The Dissertation Committee for Ming-Jun Chen
certifies that this is the approved version of the following dissertation:

Visual Perception and Quality of Distorted Stereoscopic 3D Images

Committee:

Alan C. Bovik, Supervisor

Lawrence K. Cormack, Co-Supervisor

Wilson S. Geisler

Joydeep Ghosh

Gustavo de Veciana

Visual Perception and Quality of Distorted Stereoscopic 3D Images

by

Ming-Jun Chen, B.S.; M.S.

Dissertation

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

DOCTOR OF PHILOSOPHY

The University of Texas at Austin

December 2012

Dedicated to Ling-Ying, Hsung-Hsing, and my family.

Acknowledgements

I would foremost like to thank all supports I have gotten during these years.

First, I would like to thank my wife, Ling-Ying Lu. Without the supports from her, I would have had no chance to obtain my Ph.D degree.

I want to express my deepest gratitude to my advisors, Professor Alan C. Bovik and Professor Lawrence K. Cormack for their guidance in the four and half-year process of this study. Their advices and support helped me explore exciting ideas and develop a quality work. I also want to thank my committee members, Professor Wilson S. Geisler, Professor Joydeep Ghosh, and Professor Gustavo de Veciana for their valuable insights which further enhance the depth of my work.

My heartfelt thanks also go to lab mates at Laboratory for Image and Video Engineering (LIVE) including Sina Jahabin, Joonsoo Lee, Anush Moorthy, Rajiv Soundararajan, Ajay Gopinath, Anish Mittal, Michele Saad, Chao Chen, Che-Chun Su, Gautam Muralidhar, Lark Kwon Choi, Deepti Ghadiyaram and Dinesh Jayaraman for all discussions and fun that make LIVE a memorable place.

I would like to thank my friends for their supports to me and my family. Your supports helped me and my wife to be able to raise a newborn baby by ourselves.

Especially, I would like to thank Torry Wen and Mao-Lun Weng for their kindness help to take care of my wife and daughter when I am occupied. I would like to thank my best friend Chun-Neng Huang for taking my calls whenever I need to have a talk.

A special thanks to the kind financial support from Texas instrument. A major portion of my study years was supported by their university funding program. Also, I would like to thank TI members Minhou Zhou, Do-Kyoung Kwon, Franscis Yoo, Hamid Sheikh, and Greg Hewes. I had great time working with them during my summer internships at TI and their encouragements also contribute to my PhD work.

Finally, I would like to thank my family. My older brother Ming-Hsien Chen and my sister in law Yi-Chen Wu who take care of my parents when I stay in US; and my father and mother in law who always provide unconditional supports.

Visual Perception and Quality of Distorted Stereoscopic 3D Images

Ming-Jun Chen, Ph.D.
The University of Texas at Austin, 2012

Supervisors: Alan C. Bovik
Lawrence K. Cormack

This dissertation focuses on the investigation of human perception of stereoscopic 3D image quality and the development of automatic stereoscopic 3D image quality assessment frameworks. In order to assess human perception of visual quality, a human study was conducted and interactions between image quality, depth quality, visual comfort, and 3D viewing quality were inferred. The results indicate that the overall 3D viewing quality can be well predicted from only image quality and depth quality. Between image and depth quality, image quality seems to be the main factor that enables accurate prediction of overall 3D viewing quality. Two other human studies were conducted to study the effect of masking on stereoscopic distortions. Binocular suppression was observed in the stereo images which were distorted by blur, JPEG compression, or JPEG2K compression, however, no such

suppression was observed for stereo images distorted by white noise. Further, a facilitation effect was also observed against disparity variation for blur and JPEG2K distorted stereo images while no depth masking effect was observed. Based on these results, I proposed an automatic full-reference (FR) 3D quality assessment framework. In this framework, I used Gabor filterbank responses to model stimulus strength and then synthesize a Cyclopean image from a stereo image pair. Because the quality of this synthesized view is similar to that of a Cