted stereopairs, which allows us to directly examine the impact of asymmetric distortions on depth perception. This may also help us better understand what are the key factors that affect depth quality in stereoscopic images.

4.3.2 Subjective Test

The subjective test was conducted in the Lab for Image and Vision Computing at University of Waterloo with the same test environment, the same 3D display system, and the same viewing conditions as described in Section 4.2. Thus here we only describe some important differences from Subjective Study I. Twenty-two naive subjects, 11 males and 11 females aged from 21 to 34, participated in the study and no one failed the vision test. As a result, a total of twenty-two subjects proceeded to the formal test. The training process is fairly straightforward. Twelve stereopairs with different depth configurations including polarities and levels were presented to the subjects. Subjects were asked to speak out their judgements for these training stereopairs as an exercise. Then a multi-stimulus method was adopted to obtain subjective judgements for all test stereopairs. Each stimulus contains six stereopairs with the same depth level and the same image content but different depth polarity or image distortion. All stimuli were shown once and the order of stimuli was randomized. 75 stimuli were evaluated in one session and each session was controlled to

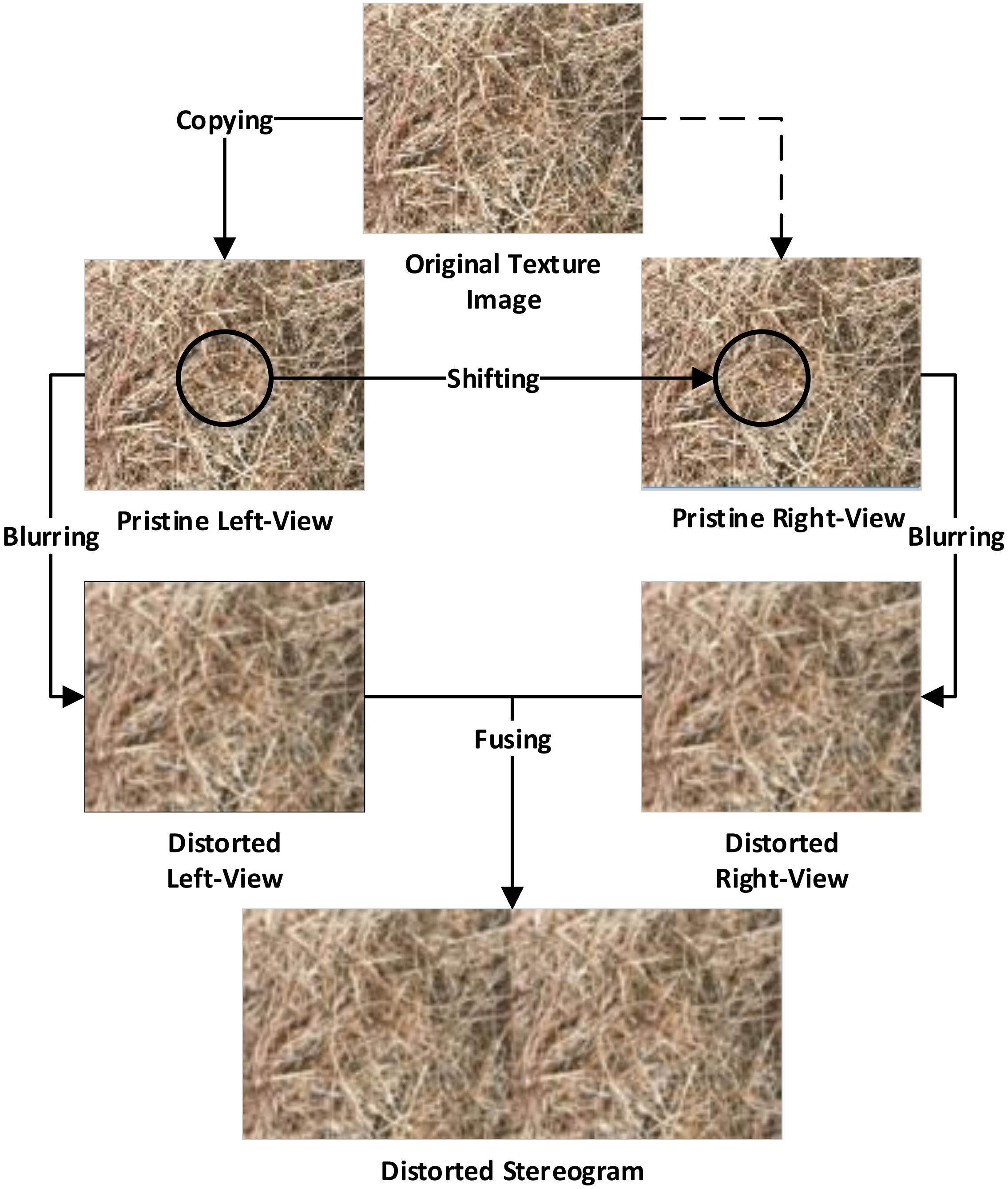


Figure 4.6: Procedure of generating a symmetrically blurred stereoscopic image in Subjective Study II.

be within 20 minutes. Similarly, subjects only needed to speak out their judgements and an instructor was responsible for recording subjective results.

We observe a significant variation between subjects' behaviors, which is expected as humans exhibit a wide variety of stereoacuity and stereosense [169]. The rest of this section focuses on the impact of depth level, depth polarity, image content and image distortion. More detailed analysis of the other aspects of the subjective data will be discussed in the future work.

4.3.3 Depth Perception Difficulty Index (DPDI)

For each test image, there are 3 possible ground-truth polarity answers - inner, outer and flat. Meanwhile, pooling the subjective judgements on the image leads us to four percentage values, denoted by $\{P_{\text{in}}, P_{\text{out}}, P_{\text{flat}}, P_{\text{unable}}\}$, corresponding to the percentages of subject judgements of inner, outer, flat and unable to decide, respectively, and $P_{\text{in}} + P_{\text{out}} + P_{\text{flat}} + P_{\text{unable}} = 1$. Given these values, we define a novel measure named Depth Perception Difficulty Index (DPDI), which indicates how difficult it is for an average subject to correctly perceive the depth information in the image. Specially, if the ground-truth is an inner image, we define

$$\begin{aligned} \text{DPDI} &= \min\{1, P_{\text{flat}} + P_{\text{unable}} + 2P_{\text{out}}\} \\ &= 1 - \max\{0, P_{\text{in}} - P_{\text{out}}\}. \end{aligned} \tag{4.1}$$

Similarly, for an outer image

$$\begin{aligned} \text{DPDI} &= \min\{1, P_{\text{flat}} + P_{\text{unable}} + 2P_{\text{in}}\} \\ &= 1 - \max\{0, P_{\text{out}} - P_{\text{in}}\}. \end{aligned} \tag{4.2}$$

This DPDI is bounded between 0 and 1. The values of DPDI in some extreme cases are as follows: when we have $\{1, 0, 0, 0\}$ for inner images or $\{0, 1, 0, 0\}$ for outer images, DPDI equals 0; when we have $\{0.25, 0.25, 0.25, 0.25\}$, which is equivalent to the case of random guess, DPDI equals 1.

4.3.4 Analysis and Key Observations

Table 4.6 shows the mean DPDI values for different depth levels for the cases of all images, inner images and outer images. Unsurprisingly, DPDI drops with increasing depth in each test group. A much more interesting observation here is that with a given level of depth, inner images generally have lower DPDI values and the difference in mean DPDI values between inner and outer images increase with the level of depth. This indicates that it is easier for humans to perceive depth information when objects appear to be behind the screen than in the opposite case.

Table 4.6: Average DPDI values of different depth levels

Depth Levels	Inner	Outer	Outer − Inner	All
Level 1	0.9196	0.9146	-0.0050	0.9171
Level 2	0.7605	0.7883	0.0278	0.7744
Level 3	0.5829	0.6721	0.0892	0.6275
Level 4	0.4095	0.5732	0.1637	0.4914
Level 5	0.3409	0.5008	0.1599	0.4209
Level 6	0.2811	0.4474	0.1663	0.3643

Table 4.7 reports the mean DPDI values for different background image contents. First, it appears that DPDI is highly image content dependent as it varies significantly across content. In general, DPDI decreases with the increase of high-frequency details, which is consistent with the previous vision research [117] that stereo gain is higher for the high spatial-frequency system than the low spatial-frequency system. Second, although inner images always have higher DPDI values, the gap between inner and outer images is image content dependent.

Table 4.8 shows the mean DPDI values of different distortion types and levels. First, across distortion types, noise contamination has more impact on depth perception than JPEG compression and Gaussian blur. Second, more interestingly, although the cases of symmetric distortions double the total amount of distortions than asymmetric distortions (because the same level of distortions is added to both views), the DPDI gap between asymmetric and symmetric distortions is distortion type dependent. The gaps in the case

Table 4.7: Average DPDI values of different image contents

Image Contents	Inner	Outer	Outer − Inner	All
Bark	0.4831	0.5793	0.0962	0.5339
Brick	0.7562	0.9226	0.1664	0.8209
Flower	0.4232	0.4985	0.0753	0.4545
Food	0.4948	0.6007	0.1059	0.5448
Grass	0.2646	0.4315	0.1669	0.3620
Water	0.8712	0.8846	0.0134	0.8794

of noise contamination is much higher than those of Gaussian blur and JPEG compression. The point worth noting is that adding blur or JPEG compression to one view of stereopair results in similar difficulty in depth perception as adding the same level of distortion to both views. This is quite different from the distortion type dependency in 3D image quality perception, as shown in Figure 4.4 (a). It is interesting to note that some of our new observations are somehow implicitly consistent with previous vision studies [44, 49]. For example, in [49], Hess *et al.* found that stereoacuity was reduced when one view was severely blurred by filtering off high spatial frequencies and loss of acuity was much less severe when both views are blurred.

Table 4.8: Average DPDI values of different distortion types and levels

Distortions	All	Level 1	Level 2	Level 3	Level 4
Noise Sym.	0.7986	0.6275	0.7412	0.8838	0.9419
Noise Asym.	0.6504	0.5215	0.6477	0.7058	0.7500
Blur Sym.	0.5470	0.4962	0.4912	0.5492	0.6515
Blur Asym.	0.5431	0.3902	0.4975	0.6048	0.6528
JPEG Sym.	0.5660	0.4444	0.5278	0.6187	0.6730
JPEG Asym.	0.5473	0.4470	0.5265	0.5871	0.6679

4.3.5 Impact of Eye Dominance

Eye dominance is a common visual phenomenon, referring to the tendency to prefer the input from one eye to the other, depending on the human subject [67]. When studying visual quality of asymmetrically distorted images, it is important to understand if eye dominance plays a significant role in the subjective test results. For this purpose, we carried out a separate analysis on the impact of eye dominance in the depth perception of asymmetrically distorted stereoscopic images. The side of the dominant eye under static conditions was checked first by Rosenbach's test [107]. This test examines which eye determines the position of a finger when the subject is asked to point to an object. Among twenty subjects who finished the formal test Subjective Study II, ten subjects (6 males, 4 females) had a dominant left eye, and the others (5 males, 7 females) are right-eye dominant.

The DPDI for each image in Waterloo-IVC 3D Depth Quality Database were computed for left-eye dominant subjects and right-eye dominant subjects, denoted as $\mathrm{DPDI_L}$ and $\mathrm{DPDI_R}$, respectively. We employed the one-sample t-test to obtain a test decision for the null hypothesis that the difference between $\mathrm{DPDI_L}$ and $\mathrm{DPDI_R}$, i.e., $\mathrm{DPDI_D} = \mathrm{DPDI_L} - \mathrm{DPDI_R}$, comes from a normal distribution of zero-mean and unknown variance. The alternative hypothesis is that the population distribution does not have a mean equaling zero. The result h is 1 if the test rejects the null hypothesis at the 5% significance level, and 0 otherwise. The returned p-values for symmetric and asymmetric images are 0.3448 and 0.3048, respectively, thus the null hypothesis cannot be rejected at the 5% significance level, which indicates that the impact of eye dominance in the perception of depth quality of asymmetrically distorted stereoscopic images is not significant.

It is worth noting that in Section 3.4 we found that the eye dominance effect does not have strong impact on the perceived image content quality of stereoscopic images. Our two observations are consistent with the "stimulus" view of rivalry that is widely accepted in the field of visual neuroscience [15]. A comprehensive review and discussion on "stimulus" rivalry versus "eye" rivalry can be found in [15, 81].

4.4 Objective Study: Prediction Of Depth Perception Difficulty Index

4.4.1 DPDI Prediction Model

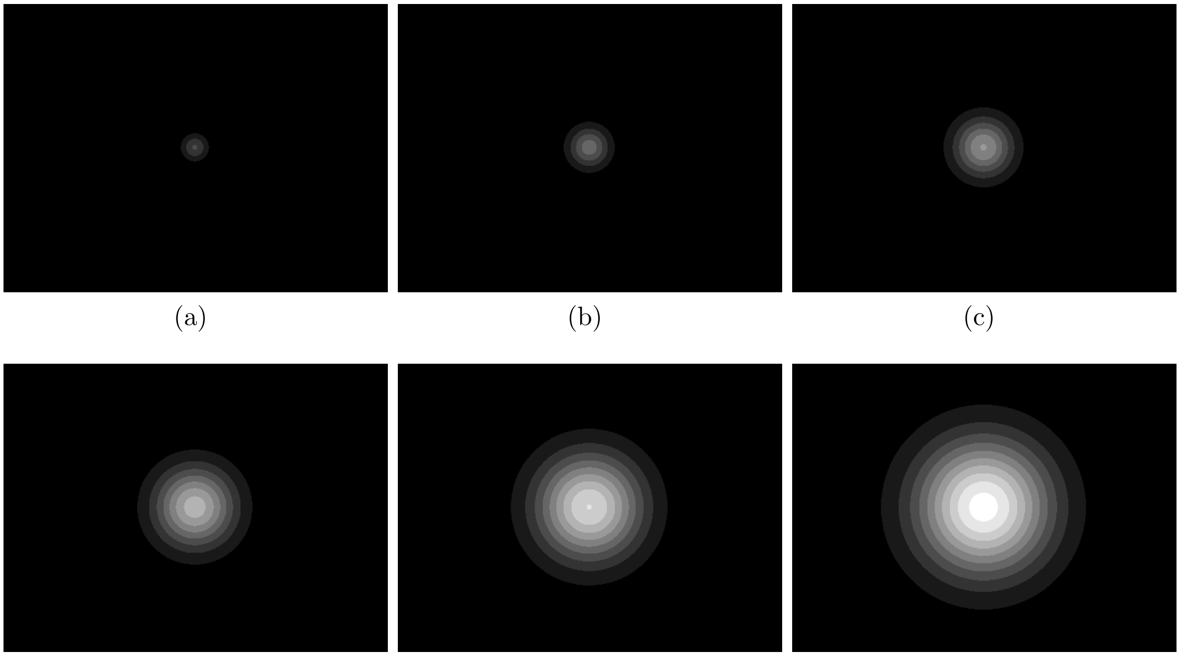
We opt to use a multiple-stage approach in the design of an objective DPDI predictor. The first stage aims to predict the DPDI for different depth levels $H_{L(evel)}$ and image contents $H_{C(ontent)}$, while in the second stage, a patch-structure representation is developed to predict the DPDI for different distortion types and levels $H_{D(istortion)}$. Finally, these components are combined to yield an overall DPDI prediction model.

In Section 4.3.4, DPDI is found to increase with the depth level monotonically. Here we look for an efficient approach to predict DPDI values of different levels using stereo matching, which is an active research area in computer vision over the last few decades [115]. Specially, given a stereopair of $\mathbf{x_l}$ and $\mathbf{x_r}$ for the left-view and right-view reference images, respectively, we first estimate the disparity map $\mathbf{D}_{lr}$, which is simply done by using MATLAB®'s utility `disparityMap` [52]. The ground truth and the estimated disparity maps for different depth levels are shown in Figure 4.7 and Figure 4.8, respectively. The mean values of the ground truth and the estimation errors of the disparity maps are reported in Table 4.9, where it can be seen that the estimations on the disparities are accurate, allowing us to design a simple approach to predict how DPDI changes with depth levels. We denote $\mu_{\mathbf{D}_{lr}}$ as the mean of disparity values and apply a nonlinear mapping on $\mu_{\mathbf{D}_{lr}}$ to predict the DPDI values of different depth levels:

$$H_L = \frac{\alpha}{(\mu_{\mathbf{D}_{lr}})^{\beta} + \gamma}\,, \tag{4.3}$$

where the best parameters are found to be $\alpha = 0.4$, $\beta = 1$ and $\gamma = 0.47$.

In Section 4.3.4, we found that DPDI is highly image content dependent as it varies significantly across content. In general, DPDI decreases with the increase of high-frequency details or energy. We measure the energy by computing the local variances at each spatial location, i.e., the variances of local image patches extracted around each spatial location,

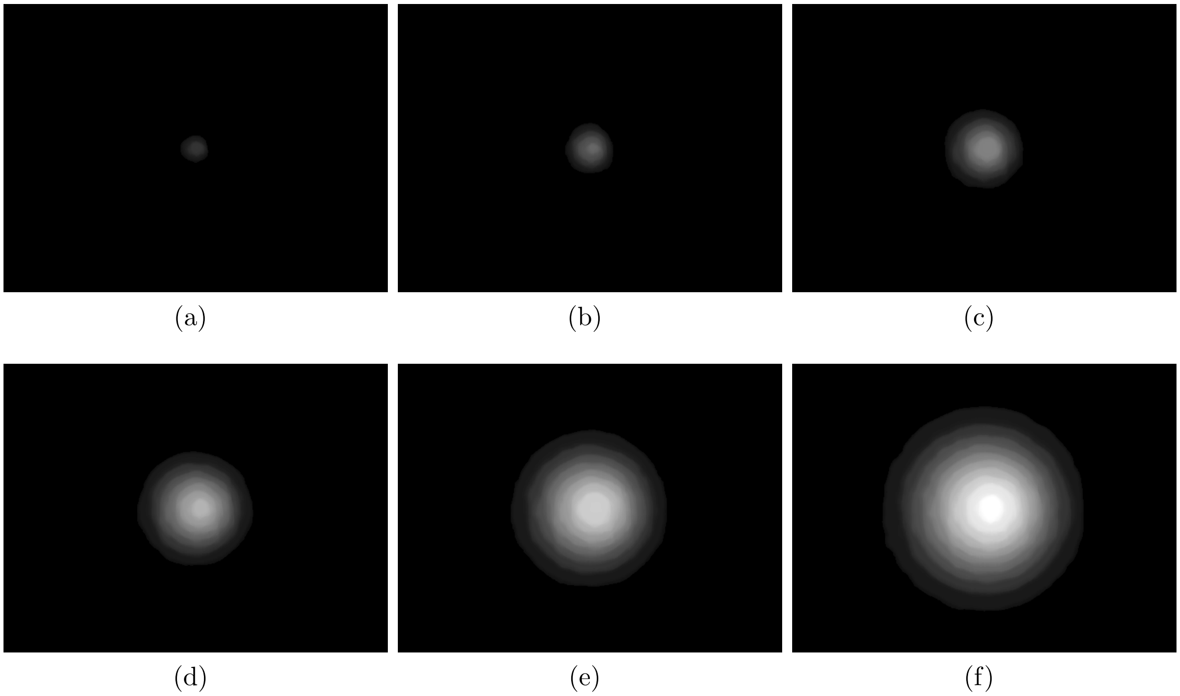

(a) (b) (c)

(d) (e) (f)

Figure 4.7: The ground truth disparity maps for different depth levels from the pristine Bark stereopair used in Subjective Study II. Brighter pixels represent higher disparity values. (a) Level 1. (b) Level 2. (c) Level 3. (d) Level 4. (e) Level 5. (f) Level 6.

Table 4.9: Ground truth (G.T.) and mean estimation errors of disparity maps for different image contents

Levels	**G. T.**	Bark	Brick	Flowers	Food	Grass	Water
Depth 1	**0.0080**	0.0012	0.0020	0.0015	0.0017	0.0016	0.0022
Depth 2	**0.0349**	0.0031	0.0030	0.0024	0.0030	0.0026	0.0038
Depth 3	**0.1045**	0.0034	0.0034	0.0036	0.0035	0.0032	0.0063
Depth 4	**0.2553**	0.0045	0.0038	0.0040	0.0045	0.0037	0.0070
Depth 5	**0.5402**	0.0042	0.0044	0.0038	0.0045	0.0021	0.0075
Depth 6	**1.0370**	0.0052	0.0037	0.0026	0.0032	0.0319	0.0081


(a) (b) (c)

(d) (e) (f)

Figure 4.8: The estimated disparity maps for different depth levels from the pristine Bark stereopair used in Subjective Study II. Brighter pixels represent higher disparity values. (a) Level 1. (b) Level 2. (c) Level 3. (d) Level 4. (e) Level 5. (f) Level 6.

for which an 11×11 circular-symmetric Gaussian weighting function $\mathbf{w} = \{w_i | i = 1, 2, \cdot, N\}$ with standard deviation of 1.5 samples, normalized to unit sum ($\sum_{i=1}^{N} w_i = 1$), is employed. The mean of local variances is used to measure the energy E. Empirically, we observe that E shows a high dependency with DPDI for different image contents at different levels of complexity. The relationship can be well accounted for by the following nonlinear mapping:

$$H_C = \frac{\tau}{\log(E^{\lambda})}\,, \tag{4.4}$$

where the best parameters are found to be $\tau = 21.9$ and $\lambda = 6$.

Any image patch can be represented in a unique and adaptive way by three conceptually independent components: mean intensity, signal strength and signal structure [78]. This novel representation has been found to be useful in IQA of multi-exposure image fusion [79] and contrast changed images [147]. In this work, we show that this representation can well explain the distortion type dependency observations we described in Section 4.3.4.

Given a $\sqrt{N} \times \sqrt{N}$ local image patch $\mathbf{x}$ that is represented as an N-dimensional vector, we decompose it by

$$\begin{aligned} \mathbf{x} &= \mu_{\mathbf{x}} + ||\mathbf{x} - \mu_{\mathbf{x}}|| \cdot \frac{\mathbf{x} - \mu_{\mathbf{x}}}{||\mathbf{x} - \mu_{\mathbf{x}}||} \\ &= c_1^{\mathbf{x}} \cdot \mathbf{v}_1^{\mathbf{x}} + c_2^{\mathbf{x}} \cdot \mathbf{v}_2^{\mathbf{x}}\,, \end{aligned} \tag{4.5}$$

where $||\cdot||$ denotes the l^2 norm of a vector, $\mu_{\mathbf{x}}$ is the mean intensity of the patch. $\mathbf{x}$ is now represented as a linear combination of two unit-length vectors,

$$\mathbf{v}_1^{\mathbf{x}} = \frac{1}{\sqrt{N}} \cdot \mathbf{1} \qquad \text{and} \qquad \mathbf{v}_2^{\mathbf{x}} = \frac{\mathbf{x} - \mu_{\mathbf{x}}}{||\mathbf{x} - \mu_{\mathbf{x}}||}\,, \tag{4.6}$$

each associated with a coefficient

$$c_1^{\mathbf{x}} = \sqrt{N}\mu_{\mathbf{x}} \qquad \text{and} \qquad c_2^{\mathbf{x}} = ||\mathbf{x} - \mu_{\mathbf{x}}||\,, \tag{4.7}$$

here $\mathbf{1}$ denotes a column vector with all entries equaling 1. Since $\mathbf{v}_1^{\mathbf{x}}$ is fixed, each source patch $\mathbf{x}$ can be uniquely represented by three components $c_1^{\mathbf{x}}$, $c_2^{\mathbf{x}}$ and the unit-length vector

$\mathbf{v}_2^{\mathbf{x}}$, which denote the mean intensity, signal strength and signal structure, respectively. The representation or decomposition is adaptive, where the basis $\mathbf{v}_2^{\mathbf{x}}$ points to a specific direction in the signal space and is adapted to the input signal.

Now assume $\mathbf{x}$ and $\mathbf{y}$ are the co-located patches in the reference and distorted images, respectively. Then from Equation (4.5), we have $\mathbf{v}_2^{\mathbf{x}}$ and $\mathbf{v}_2^{\mathbf{y}}$, which represent the signal structures of the reference and distorted images, respectively. We denote the angle between the orientations of $\mathbf{v}_2^{\mathbf{x}}$ and the structural distortion vector $(\mathbf{v}_2^{\mathbf{y}} - \mathbf{v}_2^{\mathbf{x}})$ as $\Delta\theta$. Then $\cos\Delta\theta$ can be computed as

$$\cos\Delta\theta = \frac{|\mathbf{v}_2^{\mathbf{x}} \cdot (\mathbf{v}_2^{\mathbf{y}} - \mathbf{v}_2^{\mathbf{x}})|}{||\mathbf{v}_2^{\mathbf{x}}||||\mathbf{v}_2^{\mathbf{y}} - \mathbf{v}_2^{\mathbf{x}}||}\,, \tag{4.8}$$

and $\Delta\theta$ can be subsequently obtained through an arc-cosine function. Note that $\Delta\theta$ is the angle between two orientations and thus has a dynamic range between 0 and $\frac{\pi}{2}$. Table 4.10 reports the mean values of $\Delta\theta$ for each distortion type and level. Interestingly, the results show a strong distortion type dependency of $\Delta\theta$. In particular, for noise contaminated image, $\Delta\theta$ is close to $\frac{\pi}{2}$ (90°); for blurred image, $\Delta\theta$ is below $\frac{\pi}{6}$ (30°); and for JPEG compressed image, $\Delta\theta$ typically lies between $\frac{\pi}{4}$ (45°) and $\frac{\pi}{3}$ (60°).

Table 4.10: Mean Values of $\Delta\theta$ for different distortion types and levels

Distortion	Noise	Blur	JPEG
Level 1	83.27°	24.05°	62.87°
Level 2	81.98°	19.00°	58.02°
Level 3	80.95°	15.44°	50.74°
Level 4	79.96°	12.71°	48.59°

Some qualitative explanations of this phenomenon are as follows. When left- and right-views are both noise contaminated, the distortion vectors $\mathbf{v}_2^{\mathbf{y}} - \mathbf{v}_2^{\mathbf{x}}$ are orthogonal to the original vectors $\mathbf{v}_2^{\mathbf{x}}$, thus the original necessary information used to establish stereoscopic cues is affected by independent noise only. In this case, the impact of distortion on the depth quality is additive. As such, the gap of DPDI between noise added to one-view and two-views is much higher than those of Gaussian blur and JPEG compression because twice amount of noise is added. On the other hand, when left- and right-views are either blurred or JPEG compressed, the distortion vectors $\mathbf{v}_2^{\mathbf{y}} - \mathbf{v}_2^{\mathbf{x}}$ can be decomposed into two orthogonal components, one of which aligns with and the other is orthogonal to the original vector. The original necessary information used to establish stereoscopic cues is affected by not only the relative strength of these two components, but also the consistency of such relative strengths on the left- and right-views. When deterministic distortions such as blur or JPEG compression are applied equally to both views, high consistency is expected. In this case, the impact of blurriness or compression artifacts on the depth quality is more dependent on the lower quality view with more structural distortions. As such, the gap of DPDI between one-view and two-views is reduced.

The above analysis shows that this patch-structure representation provides useful cues to account for the distortion type dependency we observed in Section 4.3.4. This inspires us to develop an objective model to automatically predict DPDI for different distortion types and levels.

Let $(\mathbf{x_l}, \mathbf{y_l})$ and $(\mathbf{x_r}, \mathbf{y_r})$ be the co-located patches in the reference and distorted left- and right-views images, respectively. Let d_l and d_r denote the local distortion measures for $(\mathbf{x_l}, \mathbf{y_l})$ and $(\mathbf{x_r}, \mathbf{y_r})$, respectively. We characterize the local measure of DPDI by

$$H_D = \left(|d_l|^p + |d_r|^p\right)^{\frac{1}{p}} \ , \tag{4.9}$$

where $p \geq 0$ is an exponent parameter. With various choices of the value of p, this general formulation leads to a family of combination rules with different physical meanings. The larger the p value, the more emphasis is put on the patches that have relatively larger distortion between left- and right-view. Specifically, $p = 1$ corresponds to length-weighted averaging; $p = 2$ corresponds to energy-weighted averaging; and $p = \infty$ corresponds to picking the patch that has the larger distortion.

It remains to determine the value of p. Instead of fixing p to be a constant for each distortion types and levels, here we propose an automatic approach that chooses p at each spatial location adaptively. From Equation (4.5), we have $(\mathbf{v}_2^{\mathbf{x_l}}, \mathbf{v}_2^{\mathbf{y_l}})$ and $(\mathbf{v}_2^{\mathbf{x_r}}, \mathbf{v}_2^{\mathbf{y_r}})$ for left-view and right-view, respectively. We denote $\Delta\theta_l$ as the angle between $\mathbf{v}_2^{\mathbf{x_l}}$ and $\mathbf{v}_2^{\mathbf{y_l}} - \mathbf{v}_2^{\mathbf{x_l}}$ and $\Delta\theta_r$ as the angle between $\mathbf{v}_2^{\mathbf{x_r}}$ and $\mathbf{v}_2^{\mathbf{y_r}} - \mathbf{v}_2^{\mathbf{x_r}}$. $\cos\Delta\theta_l$ and $\cos\Delta\theta_r$ are computed using

Equation (4.8), then p is determined by

$$p = (1 + \cos\Delta\theta_l + \cos\Delta\theta_r)^2 \,. \tag{4.10}$$

The motivation of using adaptive-p in Equation (4.9) is to automatically adjust the weighting strategies of the left- and right-views for symmetric and asymmetric distortions at different distortion types and levels. In particular, when both views are noise contaminated, $\Delta\theta_l$ and $\Delta\theta_r$ are close to $\frac{\pi}{2}$, $\cos\Delta\theta_l$ and $\cos\Delta\theta_r$ are close to 0, and thus p is close to 1, then we have

$$H_D \simeq (|d_l| + |d_r|) \;; \tag{4.11}$$

when only one view (e.g. left-view) is noise contaminated and the other one (e.g. right-view) is pristine, $\Delta\theta_r$ is 0 and $\cos\Delta\theta_r$ is 1, p goes relatively larger, then we have

$$H_D \simeq \max\{|d_l|, |d_r|\} \;. \tag{4.12}$$

When both views are blurred or JPEG compressed, $\Delta\theta_l$ and $\Delta\theta_r$ are close to 0 or around $\frac{\pi}{4}$, $\cos\Delta\theta_l$ and $\cos\Delta\theta_r$ are close to 1 (for higher JPEG compression levels), thus p is relatively large; when only one view (e.g. left-view) is blurred or JPEG compressed and the other one (e.g. right-view) is pristine, $\Delta\theta_r$ is 0 and $\cos\Delta\theta_r$ is 1, and thus p is also a larger number. In both cases, we have

$$H_D \simeq \max\{|d_l|, |d_r|\} \;. \tag{4.13}$$

As such, the value of p is automatically determined, without recognizing the distortion types explicitly.

Once the value of p is determined at each spatial location, the local H_D measure is computed using Equation (4.9). The global H_D measure is the average of the local H_D across all spatial locations. Finally, the three components, H_L, H_C and H_D are combined to yield an overall DPDI prediction

$$H = H_L \cdot H_C \cdot H_D \,. \tag{4.14}$$

4.4.2 Validation

We use the new Waterloo-IVC 3D Depth Quality Database to test the proposed DPDI prediction model. First, DPDI predictions from H_D only are computed for each depth level and each image content. PLCC, SRCC and KRCC between the observed and the predicted DPDI values are reported in Table 4.11 and Table 4.12, where the results are summarized as the average performance for each image content. The direct averaging (Ave.) method corresponds to the case of $p = 1$ in Equation (4.9), while in the adaptive-p (Adpt.) method the value of p is adaptively determined using Equation (4.10) to Equation (4.13). PSNR, SSIM [153], MS-SSIM [158] and IW-SSIM [154] are employed to create the base single-view distortion measurements d, where we let $d = 50 - \text{PSNR}$ and $d = 1 - \text{SSIM}$ or its derivatives. For fairness, a global approach to compute H_D is adopted, i.e., an average spatial pooling on $\cos \Delta\theta_l$ and $\cos \Delta\theta_r$ is applied to the left- and right-view, respectively, and thus p is determined globally. From Table 4.11, it can be observed that the adaptive-p model outperforms the direct averaging method in almost all cases. In addition, MS-SSIM and IW-SSIM pronounce larger improvements than PSNR and SSIM. Considering the performance and computational complexity, MS-SSIM is chosen as the distortion measurement method in the subsequent tests.

Table 4.11: Performance comparison of DPDI estimations using different base 2D distortion measures. Ave.: direct averaging; Adpt.: adaptive-p model.

PLCC												
	Bark		Brick		Flower		Food		Grass		Water	
2D-IQA	Ave.	Adpt.	Ave.	Adpt.	Ave.	Adpt.	Ave.	Adpt.	Ave.	Adpt.	Ave.	Adpt.
PSNR	0.3770	0.5700	0.1643	0.3330	0.5028	0.7621	0.2673	0.4848	0.3961	0.7177	0.2358	0.3168
SSIM	0.5272	0.6458	0.2937	0.4457	0.5182	0.7420	0.3423	0.5054	0.4638	0.6152	0.2167	0.3275
MS-SSIM	0.6046	0.7365	0.4139	0.6013	0.6435	0.7647	0.4364	0.5575	0.5932	0.6964	0.4426	0.5519
IW-SSIM	0.5921	0.7061	0.5521	0.6230	0.6296	0.7079	0.3647	0.5236	0.5849	0.6639	0.5575	0.5645
SRCC												
2D-IQA	Ave.	Adpt.	Ave.	Adpt.	Ave.	Adpt.	Ave.	Adpt.	Ave.	Adpt.	Ave.	Adpt.
PSNR	0.3925	0.5563	0.1781	0.3029	0.5787	0.7104	0.2834	0.4101	0.5131	0.5942	0.2140	0.1862
SSIM	0.4654	0.6142	0.1753	0.2984	0.4771	0.6037	0.2566	0.3501	0.3319	0.4016	0.2784	0.2253
MS-SSIM	0.5588	0.6587	0.3277	0.4830	0.5785	0.6310	0.3013	0.3679	0.4881	0.5001	0.3924	0.4207
IW-SSIM	0.5133	0.6204	0.3673	0.4840	0.5552	0.6000	0.2637	0.3254	0.4868	0.4978	0.5094	0.5443
KRCC												
2D-IQA	Ave.	Adpt.	Ave.	Adpt.	Ave.	Adpt.	Ave.	Adpt.	Ave.	Adpt.	Ave.	Adpt.
PSNR	0.2969	0.4216	0.1364	0.2184	0.4389	0.5546	0.2104	0.2985	0.3839	0.4468	0.1606	0.1411
SSIM	0.3470	0.4710	0.1360	0.2165	0.3519	0.4563	0.1834	0.2530	0.2497	0.2988	0.2165	0.1790
MS-SSIM	0.4157	0.5017	0.2391	0.3550	0.4328	0.4804	0.2232	0.2725	0.3596	0.3740	0.2955	0.3099
IW-SSIM	0.3779	0.4666	0.2652	0.3559	0.4108	0.4553	0.1912	0.2405	0.3584	0.3712	0.3903	0.4138

Table 4.12: Performance comparison of DPDI estimations using different base 2D distortion measures (Averaged prediction performance from Table 4.11). Ave.: direct averaging; Adpt.: adaptive-p model.

	PLCC		SRCC		KRCC	
2D-IQA	Ave.	Adpt.	Ave.	Adpt.	Ave.	Adpt.
PSNR	0.3239	0.5307	0.3600	0.4600	0.2712	0.3468
SSIM	0.3937	0.5469	0.3308	0.4156	0.2474	0.3124
MS-SSIM	0.5224	**0.6514**	0.4411	**0.5102**	0.3276	**0.3822**
IW-SSIM	0.5468	**0.6315**	0.4493	**0.5120**	0.3323	**0.3839**

Table 4.13 shows PLCC, SRCC and KRCC results for DPDI predictions from all individuals and combinations of H_L, H_C and H_D for all stereopairs and each distortion type. It can be seen that DPDI prediction performance from H_L only and H_C only are similar, which indicates that depth level and image content are about equally important to DPDI estimation, and their combination, not surprisingly, provides a relatively better DPDI prediction performance. It can also be observed that DPDI predictions from H_D only are not as good as those from H_L only or H_C only, even though the adaptive-p method reduces the prediction bias. When H_L, H_C and their combination are combined with H_D, significant improvements are obtained, but in the case of using the adaptive-p method only and not in the case of direct averaging.

The best prediction happens in the case that all DPDI prediction components are included. For all images and each distortion type, the proposed method, when combined with MS-SSIM as the base 2D distortion measure, without attempting to recognize the distortion types or giving any specific treatment for any specific distortion type, leads to highly promising DPDI prediction performance.

Moreover, Table 4.14 and Table 4.15 reports PLCC, SRCC and KRCC values of the overall DPDI predictions for different depth levels and different image contents, respectively. Interestingly, it can be observed that improvements are most pronounced at the middle depth levels (Level 2 and Level 3) or at the textural contents of middle complexity (Bark and Flower), which indicates that the impact of symmetric and asymmetric distortions on the perception of depth is more significant in these “middle” cases.

Table 4.13: Performance comparison of DPDI estimations using different combinations of prediction components

	All			Noise		
Method	PLCC	SRCC	KRCC	PLCC	SRCC	KRCC
H_L	0.5773	0.6160	0.4576	0.6253	0.6686	0.5034
H_C	0.5452	0.5331	0.4087	0.3433	0.3117	0.2338
$H_L + H_C$	0.7278	0.7393	0.5680	0.6840	0.6972	0.5307
H_D for $p = 1$	0.3189	0.3441	0.2390	0.4102	0.3508	0.2471
H_D for adaptive p	0.4077	0.3917	0.2716	0.4040	0.3931	0.2797
$H_L + H_D$ for $p = 1$	0.5885	0.5873	0.4212	0.6933	0.6853	0.5054
$H_L + H_D$ for adaptive p	0.6555	0.6583	0.4810	0.7203	0.7216	0.5403
$H_C + H_D$ for $p = 1$	0.5173	0.5249	0.3724	0.5266	0.5101	0.3691
$H_C + H_D$ for adaptive p	0.6258	0.6299	0.4534	0.5489	0.5297	0.3873
$H_L + H_C + H_D$ for $p = 1$	0.7104	0.7160	0.5327	0.8046	0.7861	0.6072
$H_L + H_C + H_D$ for adaptive p	0.7970	0.8018	0.6190	0.8118	0.7964	0.6199

	Blur			JPEG		
Method	PLCC	SRCC	KRCC	PLCC	SRCC	KRCC
H_L	0.6281	0.6676	0.5051	0.4988	0.5752	0.4291
H_C	0.6200	0.6168	0.4816	0.6832	0.6663	0.5153
$H_L + H_C$	0.7995	0.8072	0.6374	0.7543	0.7616	0.5929
H_D for $p = 1$	0.1913	0.1748	0.1232	0.5022	0.4899	0.3427
H_D for adaptive p	0.2313	0.2307	0.1627	0.6112	0.5853	0.4178
$H_L + H_D$ for $p = 1$	0.4885	0.4952	0.3475	0.6722	0.6755	0.4902
$H_L + H_D$ for adaptive p	0.6160	0.6169	0.4424	0.7366	0.7421	0.5534
$H_C + H_D$ for $p = 1$	0.4392	0.4165	0.2959	0.6326	0.6262	0.4493
$H_C + H_D$ for adaptive p	0.6455	0.6262	0.4580	0.7151	0.6878	0.5090
$H_L + H_C + H_D$ for $p = 1$	0.6698	0.6758	0.4959	0.7713	0.7761	0.5844
$H_L + H_C + H_D$ for adaptive p	0.8205	0.8248	0.6452	0.8415	0.8404	0.6577

Table 4.14: Performance comparison of DPDI Estimations for different depth levels

Depth Levels	Level 1			Level 2			Level 3		
Method	PLCC	SRCC	KRCC	PLCC	SRCC	KRCC	PLCC	SRCC	KRCC
$H_L + H_C + H_D$ for $p = 1$	0.4830	0.5664	0.4342	0.6630	0.6470	0.4805	0.6786	0.6714	0.4959
$H_L + H_C + H_D$ for adaptive p	0.7294	0.6696	0.5280	0.8045	0.8182	0.6398	0.8140	0.8197	0.6379

Depth Levels	Level 4			Level 5			Level 6		
Method	PLCC	SRCC	KRCC	PLCC	SRCC	KRCC	PLCC	SRCC	KRCC
$H_L + H_C + H_D$ for $p = 1$	0.6435	0.6218	0.4485	0.6564	0.6403	0.4586	0.6258	0.5980	0.4292
$H_L + H_C + H_D$ for adaptive p	0.8201	0.7873	0.5914	0.8288	0.7956	0.5926	0.8224	0.7765	0.5791

Table 4.15: Performance comparison of DPDI Estimations for different image contents

Image Contents	Bark			Brick			Flower		
Method	PLCC	SRCC	KRCC	PLCC	SRCC	KRCC	PLCC	SRCC	KRCC
$H_L + H_C + H_D$ for $p = 1$	0.7009	0.7026	0.5190	0.5747	0.5741	0.4198	0.7190	0.7248	0.5392
$H_L + H_C + H_D$ for adaptive p	0.8288	0.8267	0.6382	0.6299	0.6620	0.4933	0.8146	0.8025	0.6169
Image Contents	Food			Grass			Water		
Method	PLCC	SRCC	KRCC	PLCC	SRCC	KRCC	PLCC	SRCC	KRCC
$H_L + H_C + H_D$ for $p = 1$	0.6104	0.6035	0.4384	0.6529	0.6302	0.4610	0.6745	0.5991	0.4589
$H_L + H_C + H_D$ for adaptive p	0.7645	0.7572	0.5664	0.7490	0.7073	0.5312	0.6711	0.6274	0.4930

4.5 Discussions

A main issue with the traditional subjective testing approaches such as that used in our first subjective test (Section 4.2) is the difficulty in singling out the contribution of stereo cues in depth perception, and the subjective scores collected through such experiments show strong correlations between 3D image quality and depth quality scores, even though they are substantially different perceptual attributes. The second subjective testing method introduced in Section 4.3 is an attempt to overcome this problem. To observe it more closely, for each stereopair on the new Waterloo-IVC 3D Depth Database, we estimate its 3DIQ using the binocular rivalry-inspired weighting method presented in [142] and its DPDI using the proposed DPDI prediction model, respectively. Table 4.16 shows the PLCC, SRCC and KRCC values between the predicted 3DIQ and DPDI. It is important to note that the correlations between 3DIQ and DPDI predictions are relatively low, which is quite different from the observations in the first subjective test we discussed in Section 4.2.3. This result suggests that our new subjective testing approach is able to provide more independent information on the depth perception aspect of 3D visual perception.

In Section 4.2.3, we described that there exists a strong distortion type dependency with 3D image quality [142]. Then in Section 4.3.4, a different distortion type dependency in depth perception has been discovered. The discovery of such a distortion type dependency in depth perception not only has scientific values in understanding depth perception in the HVS, but is also desirable in the practice of 3D video compression and transmission. The distortions involved in 3D video coding/communication are not only compression artifacts. The practical encoder/decoder also needs to decide on whether inloop/out-of-loop deblocking filters need to be turned on, and whether mixed-resolutions

Table 4.16: Correlations between 3DIQ and DPDI predictions

Depth Levels	PLCC	SRCC	KRCC
Depth 1	0.5680	0.5512	0.4020
Depth 2	0.5673	0.5496	0.4014
Depth 3	0.5669	0.5509	0.4011
Depth 4	0.5498	0.5274	0.3842
Depth 5	0.5521	0.5202	0.3782
Depth 6	0.4989	0.4476	0.3333
Average	0.5505	0.5245	0.3834

of the left/right-views should be used. Mixed-resolution coding, asymmetric transform-domain quantization coding, and postprocessing techniques (deblocking or blurring) can be employed individually or collectively. Previously in [133, 17] and [3], the extent of the downsampling ratio that can be applied to a low quality view without a noticeable degradation on the 3D quality has been investigated. In [3], symmetric stereoscopic video coding, asymmetric quantization coding and mixed-resolution coding have been compared and the results suggested that mixed-resolution coding achieves the best coding efficiency. In [143], different levels of Gaussian blurring are applied after asymmetric quantization and a significant bit rate reduction has been achieved for this joint asymmetric compression and postprocessing method. However, here our new observations indicate that asymmetric compression and asymmetric blurring will influence the perceived 3D depth quality, i.e., adding blur or JPEG compression to one view of stereopair has similar effect in depth perception as adding the same level of distortion to both views. This is quite different from the distortion type dependency in 3D image quality perception. Therefore, the current study suggests that mixed-resolution coding, asymmetric transform-domain quantization coding, and postprocessing schemes need to be carefully reexamined and redesigned to maintain a good tradeoff between perceptual 3D image quality and depth quality. One possible solution is that a threshold on H may be used as a constraint in the process of asymmetrical bit allocation, ensuring that the quality of depth perception will not be severely affected.

4.6 Summary

In this chapter, we have carried out two subjective studies on depth perception of stereoscopic 3D images. The first one follows a traditional framework where subjects are asked to rate depth quality directly on distorted stereopairs. The second one uses a novel approach, where the stimuli are synthesized independent of the background image content and the subjects are asked to identify depth changes and label the polarities of depth. Our analysis shows that the second approach is much more effective at singling out the contributions of stereo cues in depth perception, through which we have several interesting findings regarding distortion type dependency, image content dependency, and the impact of symmetric and asymmetric distortions on the perception of depth. Furthermore, we propose a novel computational model for DPDI prediction. Our results show that the proposed model, without explicitly identifying image distortion types, leads to highly promising DPDI prediction performance. We believe these findings provide useful insights in the future development of comprehensive 3D QoE models for stereoscopic images, which have great potentials in real-world applications such as asymmetric compression of stereoscopic 3D videos.

Chapter 5

Asymmetrically Compressed Stereoscopic 3D Videos: Quality Assessment and Rate-Distortion Performance Evaluation

In this chapter, we first carry out subjective quality assessment experiments on two databases that contain various asymmetrically compressed stereoscopic 3D videos obtained from mixed-resolutions coding, asymmetric transform-domain quantization coding, their combinations, and multiple choices of postprocessing techniques. We compare these asymmetric stereoscopic video coding schemes with symmetric coding methods and verify their potential coding gains. We observe a strong systematic bias when using direct averaging of 2D video quality of both views to predict 3D video quality. We then apply a binocular rivalry inspired model to account for the prediction bias, leading to a significantly improved quality prediction model of stereoscopic videos. The model allows us to quantitatively predict the coding gain of different variations of asymmetric video compression, and provides new insight on the development of high efficiency 3D video coding schemes.

5.1 Introduction

In Chapter 3, we presented a systematic study on quality prediction of asymmetrically distorted stereoscopic 3D images and revealed a strong distortion type dependent prediction bias when predicting quality of asymmetrically distorted stereoscopic images from single-views. Studying the impact of asymmetric distortions on the quality of stereoscopic images not only has scientific values in understanding the HVS, but is also desirable in the practice of 3D video compression and transmission. The distortions involved in 3D video coding/communication are not only compression artifacts. The practical encoder/decoder also needs to decide on whether deblocking filters need to be turned on, and whether mixed-resolutions of the left/right-views should be used. Mixed-resolutions coding, asymmetric transform-domain quantization coding, and postprocessing techniques (deblocking or blurring) can be employed individually or collectively. Previously, with regard to transform-domain quantization coding, Saygili *et al.* found that asymmetric coding can perform better than symmetric coding when the lower quality view is encoded above a threshold value [112]. The subjective studies in [33] showed that stereoscopic asymmetry introduced by way of asymmetric blurriness is preferred over asymmetric blockiness, which is agreed by [5], where low-pass filtering shows no negative effect on the perceived 3D quality, sharpness and depth. In 1992, Perkins [99] introduced the idea of mixed-resolutions coding for stereoscopic video and implemented a mixed-resolutions coding scheme with a subsampling factor of 4, which can reduce the bit rate by 46% and resulted in little subjective degradation in picture quality and only moderate degradation in perceive depth. Brust *et al.* conducted subjective and objective tests on full and mixed-resolutions stereo video coding with a downsampling factor of 2 [17]. Experimental results showed that at low bit rates mixed-resolutions coded sequences have better perceptual qualities and the optimal bit rate allocation strategy is 30% to 35% of the total bit rate for the lower quality view. In [133], different vertical and horizontal spatial low-pass filtering on the right-view video were applied and subjective results showed that the perceived spatial quality and sharpness have a strong tendency towards the higher quality view and the perceived depth was unaffected. Aflaki *et al.* investigated the extent of downsampling ratios that can be applied to a low quality view without a noticeable degradation on the 3D quality [3]. They also compared the coding efficiency of symmetric quantization coding, asymmetric quantization coding and mixed-resolutions coding and found that mixed-resolutions coding can pronounce a similar 3D quality to that of symmetric coding with a significantly reduced computational complexity. Several observations in [142] provide useful implications on stereoscopic image/video coding. Specially, for JPEG compression, 3D image quality is more affected by the poorer quality view; while for blur, 3D image quality is more affected by the better quality view. Such distortion type dependency is more pronounced for strong asymmetric distortions. Moreover, for mixed-distortions types, when one view is JPEG compressed and the other is blurred, the JPEG compressed view dominates quality judgement regardless of their distortion levels. These observations suggest that simply coding one view at high rate and the other at low rate may not be a wise choice. This also suggests that a significant coding gain may be achieved by mixed-resolutions coding, followed by postprocessing techniques. However, in the literature, systematic studies on subjective and objective quality assessment of these variations of asymmetric stereoscopic video coding are still lacking, making it difficult to directly compare different coding strategies, nor to derive 3D-VQA models to guide asymmetrical 3D video coding.

In this chapter, we first carry out subjective quality assessment experiments on two databases that contain various asymmetrically compressed stereoscopic 3D videos created by mixed-resolutions coding, asymmetric transform-domain quantization coding, their combinations, and multiple choices of postprocessing techniques. We compare different variations of asymmetric stereoscopic video coding schemes with symmetric coding methods and verify their potential coding gains. We also observe a strong systematic bias when using direct averaging of 2D video quality of both views to predict 3D video quality. We then apply a binocular rivalry inspired model to account for the prediction bias, leading to a significantly improved quality model for stereoscopic videos. The model allows us to quantitatively predict the coding gain of different variations of asymmetric video compression, and provides new insight on the development of high efficiency 3D video coding schemes.

5.2 Subjective Study

5.2.1 WATERLOO-IVC 3D Video Quality Databases Phase I and Phase II

The new Waterloo-IVC 3D Video Quality Database Phase I is created from 4 pristine multi-view 3D videos, i.e., Balloons, Book, Kendo and Lovebird, which are commonly used 3D HEVC testing sequences. The new Waterloo-IVC 3D Video Quality Database Phase II is created from 6 pristine stereoscopic 3D videos, i.e., Barrier, Craft, Laboratory, Soccer, Tree and Dancer, which were collected from previous subjective 3D video quality studies [139, 100]. The details of the all test videos are given in Table 5.1. All videos are in YUV4:2:0 format. Sample frames for each test sequence are shown in Figure 5.1 and Figure 5.2.

Table 5.1: Test videos in Waterloo-IVC 3D Video Databases Phase I and Phase II

	Resolution	Length	Frames/Second	Views
Book	1024×768	6s	16.67	View 6 & View 8
Balloons	1024×768	10s	30.00	View 1 & View 3
Kendo	1024×768	10s	30.00	View 1 & View 3
Lovebird	1024×768	10s	30.00	View 4 & View 6
Barrier	1920×1080	10s	30.00	N/A
Craft	1920×1080	10s	30.00	N/A
Laboratory	1920×1080	10s	30.00	N/A
Soccer	1920×1080	10s	30.00	N/A
Tree	1920×1080	10s	30.00	N/A
Dancer	1920×1088	10s	30.00	View 1 & View 5

Waterloo-IVC 3D Video Database Phase I include stereoscopic 3D videos obtained from symmetric and asymmetric transform-domain quantization coding followed by different levels of low-pass filtering. Each single-view video was compressed using an HEVC encoder by five levels of transform-domain quantization with QP = {25, 35, 40, 45, 50} in low-delay main profile. The single-view videos were employed to generate compressed

(a) (b)

(c) (d)

Figure 5.1: Sample frames from the pristine videos used in the subjective study in Waterloo-IVC 3D Video Database Phase I. Only the right-views are shown here. (a) Balloons. (b) Book. (c) Kendo. (d) Lovebird.

(a) (b) (c) (d) (e) (f)

Figure 5.2: Sample frames from the pristine videos used in the subjective study in Waterloo-IVC 3D Video Database Phase II. Only the right-views are shown here. (a) Barrier. (b) Craft. (c) Laboratory. (d) Soccer. (e) Tree. (f) Dancer.

stereoscopic videos, either symmetrically or asymmetrically. There are 11 different kinds of combinations as listed in Table 5.2. The lower and higher QP views are assigned to the left-view or the right-view randomly. Moreover, for each QP combination, four levels of Gaussian low-pass filtering with $\sigma = \{0, 3.5, 7.5, 11.5\}$ are applied to the higher QP (lower quality) views. Altogether, there are totally 176 3D videos in the database.

Waterloo-IVC 3D Video Database Phase II include various stereoscopic 3D videos obtained from mixed-resolutions coding, asymmetric transform-domain quantization coding, their combinations, and different levels of low-pass filtering. Three choices of pre-processing, i.e., pre-downsampling by 2, pre-downsampling by 4, and pre-processed low-pass filtering with $\sigma = 2.5$, are applied to each single-view video. Then the single-view video was compressed using an HEVC encoder by different levels of transform-domain quantization with QP = {25, 30, 35, 40, 45} in low-delay main profile. The single-view videos were employed to generate compressed stereoscopic videos after upsampling if needed, either symmetrically or asymmetrically. Table 5.2 categorizes all combinations into nine groups with detailed descriptions. The lower- and higher-quality views are assigned to the left- or right-view randomly. Moreover, for each combination, two levels of Gaussian low-pass filtering with $\sigma = \{3.5, 5.5\}$ are applied to the lower-quality views. Altogether, there are totally 222 2D videos and 528 3D videos in the database.

There are two unique features of the new databases (including both Phases I and II) when compared with existing publicly known 3D-VQA databases. First, these are the only databases that allow us to perform subjective test on both 2D and 3D videos. The inclusion of 2D videos allows us to directly examine the relationship between the perceptual quality of stereoscopic video and that of its single-view videos. This is advantageous against previous studies which do not have ground truth of 2D video quality but have to rely on objective 2D-VQA measures to provide estimates. Second, these are the only databases that contain asymmetrically compressed stereoscopic videos from mixed-resolution coding, asymmetric transform-domain quantization coding and their combinations, followed by different levels of low-pass filtering. This provides the potential of a much stronger test on 3D-VQA models on their generalizability to real world applications. Such test has been largely lacking in previous studies where the development of objective 3D-VQA models only took into account asymmetric distortions of specific and very limited distortion types such

Table 5.2: Test videos on Waterloo-IVC 3D Video Databases Phase I and Phase II

Waterloo-IVC 3D Video Database Phase I			
Group	# of Videos	Description	Combination
3D.1.a	4×4	Symmetrically compressed stereoscopic videos	(QP_1, QP_2) = (25,25), (35,35), (40,40) and (50,50)
3D.1.b	4×4	Postprocessing P1 $\sigma = 3.5$	
3D.1.c	4×4	Postprocessing P3 $\sigma = 7.5$	
3D.1.d	4×4	Postprocessing P4 $\sigma = 11.5$	
3D.2.a	4×7	Asymmetrically compressed stereoscopic videos	(QP_1, QP_2) = (25,35), (25,40), (25,45), (25,50), (35,45), (35,50) and (40,50)
3D.2.b	4×7	Postprocessing P1 $\sigma = 3.5$	
3D.2.c	4×7	Postprocessing P3 $\sigma = 7.5$	
3D.2.d	4×7	Postprocessing P4 $\sigma = 11.5$	

Waterloo-IVC 3D Video Database Phase II			
Group	# of Videos	Description	Description
2D.1.a	6×4	Compressed single-view videos	QP = 30, 35, 40 and 45
2D.1.b	6×4	Postprocessing P1 $\sigma = 3.5$	
2D.1.c	6×4	Postprocessing P2 $\sigma = 5.5$	
2D.2	6×4	Compressed single-view videos with pre-processing S1 $\sigma = 2.5$	QP = S1-30, S1-35, S1-40 and S1-45
2D.3.a	6×4	Compressed single-view videos with pre-downsampling by 2	QP = D2-25, D2-30, D2-35 and D2-40
2D.3.b	6×4	Postprocessing P1 $\sigma = 3.5$	
2D.3.c	6×4	Postprocessing P2 $\sigma = 5.5$	
2D.4.a	6×3	Compressed single-view videos with pre-downsampling by 4	QP = D4-20, D4-25 and D4-30
2D.4.b	6×3	Postprocessing P1 $\sigma = 3.5$	
2D.4.c	6×3	Postprocessing P2 $\sigma = 5.5$	
3D.3.a	6×4	Symmetrically compressed stereoscopic videos	(QP_1, QP_2) = (30,30), (35,35), (40,40) and (45,45)
3D.3.b	6×4	Postprocessing P1 $\sigma = 3.5$	
3D.3.c	6×4	Postprocessing P2 $\sigma = 5.5$	
3D.4.a	6×6	Asymmetrically compressed stereoscopic videos	(QP_1, QP_2) = (30,35), (30,40), (30,45), (35,40), (35,45) and (40,45)
3D.4.b	6×6	Postprocessing P1 $\sigma = 3.5$	
3D.4.c	6×6	Postprocessing P2 $\sigma = 5.5$	
3D.5	6×10	Asymmetrically compressed stereoscopic videos with pre-processing S1 $\sigma = 2.5$	(QP_1, QP_2) = (30,S1-30), (30,S1-35), (30,S1-40), (30,S1-45), (35,S1-35), (35,S1-40), (35,S1-45), (40,S1-40), (40,S1-45) and (45,S1-45)
3D.6.a	6×8	Asymmetrically compressed stereoscopic videos with pre-downsampling by 2	(QP_1, QP_2) = (30,D2-25), (30,D2-30), (30,D2-35), (35,D2-30), (35,D2-35), (35,D2-40), (40,D2-35), (40,D2-40)
3D.6.b	6×8	Postprocessing P1 $\sigma = 3.5$	
3D.6.c	6×8	Postprocessing P2 $\sigma = 5.5$	
3D.7.a	6×8	Asymmetrically compressed stereoscopic videos with pre-downsampling by 4	(QP_1, QP_2) = (30,D4-20), (30,D4-25), (30,D4-30), (35,D4-25), (35,D4-30), (35,D4-35), (40,D4-25), (40,D4-30)
3D.7.b	6×8	Postprocessing P1 $\sigma = 3.5$	
3D.7.c	6×8	Postprocessing P2 $\sigma = 5.5$	

as compression only. Meanwhile, a broader variety of test scenarios allows us to perform a more comprehensive comparison on different variations of asymmetric stereoscopic video coding schemes with symmetric coding methods and thus to evaluate their potential coding gains.

5.2.2 Subjective Test

The subjective test was conducted in the Lab for Image and Vision Computing at University of Waterloo. The test environment has no reflecting ceiling walls and floor, and was not insulated by any external audible and visual pollution. An ASUS 27” VG278H 3D LED monitor with NVIDIA 3D Vision$^{\text{TM}}$2 active shutter glasses is used for the test. The default viewing distance was 3.5 times the screen height. In the actual experiment, some subjects did not feel comfortable with the default viewing distance and were allowed to adjust the actual viewing distance around it. The details of the viewing conditions are given in Table 3.17.

In Phase I, twenty-two naïve subjects, 12 males and 10 females aged between 22 and 35, participated in the study. In Phase II, thirty-two naïve subjects, 20 males and 12 females aged between 24 and 37, participated in the study. A 3D vision test was conducted first to verify their ability to view stereoscopic 3D content and no one failed the vision test. As a result, a total of twenty-two and thirty-two subjects proceeded to the formal test in Phase I and Phase II, respectively. While a visual acuity test was not performed in this study, a verbal confirmation was obtained prior to the experiment and subjects were asked to use their eyeglasses or contact lenses to correct their visual acuities.

The subjects were asked to evaluate their overall 3D viewing experience − 3D video quality (3DVQ) in this study. Since to visualize every stereoscopic 3D video, the subjects need to make readjustment so as to adapt to the content of the scene and establish 3D perception, using a double stimulus approach leads to interruptions of the viewing experience. To reduce this effect, we chose to use the single stimulus procedure using an 11-grade numerical categorical scale protocol. A general introduction was given at the beginning of the whole test, and more specific instructions and training session were given afterwards. The rating strategy was introduced and the subjects were required to practice by giving scores to training 2D/3D videos until they fully understood the criteria and built up their own scoring strategies. For both Phase I and Phase II tests, we use three types of videos in the training phase: pristine 2D/3D videos, moderately distorted 2D/3D videos, and highly-distorted 2D/3D videos. The subjects were told to give scores at the high end (close to 10 pts) to the pristine 2D/3D videos, at the mid-range to the moderately distorted 2D/3D videos, and at the low end (close to 0 pts) to the highly-distorted videos. With regard to Phase II, we also found that the 2D perceptual quality of left- and right-view videos are very close to each other at the same compression or postprocessing levels, and the difference in their MOS is negligible. Thus in order to control the scale of this subjective experiment, only one of the views were tested (randomly picked) in Group 2D.1 to Group 2D.4 in the formal test.

All stimuli were shown once in each test. However, there were 12 repetitions for single-view or stereoscopic videos, which means that for each subject, her/his first 12 single-view or stereoscopic videos were shown twice. The order of stimuli was randomized and the consecutive testing stereoscopic videos were from different source contents. There are three sessions for the 3D test in Phase I, while in Phase II, there are three sessions for the 2D test and nine sessions for the 3D test. Each single session, where around 80 single-view or 60 stereoscopic videos were evaluated, was finished in 15 to 20 minutes. Sufficient relaxation periods (5 minutes or more) were given between sessions. Thus in Phase I all sessions were finished in 2 hours in one day. In Phase II, the test was scheduled on two consecutive days for each subject. Day 1 (2 to 2.5 hours) was dedicated to all 2D sessions and the first three 3D sessions and Day 2 (2 to 2.5 hours) to the remaining six 3D sessions. Figure 5.3 shows the detailed procedure of our formal subjective test.

Moreover, we found that repeatedly switching between viewing 3D videos and grading on a piece of paper or a computer screen is a tiring experience. To overcome this problem, we asked the subject to speak out a score between 0 and 10, and a customized graphical user interface on another computer screen was used by the instructor to record the score. All these efforts were intended to reduce visual fatigue and discomfort of the subjects.

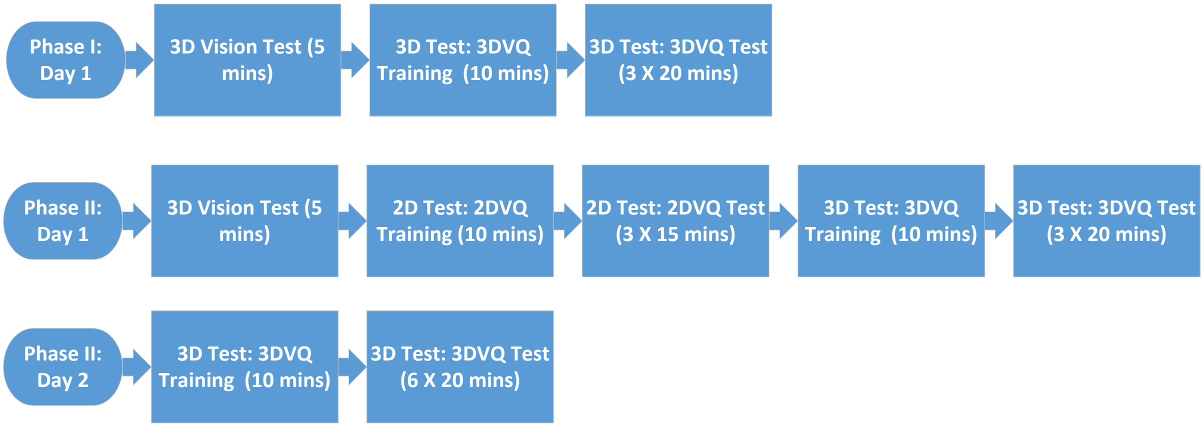


Figure 5.3: The procedure of the subjective test in Waterloo-IVC 3D Video Database Phase I and Phase II.

5.2.3 Impact of Eye Dominance

Eye dominance is a common visual phenomenon, referring to the tendency to prefer the input from one eye to the other, depending on the human subject [67]. When studying visual quality of asymmetrically compressed stereoscopic videos, it is important to understand if eye dominance plays a significant role in the subjective test results. For this purpose, we carried out a separate analysis on the impact of eye dominance in the perception of asymmetrically compressed stereoscopic videos in Phase II. The side of the dominant eye under static conditions was checked first by Rosenbach's test [107]. This test examines which eye determines the position of a finger when the subject is asked to point to an object. Among thirty-two subjects who finished the formal test in Phase II, thirteen subjects (8 males, 5 females) are left-eye dominant, and the others (12 males, 7 females) are right-eye dominant.

The 3DVQ MOS scores for each video in Phase II were computed for left-eye dominant subjects and right-eye dominant subjects, denoted as $3DVQ_{\mathrm{L}}$ and $3DVQ_{\mathrm{R}}$, respectively. We employed the one-sample t-test to obtain a test decision for the null hypothesis that the difference between $3DVQ_{\mathrm{L}}$ and $3DVQ_{\mathrm{R}}$, i.e., $3DVQ_{\mathrm{D}} = 3DVQ_{\mathrm{L}} - 3DVQ_{\mathrm{R}}$, comes from a normal distribution of zero-mean and unknown variance. The alternative hypothesis is that the population distribution does not have a mean equaling zero. The result h is 1 if the test rejects the null hypothesis at the 5% significance level, and 0 otherwise. The returned p-values for different test video groups are reported in Table 5.3. From Table 5.3, it can be seen that the null hypothesis cannot be rejected at the 5% significance level, which indicates that the impact of eye dominance in the perception of asymmetrically compressed stereoscopic videos is insignificant.

Table 5.3: p-values from the one-sample t-test for different test video groups

Group	p-values
Group 3D.3.a	0.5901
Group 3D.4.a	0.4255
Group 3D.3.b-c and Group 3D.4.b-c	0.1401
Group 3D.5	0.6637
Group 3D.6.a	0.6833
Group 3D.6.b-c	0.3172
Group 3D.7.a	0.4379
Group 3D.7.b-c	0.3495

It is worth noting that similar conclusions were reached in our earlier studies on the impact of eye dominance on the quality of asymmetrically distorted stereoscopic images in Section 3.4 and on the depth perception induced by stereo cues of asymmetrically distorted stereograms in Section 4.3.5. These observations are consistent with the "stimulus" view of rivalry that is widely accepted in the field of visual neuroscience [15]. A comprehensive review and discussion on "stimulus" rivalry versus "eye" rivalry can be found in [15, 81].

5.3 Analysis and Findings

5.3.1 Relationship between 2D and 3D Video Quality

The raw 2D video quality (2DVQ) and 3DVQ scores given by each subject were converted to Z-scores, respectively. Then the entire data sets were rescaled to fill the range from 1 to 100 and the MOS scores for each 2D and 3D video was computed after removing outliers [101]. Given the subjective 2D and 3D data, we are interested in how single-view 2D video quality predicts stereoscopic 3D video quality, especially for the case of asymmetrically compressed and post-processed stereoscopic videos. The most straightforward 2D-to-3D quality prediction method is to average the qualities of the left- and right-view videos. Table 5.4 shows PLCC, SRCC, KRCC, RMSE and MAE values between 3DVQ-MOS scores and average 2DVQ-MOS scores (Phase II Only), where the 2D-IQA/VQA measurements include PSNR, SSIM, MS-SSIM [158], IW-SSIM [154], and VQM [102] for all stereoscopic videos in Phase I and Phase II. Table 5.5 and Table 5.6 reports PLCC, SRCC, KRCC, RMSE and MAE values for different test video groups in Phase I and Phase II, respectively. PLCC, RMSE and MAE are adopted to evaluate prediction accuracy [46] and SRCC and KRCC are employed to assess prediction monotonicity [46]. Higher PLCC, SRCC and KRCC and lower RMSE and MAE values indicate better consistency with human opinions of quality. PLCC, RMSE and MAE are usually computed after a nonlinear mapping between the subjective and objective scores and the results may be sensitive to the choice of the mapping function. SRCC and KRCC are nonparametric rank order-based correlation metrics, independent of any monotonic nonlinear mapping between subjective and objective scores but do not explicitly estimate the accuracy of quality prediction.

In Table 5.5 and Table 5.6, we first compare the performance of symmetrically compressed 3D videos without postprocessing against asymmetrically compressed 3D videos without postprocessing. Unsurprisingly, accurate predictions are obtained in the category of symmetrically compressed 3D videos. By contrast, the performance drops for asymmetrically compressed 3D videos. In Section 3.5.2, we reported that for JPEG compression, average prediction overestimates 3D quality (or 3D quality is more affected by the poorer quality view). More importantly we found that for blockiness, the bias of the averaging prediction model increases with the level of distortions, and thus whether the bias is pronounced depends on the quality range being investigated. With respect to blockiness created from HEVC compression, this overestimated prediction bias is still pronounced, but not as strong as JPEG compression, which is likely due to the reduction of blocking artifacts in HEVC.

We then compare the performance of compressed 3D videos without postprocessing against compressed 3D videos with postprocessing. From Table 5.5 and Table 5.6, it can be observed that the direct averaging model performs well for 3D videos without postprocess-

Table 5.4: Performance comparison of 2D-to-3D quality prediction models on Waterloo-IVC 3D Video Databases Phase I and Phase II

Waterloo-IVC 3D Video Database Phase I					
Method	PLCC	SRCC	KRCC	RMSE	MAE
Average 2DVQ-PSNR	0.7085	0.5336	0.3849	15.4507	12.5201
Weighted 2DVQ-PSNR	0.8980	0.8366	0.6150	9.6344	7.3743
Average SSIM	0.3964	0.2872	0.2065	20.1010	16.4217
Weighted SSIM	0.8905	0.8393	0.6553	9.9615	7.7379
Average MS-SSIM	0.4072	0.2969	0.2142	19.9978	16.3647
Weighted MS-SSIM	0.8838	0.8287	0.6458	10.2448	7.9117
Average IW-SSIM	0.4833	0.2787	0.1943	19.1683	15.5577
Weighted IW-SSIM	0.8942	0.8364	0.6514	9.8035	7.5736
Average VQM	0.7912	0.6321	0.4648	13.3905	10.6730
Weighted VQM	0.9191	0.8655	0.6860	8.6273	6.6545

Waterloo-IVC 3D Video Database Phase II					
Method	PLCC	SRCC	KRCC	RMSE	MAE
Average 2DVQ-MOS	0.6912	0.6277	0.4530	8.9039	7.2380
Weighted 2DVQ-MOS	0.8829	0.8727	0.6873	5.7849	4.5634
Average PSNR	0.3699	0.3414	0.2368	11.4465	8.9958
Weighted PSNR	0.5590	0.5109	0.3552	10.2154	8.0058
Average SSIM	0.3303	0.2589	0.1777	11.6291	9.2110
Weighted SSIM	0.7571	0.7309	0.5375	8.0487	6.6116
Average MS-SSIM	0.3034	0.2503	0.1724	11.7395	9.3157
Weighted MS-SSIM	0.6813	0.6377	0.4565	9.0188	7.3775
Average IW-SSIM	0.3243	0.2459	0.1699	11.6545	9.1862
Weighted IW-SSIM	0.7677	0.7423	0.5498	7.8943	6.2250
Average VQM	0.7019	0.6287	0.4491	8.7759	6.9395
Weighted VQM	0.8496	0.8042	0.6017	6.4976	5.0903

Table 5.5: Performance comparison of 2D-to-3D quality prediction models on Waterloo-IVC 3D Video Database Phase I

	PLCC		SRCC		KRCC		RMSE		MAE	
Method	Sym.	Asym.	Sym.	Asym.	Sym.	Asym.	Sym.	Asym.	Sym.	Asym.
Average PSNR	0.9839	0.8217	0.9581	0.8226	0.8619	0.6436	5.3453	10.8551	4.1647	8.1225
Weighted PSNR	0.9839	0.8798	0.9581	0.8311	0.8619	0.6489	5.3437	9.0531	4.1615	6.9529
Average SSIM	0.9876	0.8560	0.9478	0.7960	0.8285	0.6223	4.6806	9.8468	4.0762	8.4597
Weighted SSIM	0.9877	0.8843	0.9478	0.8335	0.8285	0.6489	4.6782	8.8952	4.0733	7.6126
Average MS-SSIM	0.9851	0.8445	0.9478	0.7990	0.8285	0.6277	5.1417	10.2018	4.5230	8.7625
Weighted MS-SSIM	0.9851	0.8746	0.9478	0.8363	0.8285	0.6489	5.1391	9.2357	4.5202	7.9473
Average IW-SSIM	0.9937	0.9304	0.9581	0.9017	0.8452	0.7447	3.3373	6.9836	2.8170	5.5954
Weighted IW-SSIM	0.9937	0.9322	0.9581	0.9012	0.8452	0.7341	3.3379	6.8942	2.8181	5.5290
Average VQM	0.9876	0.8831	0.9669	0.8283	0.8787	0.6436	4.6878	8.9408	3.8242	6.7432
Weighted VQM	0.9876	0.8770	0.9669	0.8527	0.8787	0.6755	4.6893	9.1511	3.8347	7.2326
Symmetric compression: Group 3D.1.a; Asymmetric compression: Group 3D.1.b.										
	PLCC		SRCC		KRCC		RMSE		MAE	
Method	N.P.	W.P.	N.P.	W.P.	N.P.	W.P.	N.P.	W.P.	N.P.	W.P.
Average PSNR	0.9131	0.7382	0.8922	0.6580	0.7416	0.4860	9.6120	18.7419	7.2835	14.3622
Weighted PSNR	0.9146	0.8750	0.8831	0.8165	0.7268	0.6270	9.5338	10.1255	7.5099	8.2308
Average SSIM	0.9193	0.4824	0.9028	0.2621	0.7438	0.1869	9.2752	18.3211	7.6274	14.3547
Weighted SSIM	0.9358	0.8876	0.9197	0.8056	0.7629	0.6184	8.3072	9.6338	6.9238	7.5056
Average MS-SSIM	0.9132	0.5315	0.9009	0.2701	0.7416	0.1918	9.6036	17.7168	7.8953	13.8141
Weighted MS-SSIM	0.9299	0.8798	0.9202	0.7896	0.7629	0.6040	8.6678	9.9415	7.2522	7.6940
Average IW-SSIM	0.9601	0.5935	0.9446	0.3262	0.8117	0.2299	6.5882	16.8340	5.3686	13.1710
Weighted IW-SSIM	0.9641	0.8977	0.9491	0.8135	0.8180	0.6272	6.2567	9.2176	5.0623	6.9820
Average VQM	0.9265	0.8733	0.9076	0.7140	0.7607	0.5365	8.8679	10.1899	6.7266	8.2099
Weighted VQM	0.9282	0.9341	0.9041	0.8414	0.7544	0.6602	8.7715	7.4652	6.8871	5.8199
No Postprocessing: Group 3D.1.a and Group 3D.2.a; With Postprocessing: Group 3D.1.b-d and Group 3D.2.b-d.										

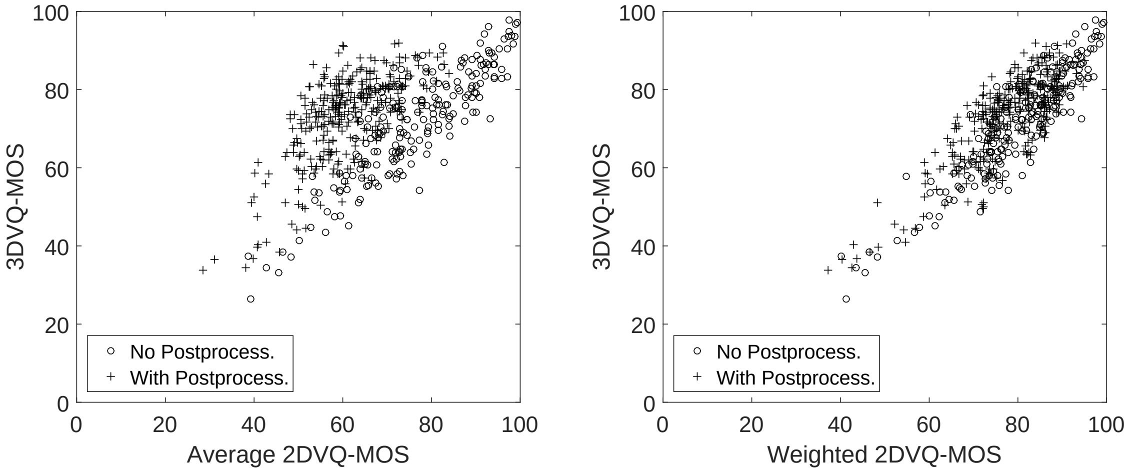
Table 5.6: Performance comparison of 2D-to-3D quality prediction models on Waterloo-IVC 3D Video Database Phase II

	PLCC		SRCC		KRCC		RMSE		MAE	
Method	Sym.	Asym.	Sym.	Asym.	Sym.	Asym.	Sym.	Asym.	Sym.	Asym.
Average 2DVQ-MOS	0.9676	0.8645	0.9470	0.8387	0.8116	0.6512	5.2708	6.3878	4.0883	5.3093
Weighted 2DVQ-MOS	0.9676	0.9002	0.9470	0.8984	0.8116	0.7225	5.2708	5.5350	4.0883	4.4378
Average PSNR	0.7717	0.3718	0.7191	0.3323	0.5217	0.2289	13.2946	11.7976	11.1093	9.6099
Weighted PSNR	0.7716	0.5027	0.7191	0.4468	0.5217	0.3122	13.2957	10.9864	11.1104	8.8680
Average SSIM	0.9236	0.5454	0.8974	0.5332	0.7246	0.3721	8.0089	10.6524	6.6367	8.6164
Weighted SSIM	0.9236	0.7625	0.8974	0.7610	0.7246	0.5602	8.0091	8.2224	6.6369	6.7762
Average MS-SSIM	0.8733	0.4658	0.8609	0.4431	0.6739	0.3025	10.1749	11.2459	8.4206	9.1572
Weighted MS-SSIM	0.8733	0.6697	0.8609	0.6534	0.6739	0.4647	10.1751	9.4382	8.4207	7.7728
Average IW-SSIM	0.9335	0.5460	0.9209	0.5253	0.7754	0.3680	7.4910	10.6476	6.4055	8.5251
Weighted IW-SSIM	0.9335	0.7836	0.9209	0.7828	0.7754	0.5834	7.4923	7.8957	6.4065	6.2687
Average VQM	0.9586	0.8107	0.9226	0.7731	0.7319	0.5788	5.9446	7.4397	5.0367	5.8921
Weighted VQM	0.9586	0.8633	0.9226	0.8457	0.7319	0.6473	5.9448	6.4137	5.0372	5.0633

Symmetric compression: Group 3D.3.a; Asymmetric compression: Group 3D.4.a, Group 3D.5, Group 3D.6.a and Group 3D.7.a.

	PLCC		SRCC		KRCC		RMSE		MAE	
Method	N.P.	W.P.	N.P.	W.P.	N.P.	W.P.	N.P.	W.P.	N.P.	W.P.
Average 2DVQ-MOS	0.8769	0.7800	0.8540	0.7394	0.6718	0.5499	6.6768	6.9184	5.4546	5.3439
Weighted 2DVQ-MOS	0.9144	0.8603	0.9097	0.8414	0.7425	0.6479	5.6234	5.6356	4.4358	4.4276
Average PSNR	0.4205	0.4204	0.3811	0.3772	0.2627	0.2711	12.6039	10.0312	10.1841	7.5662
Weighted PSNR	0.5411	0.5773	0.4905	0.5385	0.3445	0.3767	11.6823	9.0271	9.4612	6.9474
Average SSIM	0.5841	0.2596	0.5694	0.2459	0.4017	0.1706	11.2760	10.6766	9.0407	8.1984
Weighted SSIM	0.7890	0.7493	0.7835	0.7084	0.5851	0.5199	8.5352	7.3217	7.1101	5.8585
Average MS-SSIM	0.5111	0.2911	0.4868	0.2286	0.3359	0.1621	11.9400	10.5769	9.7166	8.0969
Weighted MS-SSIM	0.7040	0.6708	0.6864	0.6014	0.4951	0.4295	9.8656	8.1992	8.2704	6.5705
Average IW-SSIM	0.5710	0.3875	0.5520	0.3427	0.3896	0.2391	11.4040	10.1920	9.0301	7.8493
Weighted IW-SSIM	0.8068	0.7588	0.8002	0.7159	0.6051	0.5277	8.2076	7.2011	6.5355	5.5883
Average VQM	0.8231	0.8051	0.7911	0.7155	0.5981	0.5158	7.8901	6.5569	6.1850	5.2176
Weighted VQM	0.8840	0.8326	0.8674	0.7590	0.6777	0.5500	6.4932	6.1242	5.1306	4.7356

No Postprocessing: Group 3D.3.a, Group 3D.4.a, Group 3D.5, Group 3D.6.a and Group 3D.7.a;
With Postprocessing: Group 3D.3.b-c, Group 3D.4.b-c, Group 3D.6.b-c and Group 3D.7.b-c.



(a)

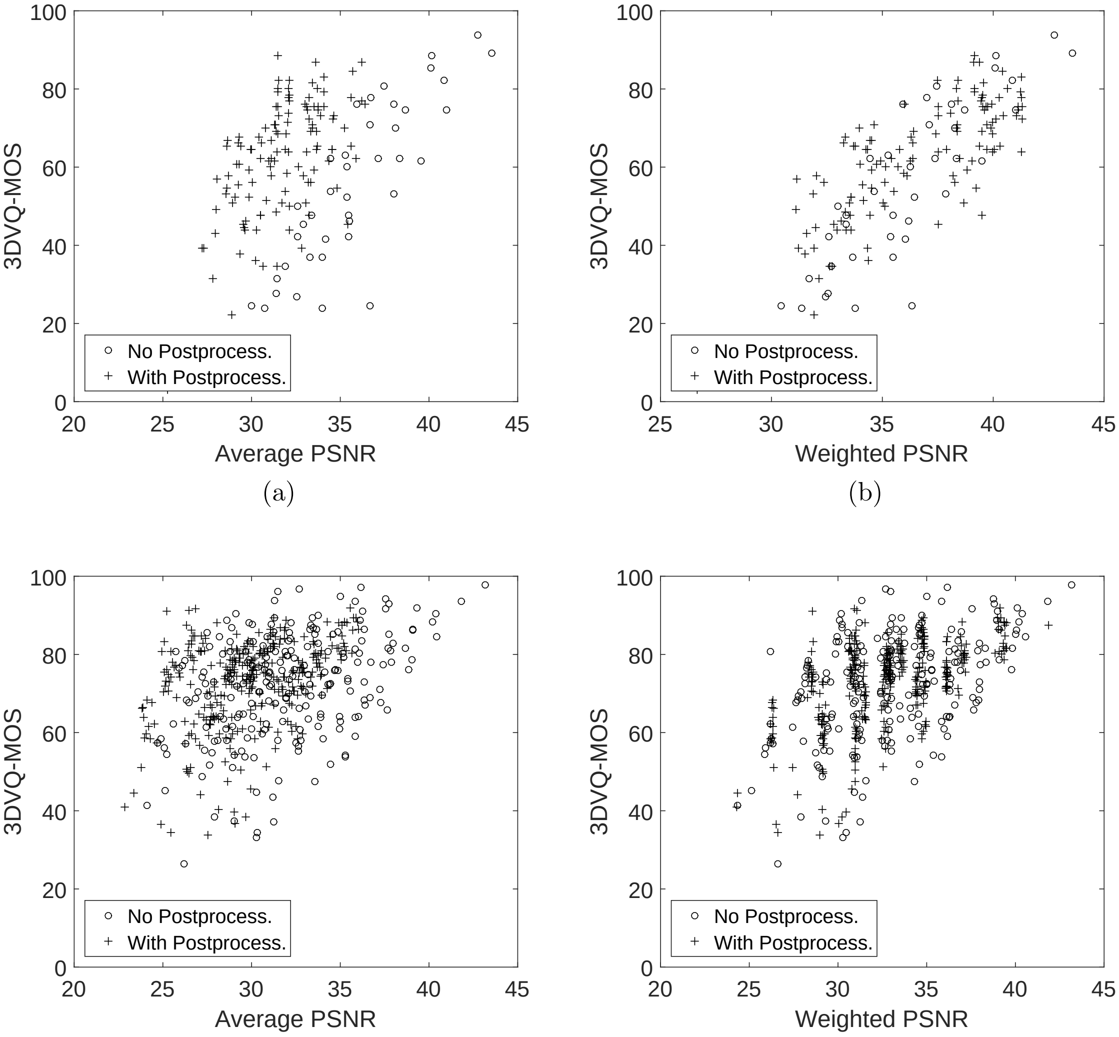
(b)

Figure 5.4: 3DVQ-MOS versus predictions from 2DVQ-MOS of 2D left- and right-views on Waterloo-IVC 3D Video Database Phase II. (a) Average of 2DVQ-MOS. (b) Weighted average of 2DVQ-MOS by the proposed method.

ing (by Gaussian blurring). By contrast, the correlation values drop significantly for videos with postprocessing. The Figure 5.4 and Figure 5.5 to Figure 5.9 show the corresponding scatter plots between 3DVQ-MOS scores and 2DVQ-MOS scores or 2D-IQA/VQA measurements, where the simple averaging prediction model generates substantial bias on many stereoscopic videos. In Section 3.5.2, we reported that for blurriness, average prediction often underestimates 3D quality (or 3D quality is more affected by the better quality view). Here the same kind of prediction bias is clearly observed, as direct averaging of state-of-the-art 2D-IQA/VQA metrics always underestimates 3D video quality for these post-processed videos.

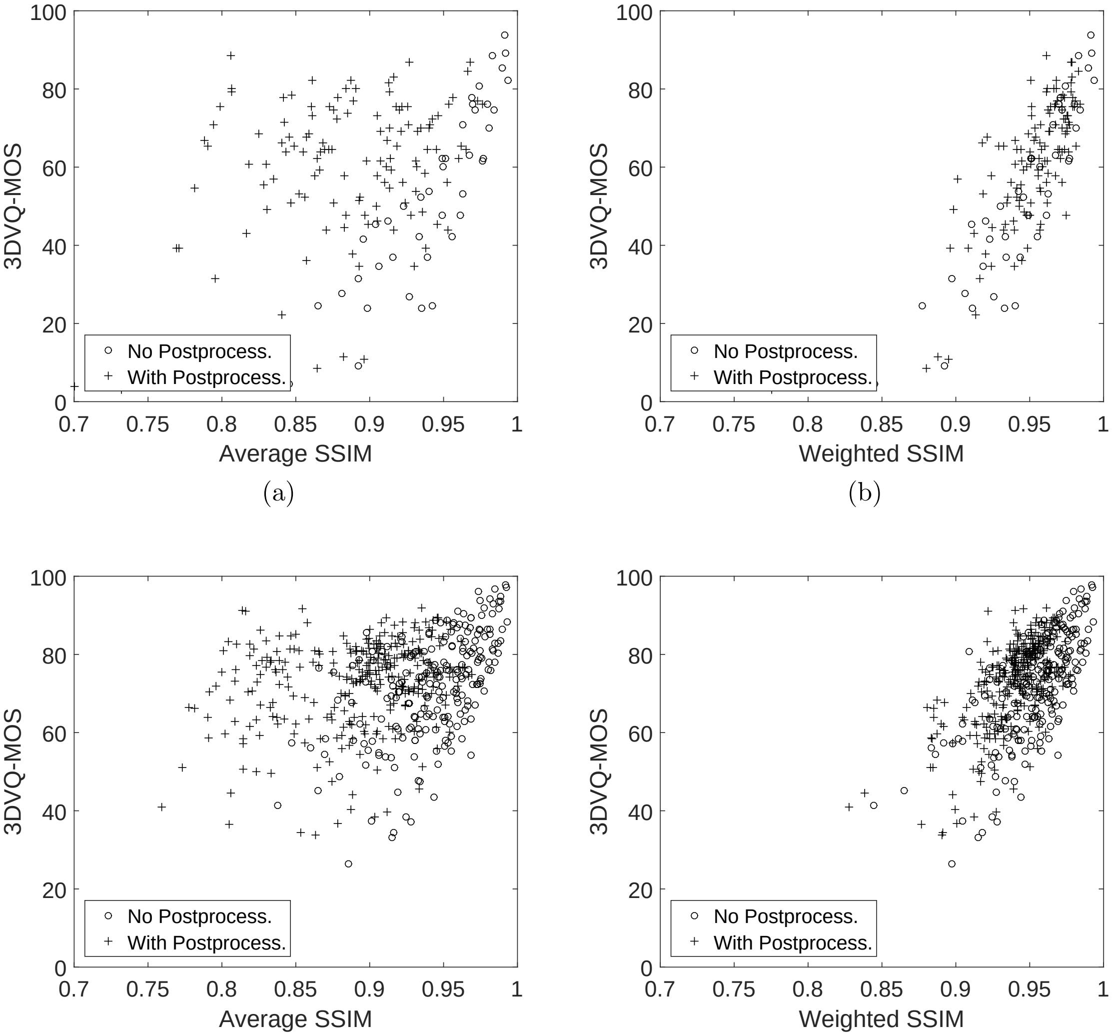
5.3.2 Quality of Asymmetric Stereoscopic Video with Postprocessing

Given the subjective data, the second question we would like to ask is how Gaussian low-pass post-filtering affects the perceptual 3D quality of asymmetrically compressed stereo-



(a) (b) (c) (d)

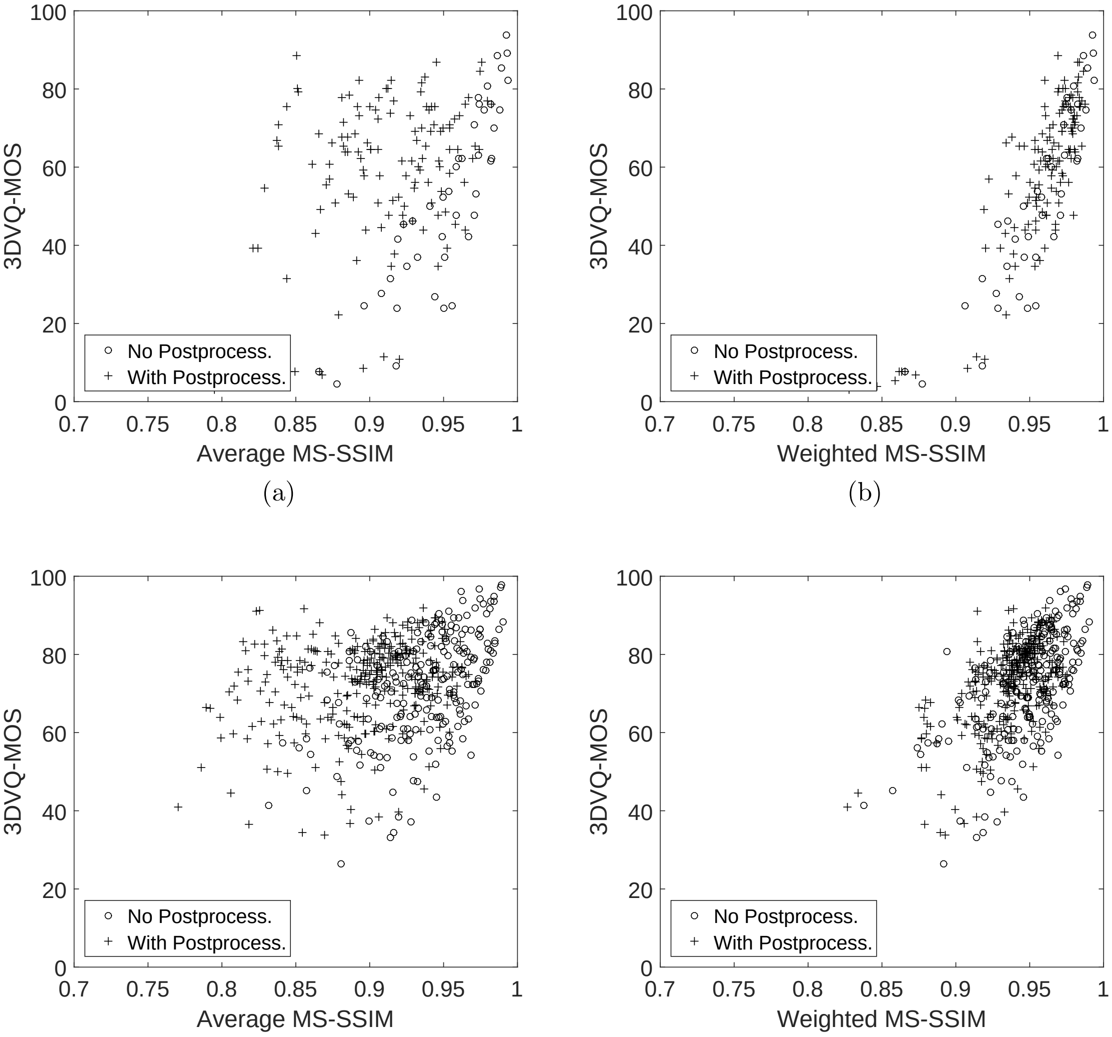
Figure 5.5: 3DVQ-MOS versus predictions from 2D-IQA/VQA estimations of 2D left- and right-views. (a) Average of PSNR on Waterloo-IVC Phase I. (b) Weighted average of PSNR by the proposed method on Waterloo-IVC Phase I. (c) Average of PSNR on Waterloo-IVC Phase II. (d) Weighted average of PSNR by the proposed method on Waterloo-IVC Phase II.



(a) (b)

(c) (d)

Figure 5.6: 3DVQ-MOS versus predictions from 2D-IQA/VQA estimations of 2D left- and right-views. (a) Average of SSIM on Waterloo-IVC Phase I. (b) Weighted average of SSIM by the proposed method on Waterloo-IVC Phase I. (c) Average of SSIM on Waterloo-IVC Phase II. (d) Weighted average of SSIM by the proposed method on Waterloo-IVC Phase II.



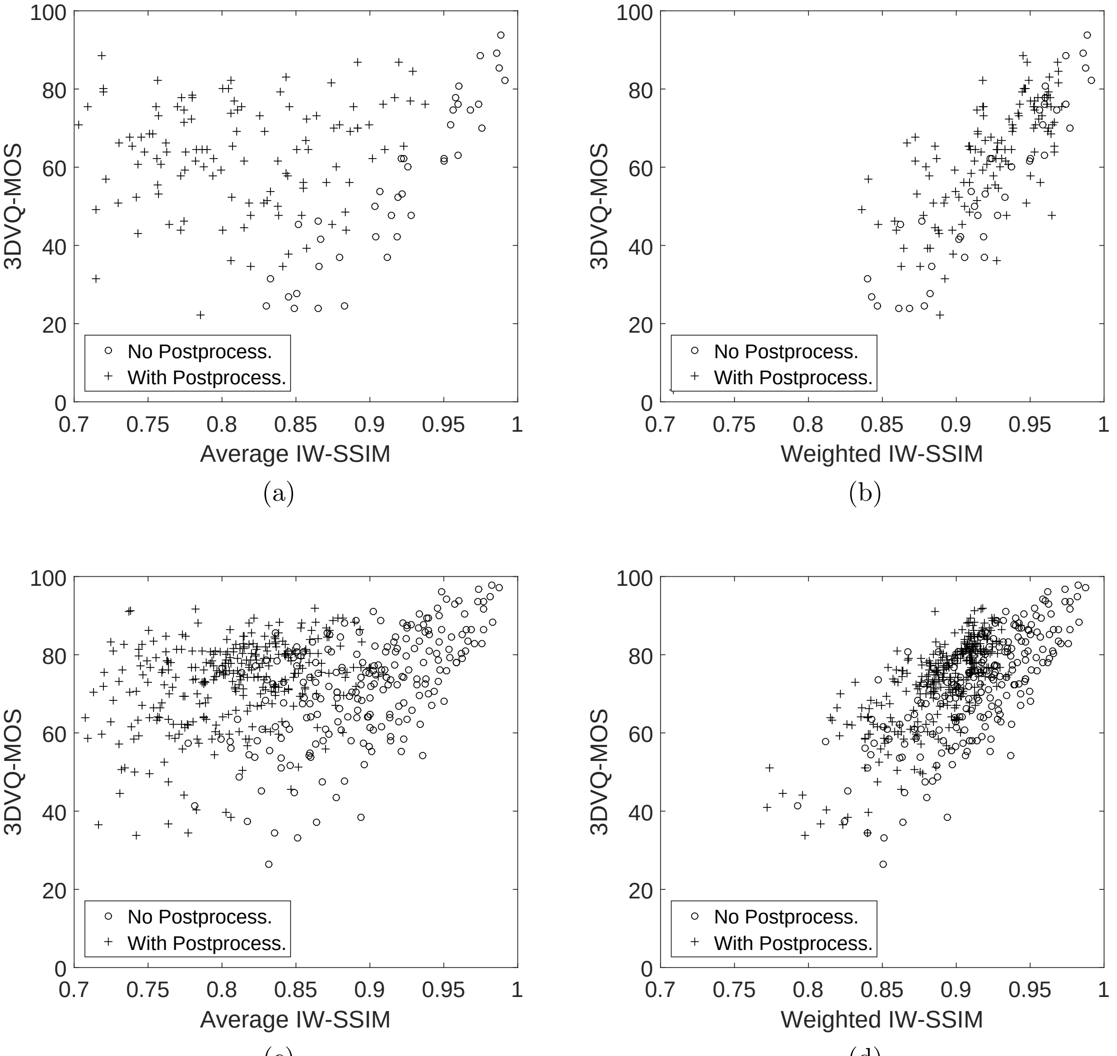
(a)

(b)

(c)

(d)

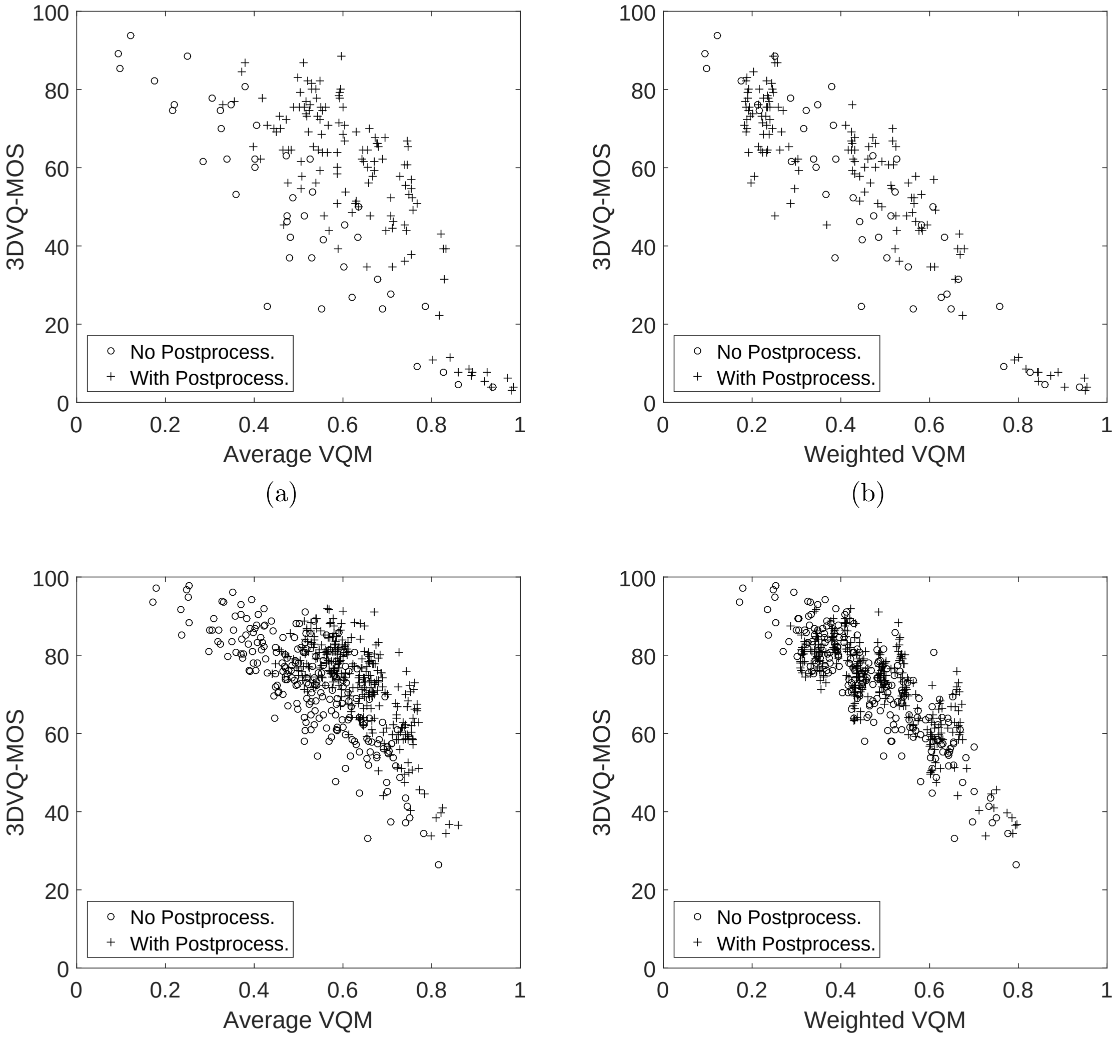
Figure 5.7: 3DVQ-MOS versus predictions from 2D-IQA/VQA estimations of 2D left- and right-views. (a) Average of MS-SSIM on Waterloo-IVC Phase I. (b) Weighted average of MS-SSIM by the proposed method on Waterloo-IVC Phase I. (c) Average of MS-SSIM on Waterloo-IVC Phase II. (d) Weighted average of MS-SSIM by the proposed method on Waterloo-IVC Phase II.



(a) (b)

(c) (d)

Figure 5.8: 3DVQ-MOS versus predictions from 2D-IQA/VQA estimations of 2D left- and right-views. (a) Average of IW-SSIM on Waterloo-IVC Phase I. (b) Weighted average of IW-SSIM by the proposed method on Waterloo-IVC Phase I. (c) Average of IW-SSIM on Waterloo-IVC Phase II. (d) Weighted average of IW-SSIM by the proposed method on Waterloo-IVC Phase II.



(a) (b)

(c) (d)

Figure 5.9: 3DVQ-MOS versus predictions from 2D-IQA/VQA estimations of 2D left- and right-views. (a) Average of VQM on Waterloo-IVC Phase I. (b) Weighted average of VQM by the proposed method on Waterloo-IVC Phase I. (c) Average of VQM on Waterloo-IVC Phase II. (d) Weighted average of VQM by the proposed method on Waterloo-IVC Phase II.

scopic videos. Table 5.7 and Table 5.8 report 3DVQ-MOS changes after applying different levels of Gaussian low-pass filtering with respect to different QP combinations and blurring levels for Phase I and Phase II, respectively. It can be observed that for symmetrically compressed 3D videos, blurring reduces perceptual 3D video quality in most cases. By contrast, for asymmetrically compressed 3D videos, blurring on the lower quality views *improves* the perceptual 3D video quality when the quality difference of left- and right-view is high. Generally, the improvement increases with the level of blurring and with the quality difference between the higher view and the lower view. Table 5.8 also includes the cases of asymmetrically compressed stereoscopic videos with pre-downsampling by factors of 2 and 4, where it can be seen that this 3DVQ-MOS improvement is less pronounced especially for the case of pre-downsampling by 4. This analysis verifies that the adoption of certain postprocessing techniques such as blurring could improve the efficiency of stereoscopic video coding but may not always work well for the cases of pre-downsampling.

Table 5.7: 3DVQ-MOS changes after applying different levels of Gaussian blurring as postprocessing on Waterloo-IVC 3D Video Database Phase I

Combinations		3DVQ-MOS Changes		
QP_l	QP_h	P1 $\sigma = 3.5$	P3 $\sigma = 7.5$	P4 $\sigma = 11.5$
25	25	-10.96	-16.15	-11.35
35	35	-0.19	-7.31	-4.81
40	40	-3.46	+0.19	-1.35
50	50	+1.54	+1.15	-0.96
25	35	-2.50	-3.27	-1.35
25	40	+7.69	+6.92	+7.12
25	45	+2.12	+16.92	+14.23
25	50	+14.81	+31.35	+27.12
35	45	+11.92	+13.08	+17.50
35	50	+11.35	+24.23	+29.62
40	50	+7.31	+16.35	+19.42

Table 5.8: 3DVQ-MOS changes after applying different levels of Gaussian blurring as post-processing on Waterloo-IVC 3D Video Database Phase II

Original Resolution					
Combinations		2DVQ-MOS		3DVQ-MOS Changes	
QP_l	QP_h	$2DVQ_h$	$2DVQ_l$	P1 $\sigma = 3.5$	P2 $\sigma = 5.5$
30	30	97.85	97.85	-12.58	-11.61
35	35	90.04	90.04	-7.40	-7.40
40	40	69.10	69.10	-1.67	-5.32
45	45	48.75	48.75	+9.68	+6.40
30	35	97.85	90.04	-1.93	-3.76
30	40	97.85	69.10	+5.72	+3.62
30	45	97.85	48.75	+7.80	+11.40
35	40	90.04	69.10	+1.66	+2.52
35	45	90.04	48.75	+5.54	+5.48
40	45	69.10	48.75	+2.35	+0.90
Downsample by 2					
Combinations		2DVQ-MOS		3DVQ-MOS Changes	
QP_l	QP_h	$2DVQ_h$	$2DVQ_l$	P1 $\sigma = 3.5$	P2 $\sigma = 5.5$
30	D2-25	97.85	85.75	-1.34	-3.66
30	D2-30	97.85	76.83	-3.76	-5.22
30	D2-35	97.85	61.18	+0.43	-0.91
35	D2-30	90.04	76.83	+1.18	-1.99
35	D2-35	90.04	61.18	+4.14	+3.39
35	D2-40	90.04	39.73	+6.45	+6.88
40	D2-35	69.10	61.18	+1.40	-0.81
40	D2-40	69.10	39.73	+2.10	+1.94
Downsample by 4					
Combinations		2DVQ-MOS		3DVQ-MOS Changes	
QP_l	QP_h	$2DVQ_h$	$2DVQ_l$	P1 $\sigma = 3.5$	P2 $\sigma = 5.5$
30	D4-25	97.85	57.58	+2.47	+0.81
30	D4-30	97.85	50.75	-0.97	-2.58
30	D4-35	97.85	39.73	+1.72	-1.02
35	D4-25	90.04	57.58	+0.59	-1.05
35	D4-30	90.04	50.75	-1.02	-2.69
35	D4-35	90.04	39.73	+1.94	-0.05
40	D4-25	69.10	57.58	+3.12	+3.49
40	D4-30	69.10	50.75	+5.97	+1.94

5.3.3 Rate-Distortion Performance of Mixed Distortions Asymmetric Stereoscopic Video

The third question we would like to ask is what is the rate-distortion (R-D) performance of different mixed-distortions asymmetric stereoscopic video coding schemes. Table 5.9 and Table 5.10 report the so-called Bjontegaard delta rate (BD-Rate) in terms of 3DVQ-MOS by comparing the mixed-distortions asymmetric stereoscopic video coding schemes with the symmetric coding method for each test sequence in Phase I and Phase II, respectively [13]. It can be seen that most mixed-distortions asymmetric stereoscopic video coding schemes achieve better R-D performance over the symmetric coding method (Group 3D.1.a in Phase I and Group 3D.3.a in Phase II).

Table 5.9: R-D Performance comparisons of asymmetric stereoscopic video coding in terms of 3DVQ-MOS on Waterloo-IVC 3D Video Database Phase I (Anchor: symmetric stereoscopic video coding)

3DVQ-MOS					
Group	Balloons	Book	Kendo	Lovebird	Average
3D.2.a	+88.3%	+37.5%	+55.8%	+164.2%	+86.4%
3D.2.b	+60.2%	+16.0%	+39.4%	+50.6%	+41.6%
3D.2.c	-16.8%	-0.6%	+23.2%	-26.1%	-5.1%
3D.2.d	-25.7%	-30.0%	-14.4%	-9.5%	-19.9%

For Phase I, the asymmetric compression-only scheme (Group 3D.2.a) degrades the R-D performance significantly due to the fact that 3D quality had a tendency towards the lower quality view with respect to blockiness [142]. However, this degradation is less pronounced and *is even reverted* when we increase the level of Gaussian low-pass post-filtering (Group 3D.2.b-d), but the optimal level of this postprocessing is content dependent.

For Phase II, similarly, the asymmetric compression plus postprocessing scheme (Group 3D.4.b-c) improves the R-D performance while the asymmetric compression-only scheme (Group 3D.4.a) reduces the R-D performance. It can also be observed that the proposed pre-downsampling plus postprocessing asymmetric coding schemes perform even better than the asymmetric compression plus postprocessing scheme, which is consistent with the

Table 5.10: R-D Performance comparisons of asymmetric stereoscopic video coding on Waterloo-IVC 3D Video Database Phase II (Anchor: symmetric stereoscopic video coding)

3DVQ-MOS							
Group	Barrier	Craft	Laboratory	Soccer	Tree	Undo	Average
3D.4.a	+8.4%	+10.1%	+11.3%	+3.9%	+7.9%	+16.8%	+9.7%
3D.4.b	-7.2%	-8.3%	-17.1%	-7.7%	-5.4%	-11.1%	-9.5%
3D.4.c	-10.5%	-12.9%	-14.4%	+0.6%	-9.1%	-12.6%	-9.8%
3D.5	-22.0%	-6.8%	-11.5%	-8.6%	-14.6%	-8.4%	-12.0%
3D.6.a	-13.8%	-8.2%	-9.0%	-17.9%	-9.9%	-9.7%	-11.4%
3D.6.b	-32.3%	-17.3%	-16.1%	-16.3%	-31.8%	-19.4%	-22.2%
3D.6.c	-17.8%	-22.1%	-11.3%	-13.7%	-25.4%	-13.9%	-17.4%
3D.7.a	-28.5%	-13.4%	-11.6%	-6.5%	-28.8%	-15.9%	-17.5%
3D.7.b	-27.0%	-14.5%	-17.2%	-14.3%	-25.6%	-14.3%	-18.8%
3D.7.c	-27.2%	-14.4%	-16.6%	-10.2%	-19.0%	-7.2%	-15.7%
Average 2DVQ-MOS							
Group	Barrier	Craft	Laboratory	Soccer	Tree	Undo	Average
3D.4.a	+12.3%	+13.4%	+11.2%	+23.8%	+19.0%	+22.3%	+17.0%
3D.4.b	+35.6%	+33.7%	+24.3%	+103.3%	+109.9%	+43.6%	+58.4%
3D.4.c	+148.9%	+79.0%	+56.9%	+208.5%	+405.1%	+97.1%	+165.9%
3D.5	-7.0%	-10.9%	-10.3%	+4.5%	-4.4%	-3.7%	-5.3%
3D.6.a	+5.9%	+7.0%	+2.7%	+34.5%	+17.3%	+19.9%	+14.5%
3D.6.b	+42.4%	+39.3%	+16.7%	+110.0%	+110.2%	+55.0%	+62.3%
3D.6.c	+132.0%	+98.7%	+68.7%	+194.3%	+366.2%	+97.7%	+159.6%
3D.7.a	+23.6%	+20.9%	+5.4%	+92.4%	+45.8%	+26.9%	+35.8%
3D.7.b	+64.4%	+62.3%	+38.9%	+124.9%	+149.1%	+60.3%	+83.3%
3D.7.c	+146.0%	+106.6%	+91.3%	+199.8%	+438.4%	+111.4%	+182.3%
Weighted 2DVQ-MOS							
Group	Barrier	Craft	Laboratory	Soccer	Tree	Undo	Average
3D.4.a	+10.1%	+10.0%	+9.6%	+19.4%	+13.0%	+17.3%	+13.2%
3D.4.b	-15.4%	-9.4%	-10.6%	-7.4%	-13.8%	-8.3%	-10.8%
3D.4.c	-9.6%	-5.2%	-12.6%	-2.8%	-12.6%	-1.9%	-7.4%
3D.5	-13.7%	-12.6%	-9.6%	-10.6%	-13.1%	-9.5%	-11.5%
3D.6.a	-16.1%	-10.1%	-10.0%	-8.2%	-16.3%	-5.4%	-11.0%
3D.6.b	-23.9%	-14.6%	-20.0%	-14.3%	-24.0%	-14.0%	-18.5%
3D.6.c	-21.0%	-10.0%	-18.0%	-12.5%	-21.9%	-11.0%	-15.7%
3D.7.a	-23.3%	-13.3%	-16.1%	-12.2%	-24.2%	-13.7%	-17.2%
3D.7.b	-24.0%	-10.8%	-15.5%	-16.2%	-24.0%	-12.8%	-17.2%
3D.7.c	-21.2%	-9.1%	-15.3%	-14.6%	-21.6%	-8.8%	-15.1%

previous subjective studies [3] that mixed-resolution stereoscopic video coding can achieve the best coding efficiency. By comparing different postprocessing levels, it is found that the scheme that applies pre-downsampling by a factor of 2 followed by postprocessing P1 ($\sigma = 3.5$), (i.e., Group 3D.6.b), provides the best overall performance. It is also interesting to note that for the case of no pre-downsampling or pre-downsampling by 2, the proposed postprocessing step increases the R-D performance significantly; while for the case of pre-downsampling by 4, the proposed postprocessing does not help in improving the R-D performance.

5.4 A Model for 2D-to-3D Quality Prediction

5.4.1 2D Video Quality Prediction

We first examine the capabilities of state-of-the-art 2D-IQA/VQA methods to predict perceptual quality of single-view 2D videos with different pre- and post- processing procedures. The tested full reference 2D-IQA/VQA methods include PSNR, SSIM, MS-SSIM, IW-SSIM, and VQM. Table 5.11 reports PLCC, SRCC and RMSE results between 2DVQ-MOS scores and 2D-IQA/VQA measurements, where it can be observed that IW-SSIM and VQM provide the most accurate quality predictions.

Table 5.11: Performance comparison of 2D-to-3D quality prediction models on Waterloo-IVC 3D Video Database Phase II

Method	PLCC	SRCC	KRCC	RMSE	MAE
PSNR	0.7122	0.6488	0.4816	16.8383	13.2672
SSIM	0.8371	0.7660	0.5853	13.1235	10.5683
MS-SSIM	0.7787	0.7116	0.5368	15.0502	12.1960
IW-SSIM	**0.8821**	**0.8474**	**0.6604**	**11.2982**	**9.0892**
VQM	**0.8997**	**0.8665**	**0.6926**	**10.4689**	**8.1712**

5.4.2 2D-to-3D Quality Prediction

The diagram of the proposed method is shown in Figure 5.10. Let ($I_{i,r,l}$, $I_{i,r,r}$) and ($I_{i,d,l}$, $I_{i,d,r}$) be the i-th left and right frames of the reference and compressed stereoscopic videos, respectively. We first create their local energy maps by computing the local variances at each spatial location, i.e., the variances of local image patches extracted around each spatial location, for which an 11×11 circular-symmetric Gaussian weighting function $\mathbf{w} = \{w_i | i = 1, 2, \cdots, N\}$ with standard deviation of 1.5 samples, normalized to unit sum ($\sum_{i=1}^{N} w_i = 1$), is employed. The resulting energy maps are denoted as $E_{i,r,l}$, $E_{i,r,r}$, $E_{i,d,l}$ and $E_{i,d,r}$, respectively. We then compute the local energy ratio maps in both views:

$$R_{i,l} = \frac{E_{i,d,l}}{E_{i,r,l}} \qquad \text{and} \qquad R_{i,r} = \frac{E_{i,d,r}}{E_{i,r,r}} \,. \tag{5.1}$$

The energy ratio maps provide useful local binocular rivalry information, which may be combined with the qualities of single-view frames to predict 3D quality. A pooling stage is necessary for this purpose. High-energy image regions are likely to contain more information content. Based on the principle exploited in [154], if the ultimate goal of visual perception is to efficiently extract useful information from the visual scene, then the more informative regions are more likely to attract visual attention, and thus should be given more importance. The modeling in [154] suggests more informative regions typically have higher energy. To emphasize on the importance of high-energy image regions in binocular rivalry, we adopt an energy weighted pooling method [156] given by

$$g_{i,l} = \frac{\sum E_{i,d,l} R_{i,l}}{\sum E_{i,d,l}} \qquad \text{and} \qquad g_{i,r} = \frac{\sum E_{i,d,r} R_{i,r}}{\sum E_{i,d,r}} \,, \tag{5.2}$$

where the summations are over the full energy and ratio maps. Here $g_{i,l}$ and $g_{i,r}$ are estimations of the level of dominance of the i-th left and right frames, respectively. Let N denotes the frame number of the entire 3D video sequence, we compute

$$g_l = \frac{1}{N} \sum_{i=1}^{N} g_{i,l} \qquad \text{and} \qquad g_r = \frac{1}{N} \sum_{i=1}^{N} g_{i,r} \,, \tag{5.3}$$

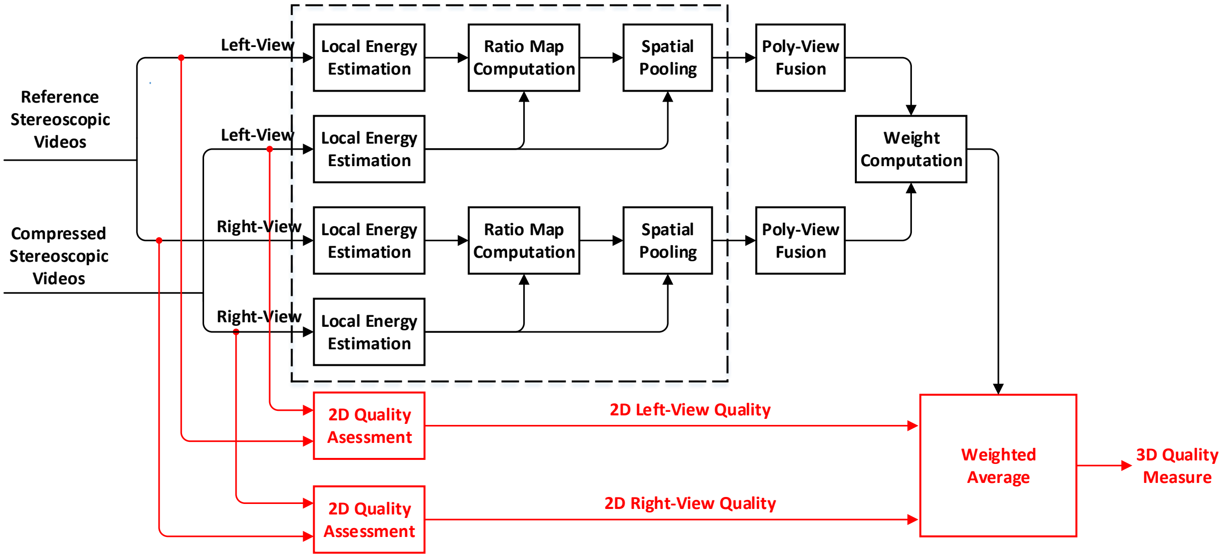


Figure 5.10: Diagram of the proposed 2D-to-3D quality prediction model.

where g_l and g_r denote the level of dominance of the left- and right-view video, respectively.

If we consider a video signal as 3-D volume data, then it can also be viewed from the side or the top. This has been explored by the poly-view fusion method, which has been shown as a simple and effective strategy to account for the temporal correlation and motion information contained in video signals [172, 171]. Instead of only estimating the level of dominance form the front-view, here we apply a poly-view fusion strategy to estimate the overall level of dominance from the front-view, the top-view, and the side-view together. We first compute the levels of dominance using Equation (5.1) to Equation (5.3) for the front-view, the top-view, and the side-view, separately, and denote them as g_l^F, g_l^T and g_l^S for the left-view video, and g_r^F, g_r^T and g_r^S for the right-view video, respectively. Then the overall level of dominance of the left- and right-view video after poly-view fusion is computed as

$$g_l^O = g_l^F + g_l^T + g_l^S \qquad \text{and} \qquad g_r^O = g_r^F + g_r^T + g_r^S \,, \tag{5.4}$$

respectively.

Given the values of g_l^O and g_r^O, the weights assigned to the left- and right-view videos are given by

$$w_l = \frac{{g_l^O}^2}{{g_l^O}^2 + {g_r^O}^2} \qquad \text{and} \qquad w_r = \frac{{g_r^O}^2}{{g_l^O}^2 + {g_r^O}^2}\,, \tag{5.5}$$

respectively.

Finally, the overall prediction of 3D video quality is calculated by a weighted average of the left- and right-view video quality:

$$Q^{3D} = w_l Q_l^{2D} + w_r Q_r^{2D}\,, \tag{5.6}$$

where Q_l^{2D} and Q_r^{2D} denote the 2D video quality of the left- and right-view videos, respectively.

5.4.3 Validation

The proposed 2D-to-3D video quality prediction model is tested on all 3D videos in the new database by applying it to the ground truth 2DVQ-MOS scores (Phase II only) and different base 2D-IQA approaches (Phase I and Phase II).

The PLCC, SRCC and RMSE values between 3DVQ-MOS and the predicted Q^{3D} value are given in Table 5.4 (all videos), Table 5.5 (different groups in Phase I) and Table 5.6 (different groups in Phase II). The corresponding scatter plots are shown in Figure 5.4 to Figure 5.9. It can be observed that the proposed 2D-to-3D model outperforms the direct averaging method significantly with respect to 2DVQ-MOS scores and all tested 2D-IQA/VQA approaches. For different levels of compressions, pre- and post- processing, the proposed method, which does not attempt to recognize the distortion types or give any specific treatment, removes or significantly reduces the 2D-to-3D quality prediction biases.

Furthermore, the R-D performance of different variations of asymmetric stereoscopic video coding in terms of the average and weighted 2DVQ-MOS for each test sequence in Phase II are reported in Table 5.10. Again, we use the BD-Rate as the test criterion, which provides a useful quantitative measure to evaluate the R-D performance. From Table 5.10, it can be seen that, compared with the bit rate savings measured by 3DVQ-MOS, the direct averaging 2DVQ-MOS generates substantial bias for all sequences and all mixed-distortions combinations. On the other hand, the proposed weighting 2DVQ-MOS significantly reduces the biases and indicates highly consistent bit rate savings with 3DVQ-MOS. This demonstrates great potentials of the proposed method to be employed in perceptually inspired R-D optimization of stereoscopic video coding systems.

5.5 Summary

The major contributions of this chapter are as follows: first, we carried out subjective quality assessment experiments on two databases (Waterloo-IVC 3D Video Database Phase I and Phase II) that contain various asymmetrically compressed stereoscopic 3D videos obtained from mixed-resolution coding, asymmetric transform-domain quantization coding, their combinations, and multiple choices of postprocessing techniques. Second, we compared different mixed-distortions asymmetric stereoscopic video coding schemes with symmetric coding methods and verified their potential coding gains. Third, we observed a strong systematic bias when using direct averaging of 2D video quality of both views to predict 3D video quality. Fourth, we proposed a model to account for the prediction bias, leading to significantly improved quality predictions of stereoscopic videos. Fifth, we show that the proposed model can help us predict the coding gain of mixed-distortions asymmetric video compression, and provides new insight on the development of high efficiency 3D video coding schemes.

Chapter 6

Quality Assessment of Multi-View-Plus-Depth Images

In this chapter, we focus on quality assessment of MVD Images. MVD representation has gained significant attention recently as a means to encode 3D scenes, allowing for intermediate views to be synthesized on-the-fly at the display site through DIBR. Automatic quality assessment of MVD images/videos is critical for the optimal design of MVD image/video coding and transmission schemes. Existing IQA methods are applicable only after the DIBR processes. Such post-DIBR measures are valuable in assessing the overall system performance, but are difficult to be directly employed in the rate-distortion optimization process in MVD image/video coding. Here we make one of the first attempts to develop a pre-DIBR IQA approach for MVD images by employing an information content weighted approach that balances between local quality measures of texture and depth images. Experiment results show that the proposed approach achieves competitive performance when compared with stat-of-the-art IQA algorithms applied post-DIBR.

6.1 Introduction

Many advanced 3D video systems are based on MVD representation [91], where typically two or three views of texture and depth videos are encoded. This allows for intermediate views to be synthesized on-the-fly at the display site from the decoded texture and depth views by means of DIBR [65]. Objective quality assessment of 3D synthesized images/videos from DIBR is a challenging problem [16]. Oh *et al.* defined a rendering view distortion function and presented an efficient depth map coding scheme [94]. Fang *et al.* proposed an analytical model to estimate the synthesized view quality in 3D video with the consideration of texture image characteristics and quality, depth map errors and rendering process [38]. Liu *et al.* conducted a subjective study of synthesized single-view videos with texture and depth compression and proposed an objective quality assessment algorithms for synthesized videos with emphasis on temporal flicker distortion induced by depth compression and view synthesis processes [75]. These existing studies are valuable but limited in one common aspect, i.e., the quality assessment process is applicable only after the DIBR process. The drawback of such methods, which we call post-DIBR approaches, is that they are difficult to be directly employed in the design and optimization of MVD based 3D video coding systems, because the actual texture and depth video signals being encoded and transmitted are the views before DIBR. Jang *et al.* proposed a fast quality metric purely for depth maps without view synthesis [57]. However, the impact of texture distortions and the interaction between texture and depth distortions are not taken into consideration, which is critical in guiding MVD coding schemes for optimal bit allocation.

The main purpose of this work is to develop a pre-DIBR IQA approach for MVD images, aiming for establishing a more convenient IQA model that can be used in the design of MVD coding schemes. A diagram that explains the difference between the two types (pre-DIBR and post-DIBR) of MVD quality assessment problems, as well as how multiple texture and depth images are used to synthesize stereoscopic 3D images are shown in Figure 6.1. In fact, in addition to all existing stereoscopic 3D-IQA models, any 2D-IQA method may also be applied for post-DIBR quality assessment by averaging 2D-IQA results of stereoscopic views. But none of them has been shown to be useful in the pre-DIBR case. Our work starts by observing how human subjects evaluate synthesized MVD images, how post-DIBR methods perform in predicting subjective quality, and how the performance varies depending on distortion types. Such observations help us develop our pre-DIBR algorithm, which demonstrates competitive performance against post-DIBR approaches.

6.2 Post-DIBR Quality Assessment

Subjective testing is critical in understanding IQA problems and validating IQA models. A highly valuable subjective study was introduced in [132], which resulted in an MCL-3D Image Database for 3D-IQA using 2D-image-plus-depth source. The database was created from nine pristine image-plus-depth source contents shown in Figure 6.2 and Figure 6.3. The resolution of texture and depth images is 1920×1080 or 1024×768. Each texture and depth image was altered by six types of distortions: additive white Gaussian noise contamination, Gaussian blur, downsampling blur, JPEG compression, JP2K compression and transmission loss. Each distortion type had four distortion levels. Three types of rendering combinations, i.e., distorted texture images and original depth maps (Texture-Distortion-Only), original texture images and distorted depth maps (Depth-Distortion-Only), and distorted texture images and distorted depth maps (Texture-Depth-Distortion), are used as the input into the DIBR software to render the distorted stereopairs. In total, there are 657 rendered stereoscopic images (including 9 "original" stereopairs). Figure 6.1 shows the building process of MCL-3D Image Database. Pair-wise comparison was adopted in the subjective test and the mean opinion score (MOS) was computed for each distorted rendered stereopair. More detailed descriptions of this database and the subjective experiment can be found in [132].

Post-DIBR methods directly work on synthesized stereoscopic pairs. Let ($R_{o,l}$, $R_{o,r}$) and ($R_{d,l}$, $R_{d,r}$) be the left- and right-view image pairs of the rendered stereoscopic images from the original and distorted texture images and depth maps, respectively. For the post-DIBR case, we are interested in investigating the appropriateness of existing 2D/3D-IQA methods to predict 3D quality of synthesized views by comparing ($R_{o,l}$, $R_{o,r}$) and ($R_{d,l}$, $R_{d,r}$).

We first test 2D-IQA methods, which can be applied to the left- and right-view images

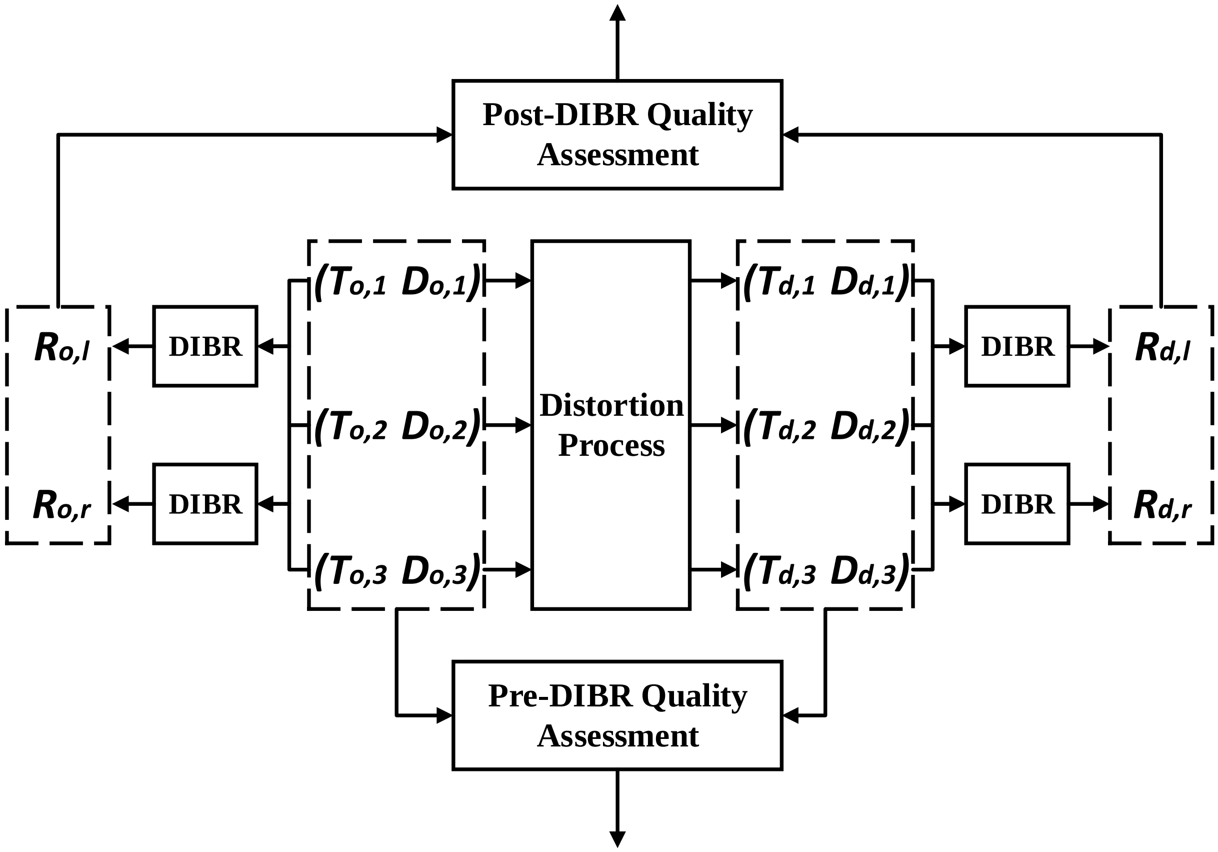
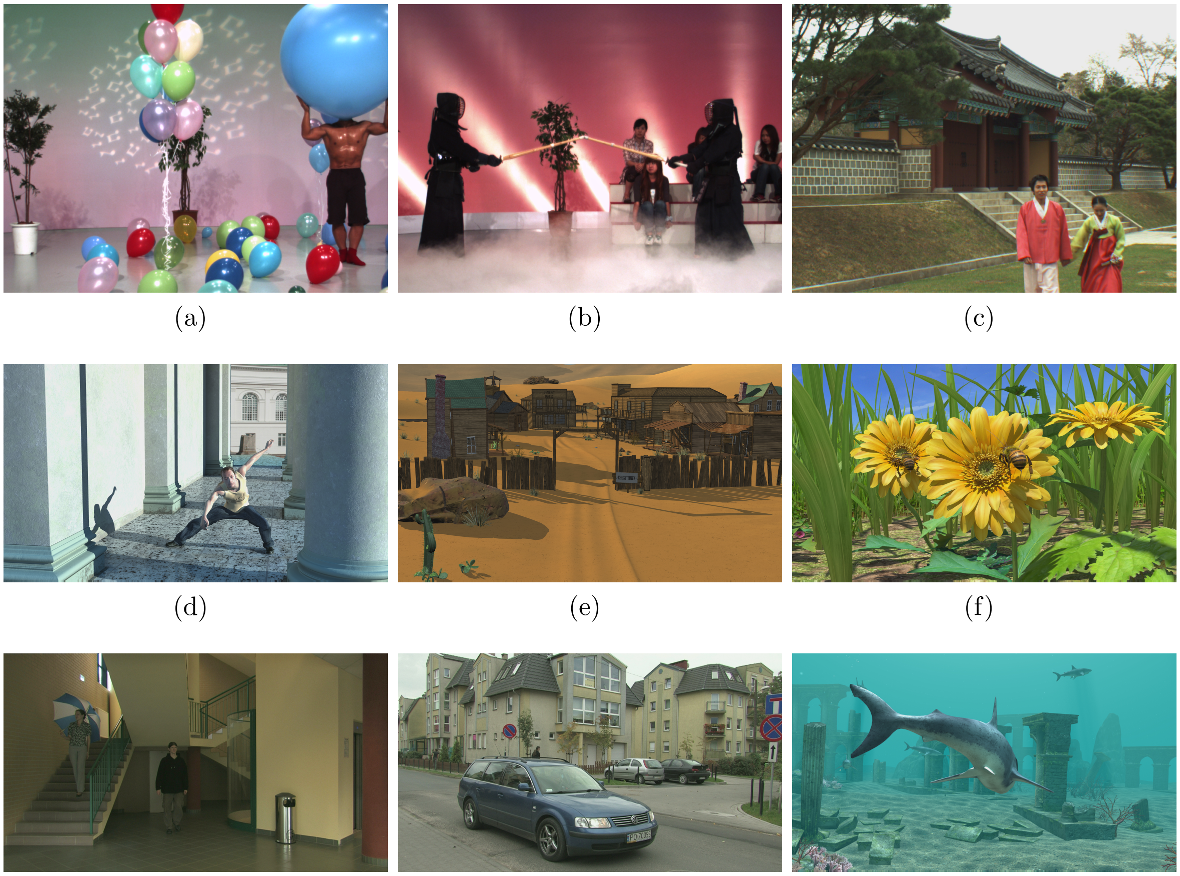


Figure 6.1: Two types of quality assessment of stereoscopic 3D images rendered from MVD representations. $(T_{o,1},\ D_{o,1})$, $(T_{o,2},\ D_{o,2})$, $(T_{o,3},\ D_{o,3})$ and $(T_{d,1},\ D_{d,1})$, $(T_{d,2},\ D_{d,2})$, $(T_{d,3},\ D_{d,3})$ are the (texture, depth) images for 3 views in the original and distorted MVD representations, respectively. $(R_{o,l},\ R_{o,r})$ and $(R_{d,l},\ R_{d,r})$ are the (left, right) views of DIBR-synthesized original and distorted stereoscopic 3D images, respectively. Type 1: post-DIBR 3D-IQA; Type 2: pre-DIBR 3D-IQA.



(a) (b) (c)

(d) (e) (f)

(g) (h) (i)

Figure 6.2: All texture images in MCL-3D database. (a) Balloon. (b) Kendo. (c) Lovebird. (d) Dancer. (e) GtFly. (f) MicroWorld. (g) PoznanHall. (h) PoznanStreet. (i) Shark.

independently and then averaged to predict 3D quality. Previous studies suggested that in the case of symmetric distortion of both views (in terms of both distortion types and levels), simply averaging state-of-the-art 2D-IQA measures of both views is sufficient to provide reasonably accurate quality predictions of stereoscopic images [23, 142] and stereoscopic videos [143]. Note that multi-view rendering and coding schemes usually generate symmetrically distorted or compressed stereopairs. The methods being tested include FR methods PSNR, SSIM [153], MS-SSIM [158], IW-SSIM [154] and IDW-SSIM [145], and NR methods BIQI [88], BLIINDS-II [109], BRISQUE [86], CORNIA [167], DIIVINE [89],

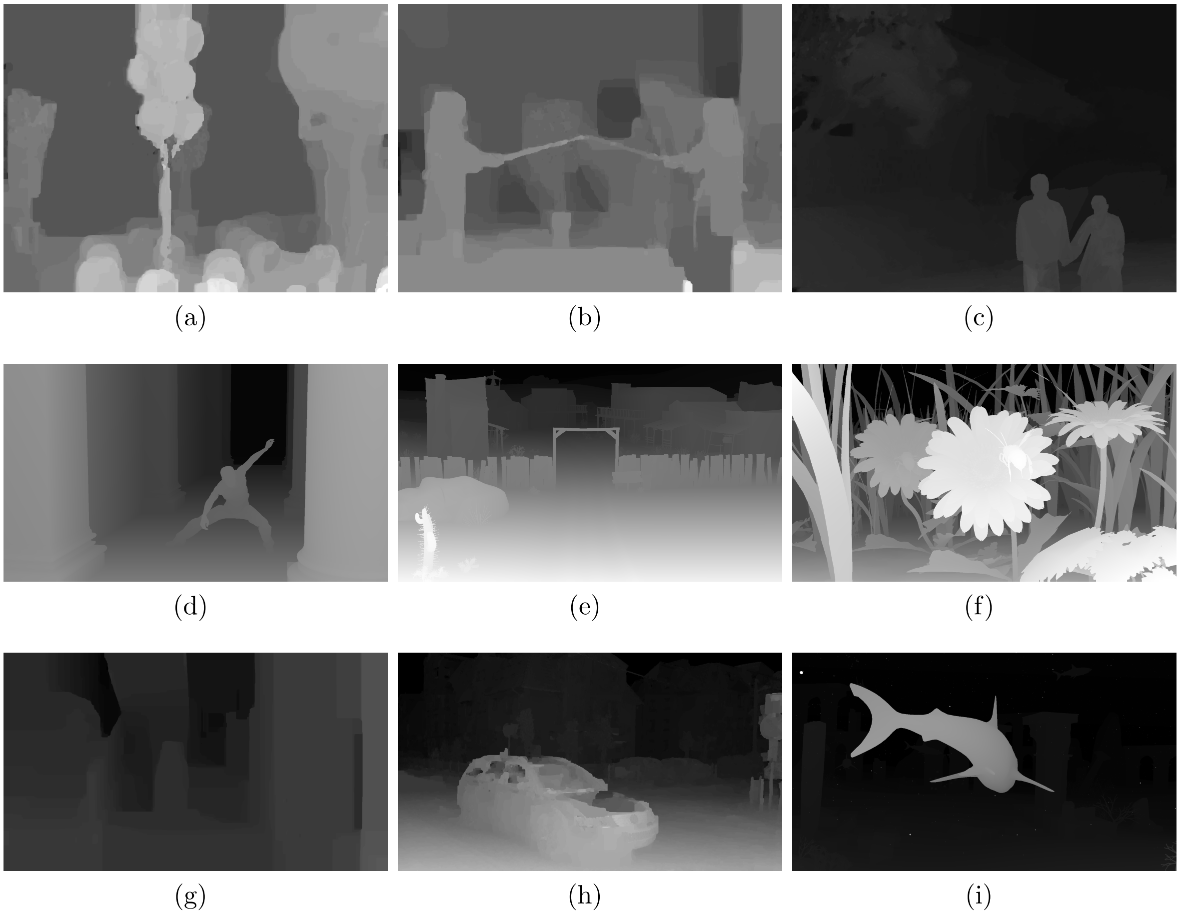
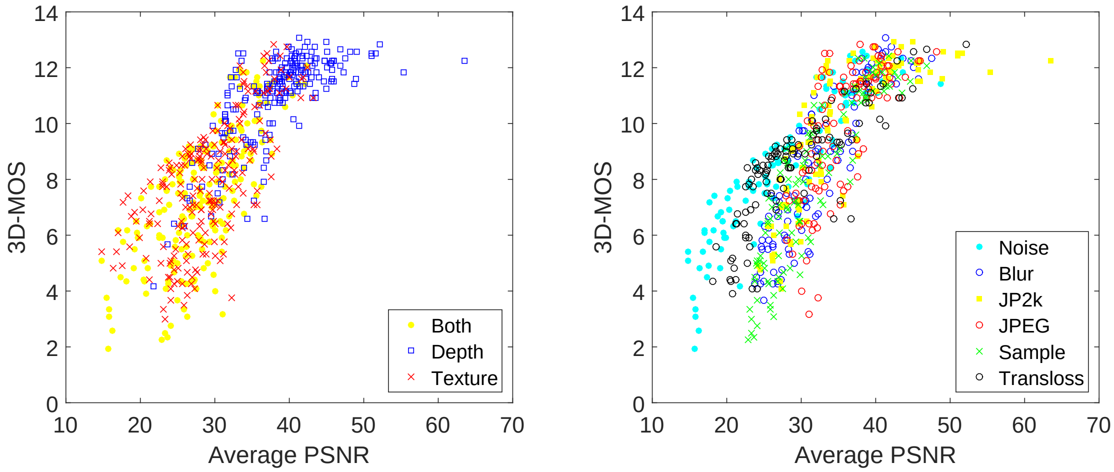

(a) (b) (c)
(d) (e) (f)
(g) (h) (i)

Figure 6.3: All depth maps in MCL-3D database. (a) Balloon. (b) Kendo. (c) Lovebird. (d) Dancer. (e) GtFly. (f) MicroWorld. (g) PoznanHall. (h) PoznanStreet. (i) Shark.

LPSI [160], M_3 [161], NIQE [87], QAC [162] and TCLT [159]. Table 6.1 reports PLCC, SRCC, KRCC, RMSE and MAE between 3D-MOS scores and the average 2D-IQA measurements and Figure 6.4 to Figure 6.8 show the corresponding scatter plots for PSNR, SSIM, MS-SSIM, IW-SSIM and IDW-SSIM. From Table 6.1 and Figure 6.4 to Figure 6.8, it can be observed that directly averaging the FR 2D-IQA measures of both views provides accurate image quality predictions of rendered stereopairs, which is consistent with previous findings in [23, 142, 143]. However, if the "reference" stereopairs are not available, there is a large drop in the performance from FR to NR models.

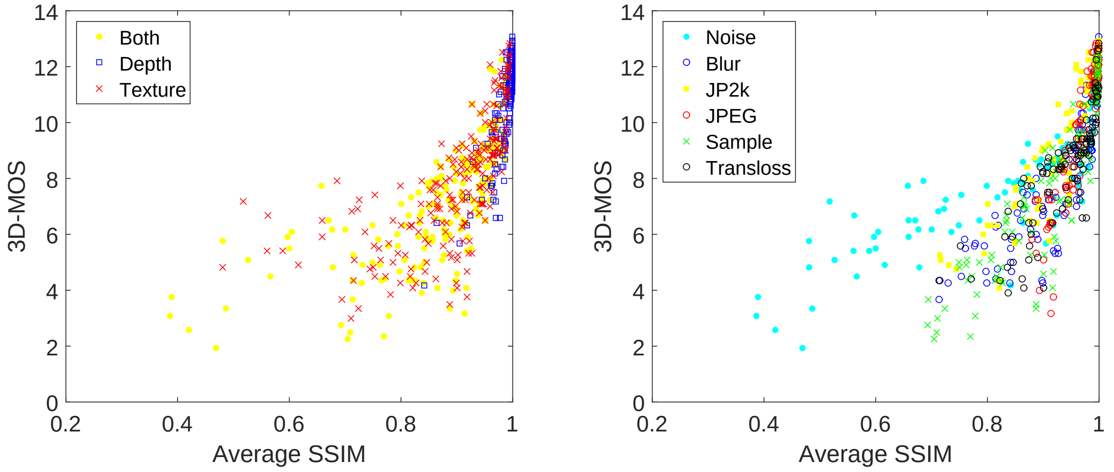
Table 6.1: Performance Comparison of 2D/3D-IQA Models on the MCL-3D Image Database

Category		Method	PLCC	SRCC	KRCC	RMSE	MAE
Post-DIBR	FR 2D-IQA	PSNR	0.8320	0.8405	0.6406	1.4435	1.1383
		SSIM [153]	0.8935	0.9034	0.7201	1.1682	0.9214
		MS-SSIM [158]	0.8644	0.8750	0.6837	1.3081	1.0197
		IW-SSIM [154]	0.9197	0.9260	0.7564	1.0213	0.8079
		IDW-SSIM [145]	0.8682	0.8792	0.6907	1.2910	1.0314
	NR 2D-IQA	BIQI [88]	0.3370	0.3135	0.2029	2.4495	2.0468
		BLIINDS-II [109]	0.5046	0.4994	0.3440	2.2461	1.8395
		BRISQUE [86]	0.6860	0.6472	0.4559	1.8930	1.5424
		CORNIA [167]	0.7401	0.7344	0.5338	1.7495	1.3900
		M_3 [161]	0.5900	0.5059	0.3498	2.1007	1.7413
		DIIVINE [89]	0.5747	0.4169	0.2779	2.1292	1.7402
		LPSI [160]	0.5027	0.4490	0.2970	2.2490	1.8744
		NIQE [87]	0.7430	0.5994	0.4431	1.7412	1.3909
		QAC [162]	0.6534	0.6475	0.4564	1.9695	1.5981
		TCLT [159]	0.6371	0.4871	0.3324	2.0054	1.6352
	FR 3D-IQA	Benoit [10]	0.6330	0.6359	0.4562	2.0142	1.6316
		Chen [25]	0.8528	0.8587	0.6627	1.3586	1.0726
		Lin [74]	0.8700	0.8690	N/A	1.2850	N/A
		Shao [121]	0.8480	0.8530	N/A	1.3800	N/A
		Yang [164]	0.8346	0.8445	0.6455	1.4331	1.1255
		You [168]	0.8922	0.9037	0.7205	1.1750	0.9271
		Zhang [174]	0.9340	0.9390	N/A	0.9300	N/A
Pre-DIBR		Proposed	0.9107	0.9164	0.7433	1.0745	0.8237



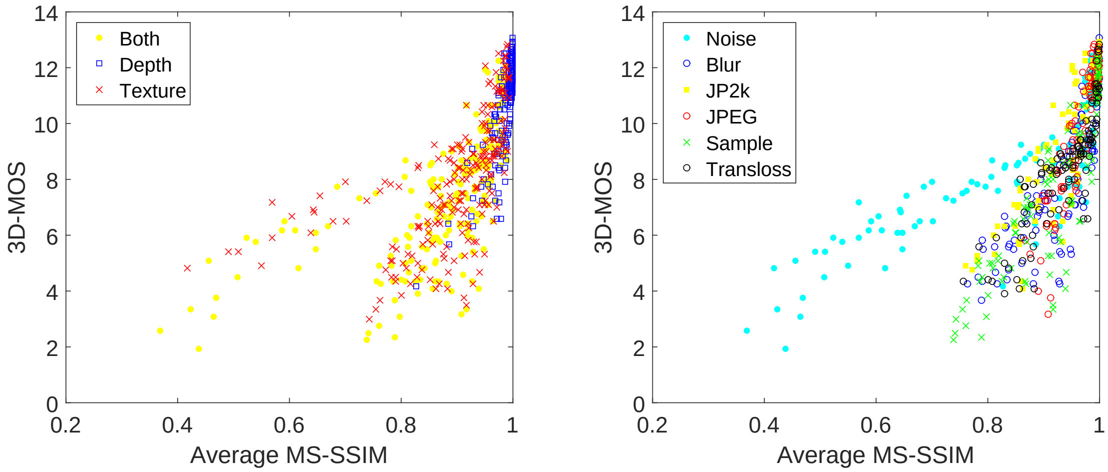
(a) (b)

Figure 6.4: 3D-MOS versus predictions from PSNR of 2D left- and right-views. (a) By different rendering combinations. (b) By different distortion types.



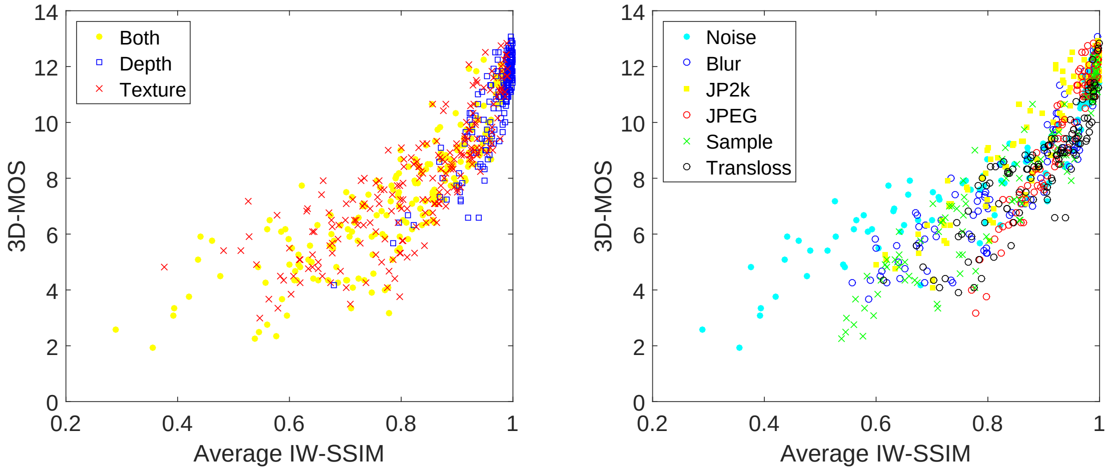
(a) (b)

Figure 6.5: 3D-MOS versus predictions from SSIM of 2D left- and right-views. (a) By different rendering combinations. (b) By different distortion types.



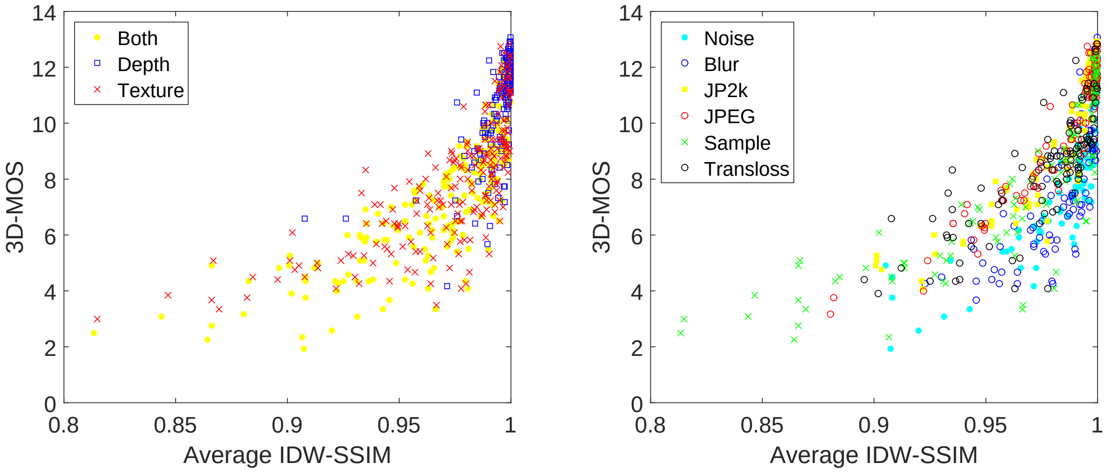
(a) (b)

Figure 6.6: 3D-MOS versus predictions from MS-SSIM of 2D left- and right-views. (a) By different rendering combinations. (b) By different distortion types.



(a) (b)

Figure 6.7: 3D-MOS versus predictions from IW-SSIM of 2D left- and right-views. (a) By different rendering combinations. (b) By different distortion types.



(a) (b)

Figure 6.8: 3D-MOS versus predictions from IDW-SSIM of 2D left- and right-views. (a) By different rendering combinations. (b) By different distortion types.

We also test some recent methods that are designed for FR 3D-IQA. The results on MCL-3D database are given in Table 6.1, where it can be seen that most FR 3D-IQA methods do not show superiority over FR 2D-IQA methods. The best result is obtained by the 3D-MAD method [174], which demonstrates similar performance as compared to IW-SSIM [154], a purely 2D algorithm.

6.3 Pre-DIBR Quality Assessment

6.3.1 Distortion type dependency

As mentioned earlier, in the rendering process of MCL-3D Image Database, three types of combinations, i.e., Texture-Distortion-Only, Depth-Distortion-Only, and Texture-Depth-Distortion, were adopted to create different kinds of distorted stereopairs. Figure 6.9 shows 3D-MOS scores for different distortion types and different rendering combinations. We compute SSIM between $(R_{o,l}, R_{d,l})$ to measure local quality/distortion of the distorted rendered left-view images. The resultant SSIM index maps for the cases of Texture-Depth-Distortion, Texture-Distortion-Only and Depth-Distortion-Only are denoted as $S_{R,l}^{T+D}$, $S_{R,l}^{T}$ and $S_{R,l}^{D}$, respectively, as shown in Column (a), (b) and (c) in Figure 6.10, where brightness indicates the magnitude of the local SSIM index (i.e., brighter = better quality).

From Figure 6.9 and Figure 6.10, it can be observed that there exists a specific distortion type dependency with respect to 3D-MOS scores for different rendering combinations. For blur, JP2K, JPEG and downsample distortion, 3D-MOS scores for Texture-Depth-Distortion and Texture-Distortion-Only images are very close, which both are strictly decreasing with the increase of distortion level. For Depth-Distortion-Only images, 3D-MOS scores are always at a high quality level and not falling down with the increasing level. Similar observations can be found in Figure 6.10. For Texture-Depth-Distortion and Texture-Distortion-Only images at Column (a) and (b), the SSIM maps exhibit almost the same local spatial variations and intensity levels, indicating a very close overall quality for both cases; for Depth-Distortion-Only images, a significantly better overall quality is presented as darker pixels only appear in a very small region. For noise and transmission loss, 3D-MOS scores lie in different quality levels for different rendering combinations with the increasing order of Texture-Depth-Distortion, Texture-Distortion-Only and Depth-Distortion-Only. Also, 3D-MOS scores decrease with the increase of distortion levels in all cases. While in Figure 6.10, the SSIM maps for Texture-Depth-Distortion and Texture-Distortion-Only images exhibit similar spatial variations but relative brighter pixels can be seen in the Texture-Distortion-Only case, indicating a better overall quality; for Depth-Distortion-Only images, an even better overall quality compared to the Texture-Distortion-Only images is pronounced as darker pixels are significantly less visible.

In general, the distortions from noise contamination and transmission loss affect the 3D quality of rendered stereopairs more significantly than from blur, JP2K, JPEG and downsampling. Another important observation is that the impact of texture image distortions are much stronger compared to depth map distortions.

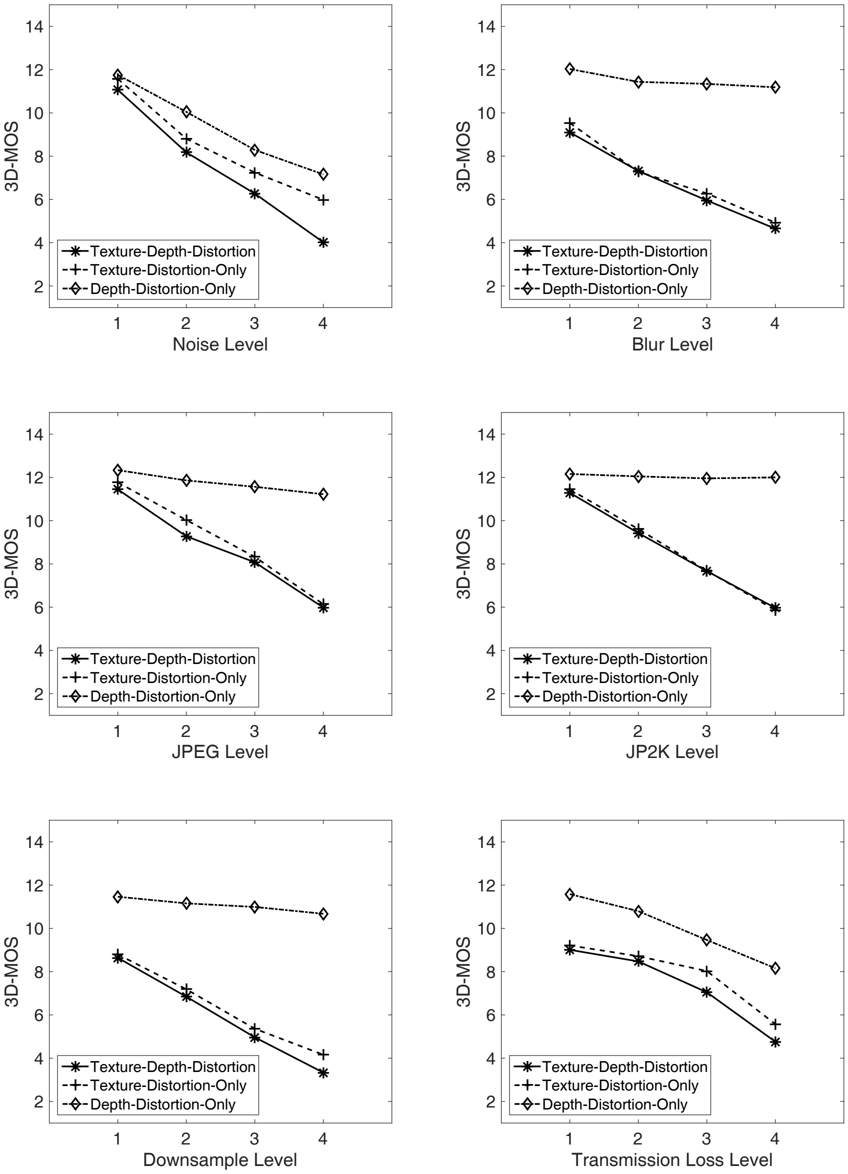
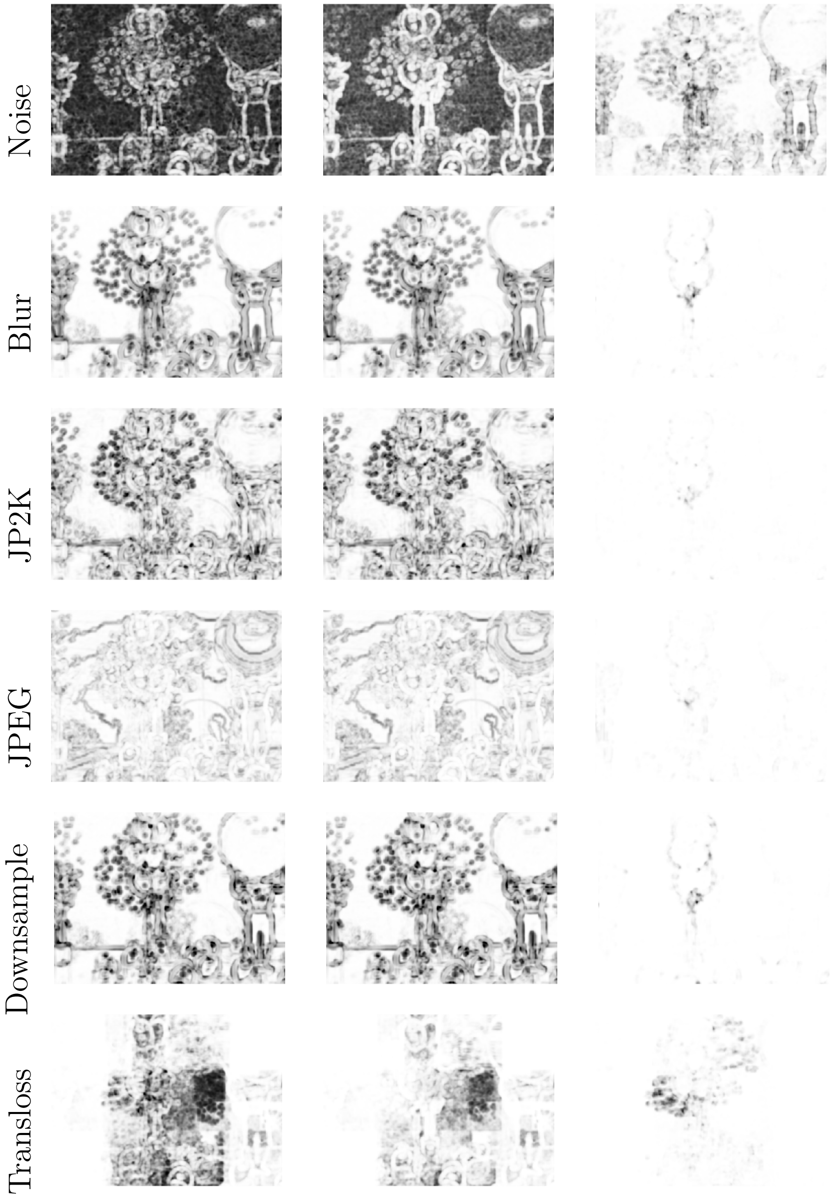


Figure 6.9: 3D-MOS scores for different distortion types and different rendering combinations. (a) Noise. (b) Blur. (c) JPEG. (d) JP2K. (e) Downsample. (f) Transmission Loss.



(a) (b) (c)

Figure 6.10: SSIM maps of the rendered Balloon images for all distortion types at level 4. (a) Rendered, Texture-Depth-Distortion. (b) Rendered, Texture-Distortion-Only. (c) Rendered, Depth-Distortion-Only.

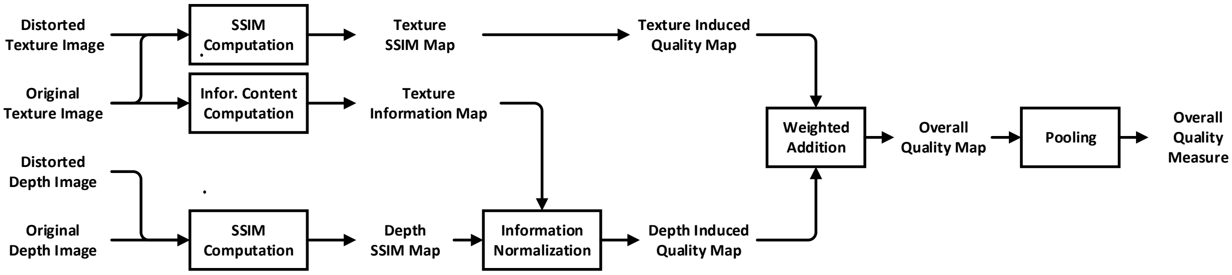


Figure 6.11: Diagram of the proposed 2D-to-3D quality prediction model.

6.3.2 Pre-DIBR 3D-IQA algorithm

Let $(T_{o,1}, D_{o,1})$, $(T_{o,2}, D_{o,2})$, $\cdots$, $(T_{o,M}, D_{o,M})$ be the original texture images and their associated depth maps, and $(T_{d,1}, D_{d,1})$, $(T_{d,2}, D_{d,2})$, $\cdots$, $(T_{d,M}, D_{d,M})$ be the corresponding distorted texture images and depth maps for M views. Figure 6.1 gives an example for the case of $M = 3$. We propose a top-down model to predict the 3D quality of synthesized views before DIBR. Specially, the prediction of Q^{3D} is calculated by directly averaging each view's quality Q_k^{3D}:

$$Q^{3D} = \frac{1}{M} \sum_{k=1}^{M} \left(Q_k^{3D}\right) . \tag{6.1}$$

The procedure to compute each view's quality Q_k^{3D} is shown in Figure 6.11. We define an overall quality map $S_{O,k}$, which is a combination of the k-th view's texture induced quality map $S_{T,k}$ and depth induced quality map $S_{D,k}$:

$$S_{O,k} = w_{T,k} S_{T,k} + w_{D,k} S_{D,k} , \tag{6.2}$$

where $w_{T,k}$ and $w_{D,k}$ are the weights assigned to $S_{T,k}$ and $S_{D,k}$, respectively. Q_k^{3D} is obtained by spatially average pooling over $S_{O,k}$.

The key step is to determine the adaptive weighting factors $w_{T,k}$ and $w_{D,k}$. Here we use an information content weighting approach, where the perceived local information content is quantified as the number of bits that can be received from a statistical image information source that passes through a noisy visual channel. Assume that the source power is P and the channel noise power is C. The mutual information between the source and destination

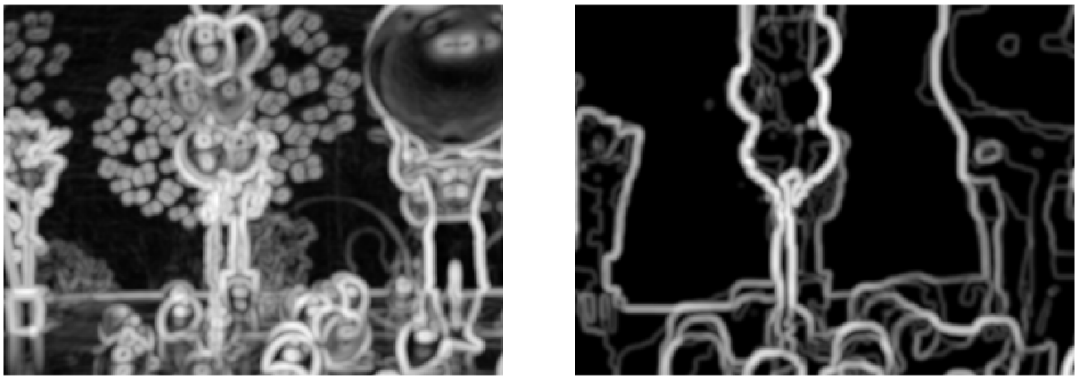

(a) (b)

Figure 6.12: Information maps of the pristine Balloon texture and depth images. (a) Texture Information Map. (b) Depth Information Map.

is

$$I = \frac{1}{2} \log \left(1 + \frac{P}{C} \right) . \tag{6.3}$$

Now assume that the source power of a local image patch can be estimated as $\sigma_{T,k}^2$ and $\sigma_{D,k}^2$ for texture images and depth maps, respectively, then the information maps from the k-th view's texture image and depth map are given by

$$I_{T,k} = \log \left(1 + \frac{\sigma_{T,k}^2}{C} \right) \quad \text{and} \quad I_{D,k} = \log \left(1 + \frac{\sigma_{D,k}^2}{C} \right) , \tag{6.4}$$

where $\sigma_{T,k}^2$ and $\sigma_{D,k}^2$ are the local variance maps by computing local variances at each spatial location. Given the information maps $I_{T,k}$ and $I_{D,k}$, the weights assigned to the k-th texture and depth induced quality maps are given by

$$w_{T,k} = \frac{\sum I_{T,k}}{\sum I_{T,k} + \sum I_{D,k}} \quad \text{and} \quad w_{D,k} = \frac{\sum I_{D,k}}{\sum I_{T,k} + \sum I_{D,k}} , \tag{6.5}$$

where a spatially average pooling is applied to both $I_{T,k}$ and $I_{D,k}$.

For the k-th original texture image $T_{o,k}$ and distorted texture image $T_{d,k}$, SSIM is used to compute local quality/distortion and the SSIM index map $S_{T,k}$ is obtained. Sample $S_{T,1}$

maps are shown in Column (a) in Figure 6.13. For all distortion types, it can be observed that the SSIM maps $S_{T,1}$ and $S^T_{R,l}$ exhibit highly similar local spatial variations. Thus we denote $S_{T,k}$ as the texture induced quality map for the k-th view.

For the k-th original depth map $D_{o,k}$ and distorted depth map $D_{d,k}$, SSIM is again employed to compute local quality/distortion and the SSIM index map $S_{D,k}$ is obtained. Sample $S_{D,1}$ maps are shown in Column (b) in Figure 6.13. Unlike texture distortion, SSIM maps $S_{D,1}$ and $S^D_{R,l}$ look significantly different in terms of spatial variations. We found that the impact of depth distortions on the rendered image is more correlated with the information map of the corresponding texture image $I_{T,k}$. As a result, when depth distortion occurs in structural regions in the texture images, the quality of rendered images is more affected. Thus we apply a spatial normalization to $S_{D,k}$ with $I_{T,k}$ and denote the resultant map $S'_{D,k}$ as the depth induced quality map for the k-th view:

$$s'_{D,k,j} = \frac{i_{T,k,j} s_{D,k,j}}{\sum_{j=1}^{N} i_{T,k,j}}\,, \tag{6.6}$$

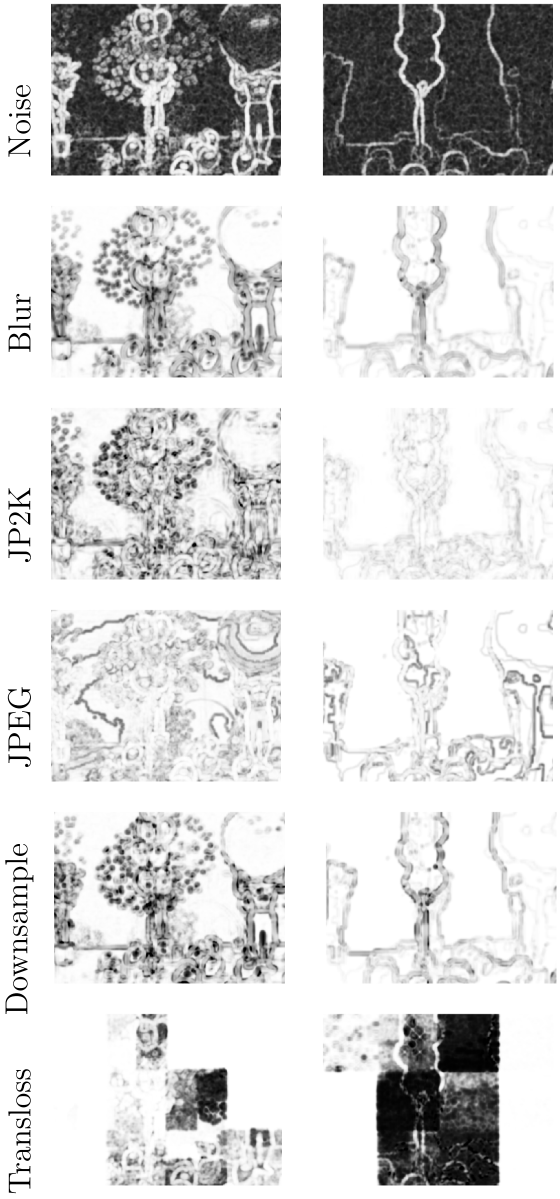
where $s_{D,k,j}$ and $s'_{D,k,j}$ are the local quality values in $S_{D,k}$ and $S'_{D,k}$, respectively, and $i_{T,k,j}$ is the weight assigned to the j-th spatial location (j-th pixel) from $I_{T,k}$. Subsequently, the overall quality map $S_{O,k}$ is combined as the weighted average between $S_{T,k}$ and $S'_{D,k}$ using Equation (6.2).

Finally, to emphasize the importance of the information content of texture images, a spatially weighted pooling with the texture information map $I_{T,k}$ is applied to the overall quality map $S_{O,k}$:

$$Q_k^{3D} = \frac{\sum_{j=1}^{N} i_{T,k,j} s_{O,k,j}}{\sum_{j=1}^{N} i_{T,k,j}}\,, \tag{6.7}$$

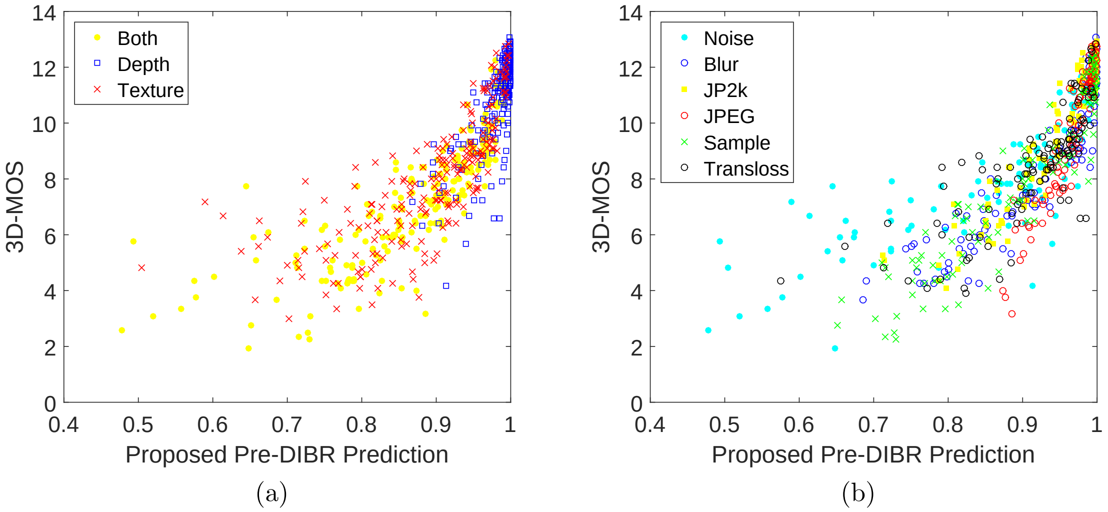
where $s_{O,k,j}$ and $i_{T,k,j}$ are the local quality value in $S_{O,k}$ and the weight assigned to the j-th spatial location from $I_{T,k}$, respectively.

We test the proposed algorithm on the MCL-3D Image Database. The PLCC, SRCC, KRCC, RMSE and MAE values between 3D-MOS and the predicted Q^{3D} value for all images are given in Table 6.1 and the corresponding scatter plots are shown in Figure 6.14, where we can see that the proposed method performs equally well or better than state-



(a) (b)

Figure 6.13: SSIM maps of the source Balloon images for all distortion types at level 4. (a) Source, Texture. (b) Source, Depth.



(a) (b)

Figure 6.14: 3D-MOS versus predictions from the proposed Pre-DIBR method. (a) By different rendering combinations. (b) By different distortion types.

of-the-art FR IQA models applied post-DIBR. Since it is applied pre-DIBR, it has great potentials to be employed in perceptually inspired rate-distortion optimization of MVD video coding systems.

6.4 Summary

We investigated the problem of objective quality assessment of MVD images, with a main focus on the pre-DIBR case. We found that although existing IQA methods can be applied post-DIBR to provide reasonable quality prediction of MVD images, they are difficult to be employed as a guiding criterion in the optimization of MVD video coding and transmission systems. We proposed a novel pre-DIBR method based on information content weighting of both texture and depth images. Experimental results show that the proposed method demonstrates competitive performance against state-of-the-art IQA models applied post-DIBR. Future work includes incorporating the proposed model in the rate-distortion optimization process of visual perception-based MVD 3D video coding schemes.

Chapter 7

Conclusion and Future Work

7.1 Conclusion

The goal of this thesis is to investigate and understand different aspects of human visual QoE of stereoscopic 3D images and videos. In this chapter, we first summarize the contributions to the scientific community that were brought forward in this thesis. We then discuss different avenues for future research. Publications during the period of the thesis are listed at the end of the chapter.

In Chapter 3, we focus on how to predict the quality of a stereoscopic 3D image from that of the 2D single-view images by carrying out a subjective quality assessment experiment on a database that contains both single-view images and stereoscopic images with symmetric and asymmetric distortion types and levels. The major contributions are as follows:

- A new subjective 3D-IQA database is created that has two unique features − the inclusion of both 2D and 3D images, and the inclusion of mixed distortion types;
- A strong distortion type dependent bias is observed when using the direct average of 2D image quality of both views to predict 3D image quality;

- A binocular rivalry inspired multi-scale 2D-to-3D model is proposed to predict the quality of stereoscopic images from that of its single-view 2D images, which without explicitly identifying image distortion types, successfully eliminates the prediction bias, leading to significantly improved quality prediction of stereoscopic 3D images;

- A binocular rivalry inspired multi-scale model to predict the quality of stereoscopic images from that of the single-view images without referring to the original left- and right-view images is proposed.

In Chapter 4, we demonstrate two subjective studies on depth perception of stereoscopic 3D images. The first one follows a traditional framework where subjects are asked to rate depth quality directly on distorted stereopairs. The second one uses a novel approach, where the stimuli are synthesized independent of the background image content and the subjects are asked to identify depth changes and label the polarities of depth. The major contributions are as follows:

- A novel subjective 3D depth experiment is carried out that are much more effective at singling out the contributions of stereo cues in depth perception;

- Several interesting findings are obtained regarding distortion type dependency, image content dependency, and the impact of symmetric and asymmetric distortions on the perception of depth;

- A novel DPDI index is developed based on the percentage of correct and incorrect subject judgements for quantifying depth perception induced by stereo cues;

- A novel computational model for DPDI prediction is proposed that leads to highly promising DPDI prediction performance without explicitly identifying image distortion types.

In Chapter 5, we focus on how to predict the quality of a stereoscopic 3D videos from that of the 2D single-view videos by carrying out subjective quality assessment experiments on two databases that contain various asymmetrically compressed stereoscopic 3D videos. The major contributions are as follows:

- Two new subjective 3D-VQA databases are built that have two unique features – the inclusion of both 2D and 3D videos, and the inclusion of asymmetrically compressed stereoscopic 3D videos obtained from mixed-resolution coding, asymmetric transform-domain quantization coding, their combinations, and multiple choices of postprocessing techniques;
- A comparison of different mixed-distortions asymmetric stereoscopic video coding schemes with symmetric coding methods is performed to verify their potential coding gains;
- A strong systematic bias when using direct averaging of 2D video quality of both views to predict 3D video quality is found;
- A model to predict the quality of stereoscopic video from that of its single-view videos is proposed, which leads to significantly improved quality prediction performance, which can help us predict the coding gain of mixed-distortions asymmetric video compression.

In Chapter 6, we investigate the problem of objective quality assessment of MVD images, with a main focus on the pre-DIBR case. We find that although existing IQA methods can be applied post-DIBR to provide reasonable quality prediction of MVD images, they are hard to be employed as a guiding criterion in the optimization of MVD video coding and transmission systems. We propose a novel pre-DIBR method based on information content weighting of both texture and depth images, which demonstrates competitive performance against state-of-the-art IQA models applied post-DIBR.

7.2 Future Work

The research work presented in this thesis aims to investigate various aspects of human visual QoE when viewing stereoscopic 3D images/videos and to develop objective quality assessment models that automatically predict visual QoE of 3D images/videos. Automatically assessing the quality of 3D visual experience is a challenging problem due to the sophistication and interaction between multiple 3D visual cues. The current work initialized several new ideas in this exciting area, but significant effort is still needed to have deep understandings of the problems and to provide useful practical solutions. Some of the possible directions for further research work are described as follows.

7.2.1 Research on Depth Quality, Visual Comfort and Overall 3D Quality-of-Experience

Subjective testing is important in understanding IQA problems and validating IQA models and becomes even more critical in the scenario of 3D-IQA. From Figure 4.4 (b), we show that human opinions on 3DIQ and DQ are highly correlated in the first subjective study. This is somewhat surprising because 3DIQ and DQ are two different perceptual attributes and the stimuli were generated to cover all combinations between picture qualities and stereo depths. Exactly the same observation can be also found with regard to 3DIQ and VC. Although many efforts were intended to reduce visual fatigue and discomfort of the subjects and to reduce the interference between different visual experience criteria (Section 3.3), the traditional experimental design is problematic because monocular cues and the spatial quality of images have strong impacts on the DQ/VC scores given by subjects. Psychologically, humans have the tendency to give high DQ/VC scores whenever the 3DIQ is good and vice versa. We then design and carry out the second subjective study where depth effect is synthesized at different depth levels before various types and levels of symmetric and asymmetric distortions are applied, in order to single out the contributions of stereo cues in depth perception. Consequently, a new distortion type dependency with respect to depth perception has been identified.

Some potential questions on depth quality, visual comfort and the overall 3D QoE are listed below. All of these are very challenging topics and the traditional IQA methodologies might not be sufficient to account for them. However, the successful experience from the depth quality study encourages us to look for possible novel approaches.

- Depth Quality: deeper understanding of the impact of monocular cues and binocular cues is lacking; the relationship between them also needs to be studied.

- Visual Comfortness: discussion on can/how we design a feasible subjective test? New experiments that are able to separate out the input of 3DIQ and depth perception need to be carefully designed.

- 3D Overall QoE: study on the tradeoff between perceptual 3DIQ and DQ is needed to account for the two different distortion type dependencies; the best way to combine 3DIQ, DQ and VC needs to be investigated.

7.2.2 Quality Assessment of Asymmetrically Distorted Videos

In Chapter 5, we propose a binocular rivalry inspired 2D-to-3D model to predict the quality of stereoscopic videos from that of its single-view 2D videos. Previously in vision science studies, most studies focus on the investigation of binocular fusion versus binocular rivalry merely with regard to stereoscopic images but none of them extends it to stereoscopic videos. The inconsistency induced by specific temporal artifacts such as flickering and ghosting [170, 173] between left- and right-view videos should not be overlooked and could also result in binocular rivalry. In Chapter 5, we apply a poly-view fusion strategy to estimate the overall level of dominance from the front-view, the top-view and the side-view together as a practical solution. In the future, more theoretical studies and experiments should be conducted to develop a complete 2D-to-3D video quality prediction model that can accurately account for the level of dominance both in spatial and temporal domain.

7.2.3 Quality Assessment of Multi-View-Plus-Depth Images

In Chapter 6, we investigate the problem of objective quality assessment of MVD images based on the MCL-3D database, which is highly valuable but limited in one aspect or another. It is desirable to build new MVD 3D images databases with the following new features: First, different rendering algorithms should be compared in the new subjective testing as such we can examine whether or not the impact of distortions in the DIBR processes is significant. Second, multiple synthesized 3D views should be included in the new database. Only one synthesized view is subject-rated in the existing MCL-3D database, making it difficult to examine and compare the potential quality difference across different views. Third, subjective tests on both 2D and 3D images should be performed. This allows us to directly examine the relationship between the perceptual quality of synthesized stereoscopic images and that of its single-view rendered images. Fourth, the selection of the distortion/degradation on the original texture images and depth maps should be more realistic, so as to guide the rate-distortion optimization process in visual perception-based MVD 3D video coding schemes.

References

[1] http://en.wikipedia.org/wiki/3D_television. 1

[2] The Vision Texture Database. http://vismod.media.mit.edu/vismod/imagery/VisionTexture/vistex.html. 124

[3] P. Aflaki, M. M. Hannuksela, and M. Gabbouj. Subjective quality assessment of asymmetric stereoscopic 3D video. *Signal, Image and Video Process.*, 9(2):331–345, February 2015. 146, 149, 174

[4] R. Akhter, Z. M. P. Sazzad, Y. Horita, and J. Baltes. No-reference stereoscopic image quality assessment. In *Proc. SPIE 7524, Stereoscopic Displays and Applications XXI*, San Jose, CA, USA, January 2010. 30, 111

[5] M. Azimi, S. Valizadeh, X. Li, L. E. Coria, and P. Nasiopoulos. Subjective study on asymmetric stereoscopic video with low-pass filtered slices. In *Proc. Int Comput., Network. and Communi.*, pages 719–723, Maui, HI, January 2012. 149

[6] A. Banitalebi-Dehkordi, M. T. Pourazad, and P. Nasiopoulos. Effect of high frame rates on 3D video quality of experience. In *Proc. IEEE Int. Conf. on Cons. Electron.*, pages 416–417, Las Vegas, NVUSA, January 2014. 26, 27

[7] A. Banitalebi-Dehkordi, M. T. Pourazad, and P. Nasiopoulos. The effect of frame rate on 3D video quality and bitrate. *3D Research*, 6(1):1–13, March 2015. 26, 27

[8] M. Barkowsky, K. Brunnström, T. Ebrahimi, L. Karam, P. Lebreton, P. Le Callet, A. Perkis, A. Raake, M. Subedar, K. Wang, L. Xing, and J. You. Subjective and

objective visual quality assessment in the context of stereoscopic 3D-TV. In *3D-TV System with Depth-Image-Based Rendering*, pages 413–437. 2013. 1

[9] Peter G. J. Barten. Formula for the contrast sensitivity of the human eye. In *Proc. SPIE 5294, Image Quality and System Performance*, pages 231–238, San Jose, CA, USA, January 2004. 80, 81

[10] A. Benoit, P. Le Callet, P. Campisi, and R. Cousseau. Quality assessment of stereoscopic images. *EURASIP Journal on Image and Video Processing*, 2008:1–13, October 2008. 20, 23, 28, 29, 30, 85, 86, 92, 93, 94, 116, 185

[11] A. Benoit, P. Le Callet, P. Campisi, and R. Cousseau. Using disparity for quality assessment of stereoscopic images. In *Proc. IEEE Int. Conf. Image Proc.*, pages 389–392, San Diego, CA, USA, October 2008. 28, 116

[12] R. Bensalma and M. C. Larabi. Towards a perceptual quality metric for color stereo images. In *Proc. IEEE Int. Conf. Image Proc.*, pages 4037–4040, Hong Kong, China, September 2010. 28

[13] G. Bjontegaard. Improvements of the BD-PSNR model. *ITU-T SG16 Q*, 6:35, 2008. 172

[14] R. Blake. Threshold conditions for binocular rivalry. *Journal of Experimental Psychology: Human Perception and Performance*, 3(2):251–257, May 1977. 74

[15] R. Blake. A primer on binocular rivalry, including current controversies. *Brain and Mind*, 2(1):5–38, April 2001. 60, 134, 159

[16] E. Bosc, R. Pepion, P. Le Callet, M. Köppel, P. Ndjiki-Nya, M. Pressigout, and L. Morin. Towards a new quality metric for 3-D synthesized view assessment. *IEEE J. of Sel. Topics in Sig. Process.*, 5(7):1332–1343, November 2011. 180

[17] H. Brust, A. Smolic, K. Mueller, G. Tech, and T. Wiegand. Mixed resolution coding of stereoscopic video for mobile devices. In *Proc. 3DTV Conf., True Vis.-Capture, Transmiss. Display 3D Video*, pages 1–4, Potsdam, Germany, May 2009. 146, 149

[18] D. C. Burr. Sensitivity to spatial phase. *Vision Research*, 20(5):391–396, 1980. 6

[19] P. Campisi, P. Le Callet, and E. Marini. Stereoscopic images quality assessment. In *Proc. European Signal Proc. Conf.*, pages 2110–2114, Poznan, Poland, September 2007. 27

[20] M. Carandini and D. J. Heeger. Normalization as a canonical neural computation. *Nature Reviews Neuroscience*, 13(1):51–62, January 2012. 101

[21] M. Carnec, P. Le Callet, and D. Barba. An image quality assessment method based on perception of structural information. In *Proc. IEEE Int. Conf. Image Proc.*, volume 2, pages 185–188, Barcelona, Spain, September 2003. 27, 116

[22] D. M Chandler and S. S. Hemami. VSNR: A wavelet-based visual signal-to-noise ratio for natural images. *IEEE Trans. Image Processing*, 16(9):2284–2298, September 2007. 11, 116

[23] M. Chen, L. K. Cormack, and A. C. Bovik. No-reference quality assessment of natural stereopairs. *IEEE Trans. Image Processing*, 22(9):3379–3391, September 2013. 19, 23, 28, 30, 31, 65, 83, 85, 111, 116, 183, 184

[24] M. Chen, D. Kwon, and A. C. Bovik. Study of subject agreement on stereoscopic video quality. In *IEEE Southwest Symp. Image Analysis & Interpretation*, pages 173–176, Santa Fe, NM, USA, April 2012. 17, 18, 24, 27, 29, 53, 56, 115, 120

[25] M. Chen, C.-C. Su, D. Kwon, L. K. Cormack, and A. C Bovik. Full-reference quality assessment of stereopairs accounting for rivalry. *Signal Processing: Image Communication*, 28(9):1143–1155, October 2013. 29, 85, 86, 92, 93, 94, 185

[26] W. Chen, J. Fournier, M. Barkowsky, and P. Le Callet. Exploration of quality of experience of stereoscopic images: Binocular depth. In *Proc. Int. Workshop Video Proc. and Quality Metrics for Consumer Electronics*, pages 116–121, Scottsdale, AZ, USA, January 2012. 2, 16, 53, 54, 56, 115, 120

[27] H. A. Chipman, E. D. Kolaczyk, and R. E. McCulloch. Adaptive bayesian wavelet shrinkage. *Journal of the American Statistical Association*, 92(440):1413–1421, 1997. 7

[28] J. Choi, D. Kim, S. Choi, and K. Sohn. Visual fatigue modeling and analysis for stereoscopic video. *Optical Engineering*, 51(1), February 2012. 17

[29] J. Choi, D. Kim, B. Ham, S. Choi, and K. Sohn. Visual fatigue evaluation and enhancement for 2D-plus-depth video. In *Proc. IEEE Int. Conf. Image Proc.*, pages 2981–2984, Hong Kong, China, September 2010. 17

[30] M. S. Crouse, R. D. Nowak, and R. G. Baraniuk. Wavelet-based statistical signal processing using hidden markov models. *IEEE Trans. Signal Processing*, 46(4):886–902, 1998. 7

[31] S. J. Daly. Visible differences predictor: an algorithm for the assessment of image fidelity. In *Proc. SPIE 1666, Human Vision, Visual Processing, and Digital Display III*, San Jose, CA, USA, February 1992. 7

[32] S. J. Daly, R. T. Held, and D. M. Hoffman. Perceptual issues in stereoscopic signal processing. *IEEE Trans. Broadcasting*, 57(2):347–361, June 2011. 16

[33] V. De Silva, H. K. Arachchi, E. Ekmekcioglu, A. Fernando, S. Dogan, A. Kondoz, and S. Savas. Psycho-physical limits of interocular blur suppression and its application to asymmetric stereoscopic video delivery. In *Proc. Int. Packet Video Workshop*, pages 184–189, Munich, Germany, May 2012. 149

[34] V. De Silva, H. K. Arachchi, E. Ekmekcioglu, and A. Kondoz. Toward an impairment metric for stereoscopic video: A full-reference video quality metric to assess compressed stereoscopic video. *IEEE Trans. Image Processing*, 22(9):3392–3404, September 2013. 24, 27, 30

[35] J. Ding and G. Sperling. A gain-control theory of binocular combination. *Proc. National Academy of Sciences of the United States of America*, 103(4):1141–1146, 2006. 74

[36] E. Dumić, S. Grgić, K. Šakić, P. M. R. Rocha, and L. A. da S. Cruz. 3D video subjective quality: a new database and grade comparison study. *Multimedia Tools and Applications*, pages 1–23, January 2016. 26, 27

[37] M. Fahle. Binocular rivalry: suppression depends on orientation and spatial frequency. *Vision Research*, 22(7):787–800, 1982. 74, 79

[38] L. Fang, N. Cheung, D. Tian, A. Vetro, H. Sun, and O. C. Au. An analytical model for synthesis distortion estimation in 3d video. *IEEE Trans. Image Process.*, 23(1):185–199, January 2014. 180

[39] D. J. Field. What is the goal of sensory coding? *Neural Computation*, 6(4):559–601, July 1994. 6

[40] M. A. Fischler and R. C. Bolles. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, 24(6):381–395, June 1981. 28

[41] J. M. Foley. Human luminance pattern-vision mechanisms: masking experiments require a new model. *Journal of the Optical Society of America A*, 11(6):1710–1719, June 1994. 100

[42] L. Goldmann, F. De Simone, and T. Ebrahimi. A comprehensive database and subjective evaluation methodology for quality of experience in stereoscopic video. In *Proc. SPIE 7526, Three-Dimen. Image Proc. and Applic.*, San Jose, CA, USA, January 2010. 26, 27

[43] L. Goldmann, F. De Simone, and T. Ebrahimi. Impact of acquisition distortions on the quality of stereoscopic images. In *Proc. Int. Workshop Video Proc. and Quality Metrics for Consumer Electronics*, pages 1–6, Scottsdale, AZ, USA, January 2010. 22, 23

[44] R. T. Goodwin and P. E. Romano. Stereoacuity degradation by experimental and real monocular and binocular amblyopia. *Investigative Ophthalmology & Visual Science*, 26(7):917–923, July 1985. 133

[45] P. Gorley and N. Holliman. Stereoscopic image quality metrics and compression. In *Proc. SPIE 6803, Stereoscopic Displays and Applications XIX*, San Jose, CA, USA, January 2008. 28, 29, 30

[46] Video Quality Experts Group et al. Final report from the video quality experts group on the validation of objective models of video quality assessment. *VQEG*, March 2000. 13, 14, 65, 121, 160

[47] K. Gu, G. Zhai, X. Yang, and W. Zhang. No-reference stereoscopic iqa approach: From nonlinear effect to parallax compensation. *Journal of Electrical and Computer Engineering*, September 2012. 111

[48] D. J. Heeger. Normalization of cell responses in cat striate cortex. *Visual Neuroscience*, 9(02):181–197, August 1992. 100

[49] R. F. Hess, C. H. Liu, and Y. Z. Wang. Differential binocular input and local stereopsis. *Vision Research*, 43(22):2303–2313, October 2003. 133

[50] C. T. E. R. Hewage and M. G. Martini. Reduced-reference quality metric for 3D depth map transmission. In *Proc. 3DTV-Conference: The True Vision-Capture, Transmission and Display of 3D Video*, pages 1–4, Tampere, Finland, June 2010. 29, 30

[51] C. T. E. R. Hewage, S. T. Worrall, S. Dogan, and A. M. Kondoz. Prediction of stereoscopic video quality using objective quality models of 2-D video. *Electronics letters*, 44(16):963–965, July 2008. 29

[52] H. Hirschmüller. Accurate and efficient stereo processing by semi-global matching and mutual information. In *Proc. IEEE Int. Conf. Computer Vision and Pattern Recognition*, volume 2, pages 807–814, San Diego, CA, USA, June 2005. 135

[53] N. S. Holliman, B. Froner, and S. P. Liversedge. An application driven comparison of depth perception on desktop 3D displays. In *Proc. SPIE 6490, Stereoscopic Displays and Virtual Reality Systems XIV*, San Jose, CA, USA, January 2007. 16, 115

[54] Y. Horita, K. Shibata, Y. Kawayoke, and Z. M. P. Sazzad. Mict image quality evaluation database. http://mict.eng.u-toyama.ac.jp/mictdb.html. 13

[55] C. Hsieh. 3D display technology and market forecast report. `http://www.displaysearch.com/cps/rde/xchg/displaysearch/hs.xsl/3d_display_technology_market_forecast_report.asp`. 1

[56] W. A. IJsselsteijn, H. de Ridder, and J. Vliegen. Subjective evaluation of stereoscopic images: effects of camera parameters and display duration. *IEEE Trans. Circuits and Systems for Video Tech.*, 10(2):225–233, March 2000. 17, 115

[57] W. Jang, T. Chung, J. Sim, and C. Kim. FDQM: Fast quality metric for depth maps without view synthesis. *IEEE Trans. Circuits and Systems for Video Tech.*, 25(7):1099–1112, July 2015. 180

[58] L. Jin, A. Boev, A. Gotchev, and K. Egiazarian. 3D-DCT based perceptual quality assessment of stereo video. In *Proc. IEEE Int. Conf. Image Proc.*, pages 2521–2524, Brussels, Belgium, September 2011. 30

[59] L. Jin, A. Gotchev, A. Boev, and K. Egiazarian. Validation of a new full reference metric for quality assessment of mobile 3DTV content. In *Proc. European Signal Proc. Conf.*, pages 1894–1898, Barcelona, Spain, August 2011. 30

[60] B. Julesz. *Foundations of Cyclopean Perception*. U. Chicago Press, 1971. 74

[61] S. Jumisko-Pyykkö, T. Haustola, A. Boev, and A. Gotchev. Subjective evaluation of mobile 3D video content: depth range versus compression artifacts. In *Proc. SPIE 7881, Multimedia on Mobile Devices 2011; and Multimedia Content Access: Algorithms and Systems V*, San Francisco Airport, CA, USA. 25, 27, 30

[62] Y. J. Jung, S. Lee, H. Sohn, H. W. Park, and Y. M. Ro. Visual comfort assessment metric based on salient object motion information in stereoscopic video. *Journal of Electronic Imaging*, 21(1), February 2012. 18

[63] H. Kalva, L. Christodoulou, and B. Furht. Evaluation of 3DTV service using asymmetric view coding based on MPEG-2. In *Proc. 3DTV Conf.*, pages 1–4, Kos Island, May 2007. 15

[64] R. G. Kaptein, A. Kuijsters, M. T. M. Lambooij, W. A. IJsselsteijn, and I. Heynderickx. Performance evaluation of 3D-TV systems. In *Proc. SPIE 6808, Image Quality and System Performance V*, San Jose, CA, USA, January 2008. 16, 115

[65] P. Kauff, N. Atzpadin, C. Fehn, M. Müller, O. Schreer, A. Smolic, and R. Tanger. Depth map creation and image-based rendering for advanced 3DTV services providing interoperability and scalability. *Signal Process.: Image Commun.*, 22(2):217–234, February 2007. 180

[66] L. Kaufman. *Sight and Mind: An Introduction to Visual Perception.* Oxford U. Press, 1974. 74

[67] A. Z. Khan and J. D. Crawford. Ocular dominance reverses as a function of horizontal gaze angle. *Vision Research*, 41(14):1743–1748, June 2001. 15, 59, 133, 158

[68] D. Kim and K. Sohn. Visual fatigue prediction for stereoscopic image. *IEEE Trans. Circuits and Systems for Video Tech.*, 21(2):231–236, February 2011. 17

[69] V. Kulyk, S. Tavakoli, M. Folkesson, K. Brunnstrom, K. Wang, and N. Garcia. 3D video quality assessment with multi-scale subjective method. In *Proc. Int. Conf. on Quality of Multimedia Experience*, pages 106–111, Klagenfurt am Wrthersee, Austria, July 2013. 16, 115

[70] M. Lambooij, W. IJsselsteijn, D. G. Bouwhuis, and I. Heynderickx. Evaluation of stereoscopic images: beyond 2D quality. *IEEE Trans. Broadcasting*, 57(2):432–444, June 2011. 17, 18, 115

[71] E. C. Larson and D. M. Chandler. The CSIQ image database. http://vision.okstate.edu/?loc=csiq. 11

[72] W. J. M. Levelt. The alternation process in binocular rivalry. *British Journal of Psychology*, 57(3-4):225–238, November 1966. 74

[73] Q. Li and Z. Wang. Reduced-reference image quality assessment using divisive normalization-based image representation. *IEEE J. of Sel. Topics in Signal Process.*, 3(2):202–211, April 2009. 100, 101

[74] Y. Lin and J. Wu. Quality assessment of stereoscopic 3D image compression by binocular integration behaviors. *IEEE Trans. Image Processing*, 23(4):1527 – 1542, April 2014. 29, 30, 185

[75] X. Liu, Y. Zhang, S. Hu, S. Kwong, C-C J. Kuo, and Q. Peng. Subjective and objective video quality assessment of 3D synthesized views with texture/depth compression distortion. *IEEE Trans. Image Processing*, 24(12):4847–4861, November 2015. 180

[76] D. G. Lowe. Object recognition from local scale-invariant features. In *Proc. IEEE Int. Conf. Computer Vision*, volume 2, pages 1150–1157, Vancouver, BC, Canada, September 1999. 28

[77] J. Lubin. A visual discrimination model for imaging system design and evaluation. *Vision Models for Target Detection and Recognition*, 2:245–357, 1995. 7

[78] K. Ma and Z. Wang. Multi-exposure image fusion: A patch-wise approach. In *Proc. IEEE Int. Conf. Image Process.*, pages 1717–1721, Quebec City, QC, Canada, September 2015. 138

[79] K. Ma, K. Zeng, and Z. Wang. Perceptual quality assessment for multi-exposure image fusion. *IEEE Trans. Image Processing*, 24(11):3345 – 3356, November 2015. 138

[80] L. Ma, S Li, F. Zhang, and K. N. Ngan. Reduced-reference image quality assessment using reorganized DCT-based image representation. *IEEE Trans. Multimedia*, 13(4):824–829, August 2011. 7

[81] A. P. Mapp, H. Ono, and R. Barbeito. What does the dominant eye dominate? a brief and somewhat contentious review. *Perception & Psychophysics*, 65(2):310–317, February 2003. 15, 60, 134, 159

[82] D. F. McAllister. *Stereo computer graphics and other true 3D technologies*. Princeton University Press, 1993. 122

[83] D. V. Meegan, L. B Stelmach, and W. J. Tam. Unequal weighting of monocular inputs in binocular combination: Implications for the compression of stereoscopic imagery. *Journal of Experimental Psychology: Applied*, 7(2):143–153, June 2001. 31, 32, 71

[84] L. M. J. Meesters, W. A. IJsselsteijn, and P. J. H. Seuntiens. A survey of perceptual evaluations and requirements of three-dimensional TV. *IEEE Trans. Circuits and Systems for Video Tech.*, 14(3):381–391, March 2004. 1, 16

[85] M. Mehrabi, E. M. Peek, B. C. Wuensche, and C. Lutteroth. Making 3D work: a classification of visual depth cues, 3D display technologies and their applications. In *Proc. 14th Australasian User Interface Conf.*, pages 91–100, 2013. 122

[86] A. Mittal, A. K. Moorthy, and A. C. Bovik. No-reference image quality assessment in the spatial domain. *IEEE Trans. Image Process.*, 21(12):4695–4708, December 2012. 95, 183, 185

[87] A. Mittal, R. Soundararajan, and A. C. Bovik. Making a completely blind image quality analyzer. *IEEE Signal Processing Letters*, 20(3):209–212, March 2013. 95, 184, 185

[88] A. K. Moorthy and A. C. Bovik. A two-step framework for constructing blind image quality indices. *IEEE Signal Processing Letters*, 17(5):513–516, May 2010. 95, 183, 185

[89] A. K. Moorthy and A. C. Bovik. Blind image quality assessment: From natural scene statistics to perceptual quality. *IEEE Trans. Image Process.*, 20(12):3350–3364, December 2011. 95, 183, 185

[90] A. K. Moorthy, C.-C. Su, A. Mittal, and A. C. Bovik. Subjective evaluation of stereoscopic image quality. *Signal Processing: Image Communication*, 28(8):870–883, September 2013. 19, 23, 30, 65, 116

[91] K. Müller, P. Merkle, and T. Wiegand. 3-D video representation using depth maps. *Proc. IEEE*, 99(4):643–656, April 2011. 180

[92] A. Ninassi, P. Le Callet, and F. Autrusseau. Subjective quality assessment - IVC database. http://www2.irccyn.ec-nantes.fr/ivcdb/. 11

[93] A. Ninassi, P. Le Callet, and F. Autrusseau. Pseudo no reference image quality metric using perceptual data hiding. In *Proc. SPIE 6057, Human Vision and Electronic Imaging XI*, San Jose, CA, USA, January 2006. 11

[94] B. T. Oh, J. Lee, and D. Park. Depth map coding based on synthesized view distortion function. *IEEE J. of Sel. Topics in Sig. Process.*, 5(7):1344–1352, November 2011. 180

[95] T. Okoshi. *Three-dimensional imaging techniques.* Elsevier, 2012. 122

[96] J. Park, S. Lee, and A. C. Bovik. 3D visual discomfort prediction: vergence, foveation, and the physiological optics of accommodation. *IEEE Journal on Selected Topics in Signal Process.*, 8(3):415–427, June 2014. 18

[97] J. Park, H. Oh, S. Lee, and A. C. Bovik. 3D visual discomfort predictor: Analysis of disparity and neural activity statistics. *IEEE Trans. Image Processing*, 24(3):1101–1114, March 2015. 18

[98] C. A. Parraga, T. Troscianko, and D. J. Tolhurst. The human visual system is optimised for processing the spatial information in natural visual images. *Current Biology*, 10(1):35–38, 2000. 7

[99] M. G. Perkins. Data compression of stereopairs. *IEEE Trans. Communications*, 40(4):684–696, April 1992. 149

[100] M. H. Pinson. The consumer digital video library [best of the web]. *IEEE Signal Processing Magazine*, 30(4):172–174, July 2013. 40, 151

[101] M. H. Pinson and S. Wolf. Comparing subjective video quality testing methodologies. In *Proc. SPIE Visual Comm. and Image Processing*, pages 573–582, Lugano, Switzerland, July 2003. 60, 159

[102] M. H Pinson and S. Wolf. A new standardized method for objectively measuring video quality. *IEEE Trans. Broadcasting*, 50(3):312–322, September 2004. 29, 116, 160

[103] N. Ponomarenko, F. Battisti, K. Egiazarian, J. Astola, and V. Lukin. Metrics performance comparison for color image database. In *Proc. Int. Workshop Video Proc. and Quality Metrics for Consumer Electronics*, pages 1–6, Scottsdale, AZ, USA, January 2009. 12, 13

[104] N. Ponomarenko and K. Egiazarian. Tampere image database 2008 TID2008. http://www.ponomarenko.info/tid2008.htm. 12, 16

[105] N. Ponomarenko and O. Ieremeiev. Tampere image database 2013 TID2013. http://www.ponomarenko.info/tid2013.htm. 12

[106] N. Ponomarenko, L. Jin, O. Ieremeiev, V. Lukin, K. Egiazarian, J. Astola, B. Vozel, K. Chehdi, M. Carli, F. Battisti, and C.-C. Jay Kuo. Image database tid2013: Peculiarities, results and perspectives. *Signal Processing: Image Communication*, 30:57–77, January 2015. 12

[107] O. Rosenbach. On monocular prevalence in binocular vision. *Med Wochenschrift*, 50:1290–1292, 1903. 59, 134, 158

[108] S. Ryu, D. H. Kim, and K. Sohn. Stereoscopic image quality metric based on binocular perception model. In *Proc. IEEE Int. Conf. Image Proc.*, pages 609–612, Orlando, FL, USA, September 2012. 28

[109] M. A. Saad, A. C. Bovik, and C. Charrier. Blind image quality assessment: A natural scene statistics approach in the DCT domain. *IEEE Trans. Image Process.*, 21(8):3339–3352, August 2012. 95, 183, 185

[110] C. P. Said and D. J. Heeger. A model of binocular rivalry and cross-orientation suppression. *PLoS Comput. Biol.*, 9(3):e1002991, March 2013. 101

[111] S. S. Sarikan, F. Olgun, R, and G. B. Akar. Quality evaluation of stereoscopic videos using depth map segmentation. In *Proc. Int. Conf. on Quality of Multimedia Experience*, pages 67–71, Mechelen,Belgium, September 2011. 30

[112] G. Saygili, C. G. Gurler, and A. M. Tekalp. Quality assessment of asymmetric stereo video coding. In *Proc. IEEE Int. Conf. Image Process.*, pages 4009–4012, Hong Kong, China, September 2010. 149

[113] Z. M. P. Sazzad, S. Yamanaka, Y. Kawayokeita, and Y. Horita. Stereoscopic image quality prediction. In *Proc. Int. Conf. on Quality of Multimedia Experience*, pages 180–185, July 2009. 20, 23, 28

[114] D. Scharstein and C. Pal. Learning conditional random fields for stereo. In *Proc. IEEE Int. Conf. Computer Vision and Pattern Recognition*, pages 1–8, Minneapolis, MN, USA, June 2007. 34

[115] D. Scharstein and R. Szeliski. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. *Int. Journal of Computer Vision*, 47(1-3):7–42, April 2002. 135

[116] A. Schertz. Source coding of stereoscopic television pictures. In *Proc. IEEE Int. Conf. Image Process.*, pages 462–464, Maastricht, The Netherlands, April 1992. 16, 115

[117] C. M. Schor and I. Wood. Disparity range for local stereopsis as a function of luminance spatial frequency. *Vision Research*, 23(12):1649–1654, 1983. 132

[118] J. Seo, X. Liu, D. Kim, and K. Sohn. An objective video quality metric for compressed stereoscopic video. *Circuits, Systems, and Signal Processing*, 31(3):1089–1107, June 2012. 30

[119] P. Seuntiens. Visual experience of 3D TV. *Doctor Doctoral Thesis, Faculty Technol. Manage., Eindhoven University of Technology*, 2006. 2, 15, 16, 18, 53, 63, 115

[120] P. Seuntiens, L. Meesters, and W. Ijsselsteijn. Perceived quality of compressed stereoscopic images: Effects of symmetric and asymmetric JPEG coding and camera separation. *ACM Trans. on Applied Perception*, 3(2):95–109, April 2006. 17, 31, 56, 71, 120

[121] F. Shao, W. Lin, S. Gu, G. Jiang, and T. Srikanthan. Perceptual full-reference quality assessment of stereoscopic images by considering binocular visual characteristics. *IEEE Trans. Image Processing*, 22(5):1940–1953, May 2013. 29, 30, 116, 185

[122] F. Shao, W. Lin, S. Wang, G. Jiang, and M. Yu. Blind image quality assessment for stereoscopic images using binocular guided quality lookup and visual codebook. *IEEE Trans. Broadcasting*, 61(2):154–165, June 2015. 111

[123] H. R. Sheikh and A. C. Bovik. Image information and visual quality. *IEEE Trans. Image Processing*, 15(2):430–444, February 2006. 28, 116

[124] H. R. Sheikh, A. C. Bovik, L. K. Cormack, and Z. Wang. Image and video quality assessment research at LIVE. http://live.ece.utexas.edu/research/quality/subjective.htm. 12

[125] H. R Sheikh, A. C. Bovik, and G. De Veciana. An information fidelity criterion for image quality assessment using natural scene statistics. *IEEE Trans. Image Processing*, 14(12):2117–2128, December 2005. 7

[126] H. R. Sheikh, M. F. Sabir, and A. C. Bovik. A statistical evaluation of recent full reference image quality assessment algorithms. *IEEE Trans. Image Processing*, 15(11):3440–3451, November 2006. 13, 95

[127] L. Shen, J. Yang, and Z. Zhang. Stereo picture quality estimation based on a multiple channel HVS model. In *Proc. IEEE Int. Congress Image and Signal Proc.*, pages 1–4, Tianjin, China. 28, 30

[128] E. Shneor and S. Hochstein. Effects of eye dominance in visual perception. In *Int. Congress Series*, volume 1282, pages 719–723, September 2005. 15

[129] E. P. Simoncelli. Modeling the joint statistics of images in the wavelet domain. In *Proc. SPIE 3813, Wavelet Applications in Signal and Image Processing VII*, Denver, CO, USA, July 1999. 7

[130] E. P. Simoncelli, W. T. Freeman, E. H. Adelson, and D. J. Heeger. Shiftable multiscale transforms. *IEEE Trans. Information Theory*, 38(2):587–607, March 1992. 101

[131] E. P. Simoncelli and D. J. Heeger. A model of neuronal responses in visual area MT. *Vision Research*, 38(5):743–761, March 1998. 100

[132] R. Song, H. Ko, and C.-C. Kuo. MCL-3D: a database for stereoscopic image quality assessment using 2D-image-plus-depth source. *arXiv preprint arXiv:1405.1403*, 2014. 22, 23, 181

[133] L. Stelmach, W. J. Tam, D. Meegan, and A. Vincent. Stereo image quality: effects of mixed spatio-temporal resolution. *IEEE Trans. Circuits and Systems for Video Tech.*, 10(2):188–193, March 2000. 146, 149

[134] C.-C. Su, A. K. Moorthy, and A. C. Bovik. Visual quality assessment of stereoscopic image and video: Challenges, Advances, and Future trends. In *Visual Signal Quality Assessment*, pages 185–212. 2015. 1

[135] N. Sun, H. Mansour, and R. Ward. HDR image construction from multi-exposed stereo LDR images. In *Proc. IEEE Int. Conf. Image Process.*, pages 2973–2976, Hongkong, September 2010. 48

[136] W. J. Tam, F. Speranza, S. YVano, K. Shimono, and H. Ono. Stereoscopic 3D-TV: Visual Comfort. *IEEE Trans. Broadcasting*, 57(2):335–346, April 2011. 17

[137] W. J. Tam, L. B. Stelmach, and P. J. Corriveau. Psychovisual aspects of viewing stereoscopic video sequences. In *Proc. SPIE 3295, Stereoscopic Displays and Virtual Reality Systems V*, San Jose, CA, USA, January 1998. 16, 115

[138] A. Tikanmaki, A. Gotchev, A. Smolic, and K. Muller. Quality assessment of 3D video in rate allocation experiments. In *Proc. IEEE Int. Symposium on Consumer Electronics*, pages 1–4, Vilamoura,Portugal, April 2008. 29

[139] M. Urvoy, M. Barkowsky, R. Cousseau, Y. Koudota, V. Ricorde, P. Le Callet, J. Gutiérrez, and N. García. Nama3ds1-cospad1: Subjective video quality assessment database on coding conditions introducing freely available high quality 3D stereoscopic sequences. In *Proc. Int. Conf. on Quality of Multimedia Experience*, pages 109–114, Yarra Valley, VIC, Australia, July 2012. 25, 27, 40, 151

[140] M. J. Wainwright and E. P. Simoncelli. Scale mixtures of gaussians and the statistics of natural images. In *Adv. Neural Information Processing Systems*, volume 12, pages 855–861, May 2000. 73

[141] B. A. Wandell. *Foundations of vision.* Sinauer Associates, 1995. 9

[142] J. Wang, A. Rehman, K. Zeng, S. Wang, and Z. Wang. Quality prediction of asymmetrically distorted stereoscopic 3D images. *IEEE Trans. Image Process.*, 24(11):3400–3414, November 2015. 145, 150, 172, 183, 184

[143] J. Wang, S. Wang, and Z. Wang. Quality prediction of asymmetrically compressed stereoscopic videos. In *Proc. IEEE Int. Conf. Image Process.*, pages 3427–3431, Quebec City, QC, Canada, September 2015. 146, 183, 184

[144] J. Wang and Z. Wang. Perceptual quality of asymmetrically distorted stereoscopic images: the role of image distortion types. In *Proc. Int. Workshop Video Proc. and Quality Metrics for Consumer Electronics*, pages 1–6, Chandler, AZ, USA, January 2014. 82, 122

[145] J. Wang, K. Zeng, and Z. Wang. Quality prediction of asymmetrically distorted stereoscopic images from single views. In *Proc. IEEE Int. Conf. Multimedia and Expo*, pages 1–6, Chengdu, Sichuan, China, July 2014. 82, 122, 183, 185

[146] K. Wang, M. Barkowsky, K. Brunnström, M. Sjöström, R. Cousseau, and P. Le Callet. Perceived 3D TV transmission quality assessment: multi-laboratory results using

absolute category rating on quality of experience scale. *IEEE Trans. Broadcasting*, 58(4):544–557, December 2012. 117

[147] S. Wang, K. Ma, H. Yeganeh, Z. Wang, and W. Lin. A patch-structure representation method for quality assessment of contrast changed images. *IEEE Signal Processing Letters*, 24(11):3345–3356, November 2015. 138

[148] X. Wang, M. Yu, Y. Yang, and G. Jiang. Research on subjective stereoscopic image quality assessment. In *Proc. SPIE 7255, Multimedia Content Access: Algorithms and Systems III*, San Jose, CA, USA, January 2009. 20, 23, 31, 32, 71

[149] Z. Wang and A. C. Bovik. A universal image quality index. *IEEE Signal Processing Letters*, 9(3):81–84, March 2002. 27, 116

[150] Z. Wang and A. C. Bovik. *Modern Image Quality Assessment.* Morgan & Claypool Publishers, March 2006. 7

[151] Z. Wang and A. C. Bovik. Mean squared error: Love it or leave it? A new look at signal fidelity measures. *IEEE Signal Processing Magazine*, 26(1):98–117, January 2009. 8

[152] Z. Wang and A. C. Bovik. Reduced- and no-reference image quality assessment. *IEEE Signal Processing Magazine*, 28(6):29–40, November 2011. 93

[153] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli. Image quality assessment: From error visibility to structural similarity. *IEEE Trans. Image Processing*, 13(4):600–612, April 2004. 7, 11, 36, 71, 116, 142, 183, 185

[154] Z. Wang and Q. Li. Information content weighting for perceptual image quality assessment. *IEEE Trans. Image Processing*, 20(5):1185–1198, May 2011. 10, 71, 73, 142, 160, 175, 183, 185, 188

[155] Z. Wang, L. Lu, and A. C. Bovik. Video quality assessment based on structural distortion measurement. *Signal Processing: Image Communication,* special issue on objective video quality metrics, 19(2):121–132, February 2004. 29

[156] Z. Wang and X. Shang. Spatial pooling strategies for perceptual image quality assessment. In *Proc. IEEE Int. Conf. Image Proc.*, Atlanta, GA, USA, October 2006. 73, 175

[157] Z. Wang and E. P. Simoncelli. Reduced-reference image quality assessment using a wavelet-domain natural image statistic model. In *Human Vision and Electronic Imaging X, Proc. SPIE*, volume 5666, San Jose, CA, January 2005. 27

[158] Z. Wang, E. P. Simoncelli, and A. C. Bovik. Multi-scale structural similarity for image quality assessment. In *Proc. IEEE Asilomar Conf. on Signals, Systems, and Computers*, pages 1398–1402, Pacific Grove, CA, November 2003. 9, 10, 11, 28, 71, 116, 142, 160, 183, 185

[159] Q. Wu, H. Li, F. Meng, K. N. Ngan, B. Luo, C. Huang, and B. Zeng. Blind image quality assessment based on multi-channel features fusion and label transfer. *IEEE Trans. Circuits and Systems for Video Tech.*, 20(3):209–219, November 2015. 95, 184, 185

[160] Q. Wu, Z. Wang, and H. Li. A highly efficient method for blind image quality assessment. In *Proc. IEEE Int. Conf. Image Process.*, pages 339–343, Quebec City, QC, Canada, September 2015. 95, 184, 185

[161] W. Xue, X. Mou, L. Zhang, A. C. Bovik, and X. Feng. Blind image quality assessment using joint statistics of gradient magnitude and laplacian features. *IEEE Trans. Image Process.*, 23(11):4850–4862, November 2014. 95, 184, 185

[162] W. Xue, L. Zhang, and X. Mou. Learning without human scores for blind image quality assessment. In *Proc. IEEE Int. Conf. Computer Vision and Pattern Recognition*, pages 995–1002, June 2013. 95, 184, 185

[163] J. Yang, C. Hou, R. Xu, and J. Lei. New metric for stereo image quality assessment based on HVS. *Int. Journal of Imaging Systems and Technology*, 20(4):301–307, November 2010. 28, 29, 30, 116

[164] J. Yang, C. Hou, Y. Zhou, Z. Zhang, and J. Guo. Objective quality assessment method of stereo images. In *Proc. 3DTV Conference: The True Vision-Capture, Transmission and Display of 3D Video*, pages 1–4, Potsdam, German, May 2009. 21, 23, 28, 30, 85, 86, 92, 93, 94, 116, 185

[165] Y. Yang and R. Blake. Spatial frequency tuning of human stereopsis. *Vision Research*, 31(7-8):1176–1189, 1991. 80

[166] S. L. P. Yasakethu, C. T. E. R. Hewage, W Fernando, and A Kondoz. Quality analysis for 3D video using 2D video quality models. *IEEE Trans. Consumer Electronics*, 54(4):1969–1976, November 2008. 29, 116

[167] P. Ye, J. Kumar, L. Kang, and D. Doermann. Unsupervised feature learning framework for no-reference image quality assessment. In *Proc. IEEE Int. Conf. Computer Vision and Pattern Recognition*, pages 1098–1105, Providence, RI, USA, June 2012. 95, 183, 185

[168] J. You, L. Xing, A. Perkis, and X. Wang. Perceptual quality assessment for stereoscopic images based on 2D image quality metrics and disparity analysis. In *Proc. Int. Workshop Video Proc. and Quality Metrics for Consumer Electronics*, pages 61–66, Scottsdale, AZ, USA, January 2010. 28, 29, 30, 85, 86, 92, 93, 94, 116, 185

[169] C. M. Zaroff, M. Knutelska, and T. E. Frumkes. Variation in stereoacuity: normative description, fixation disparity, and the roles of aging and gender. *Investigative Ophthalmology & Visual Science*, 44(2):891–900, February 2003. 131

[170] K. Zeng, A. Rehman, J. Wang, and Z. Wang. From H.264 to HEVC: Coding gain predicted by objective video quality assessment models. In *Proc. Int. Workshop Video Proc. and Quality Metrics for Consumer Electronics*, pages 1–6, scottsdale, AZ, USA, January 2013. 201

[171] K. Zeng and Z. Wang. 3D-SSIM for video quality assessment. In *Proc. IEEE Int. Conf. Image Process.*, pages 621–624, Orlando, FL, US, September 2012. 176

[172] K. Zeng and Z. Wang. Polyview fusion: A strategy to enhance video-denoising algorithms. *IEEE Trans. Image Processing*, 21(4):2324–2328, April 2012. 176

[173] K. Zeng, T. Zhao, A. Rehman, and Z. Wang. Characterizing perceptual artifacts in compressed video streams. In *Proc. SPIE 9014, Human Vision and Electronic Imaging XIX*, San Francisco, CA, USA, February 2014. 201

[174] Y. Zhang and D. M. Chandler. 3D-MAD: A full reference stereoscopic image quality estimator based on binocular lightness and contrast perception. *IEEE Trans. Image Process.*, 24(11):3810–3825, November 2015. 185, 188

[175] J. Zhou, G. Jiang, X. Mao, M. Yu, F. Shao, Z. Peng, and Y. Zhang. Subjective quality analyses of stereoscopic images in 3DTV system. In *Proc. SPIE Visual Comm. and Image Processing*, pages 1–4, November 2011. 21, 23

[176] Z. Zhu and Y. Wang. Perceptual distortion metric for stereo video quality evaluation. *WSEAS Tran. Signal Process.*, 5(7):241–250, July 2009. 29, 30, 116

[177] M. Zwicker, S. Yea, A. Vetro, C. Forlines, W. Matusik, and H. Pfister. Display pre-filtering for multi-view video compression. In *Proc. Int. Conf. Multimedia*, pages 1046–1053, 2007. 16, 115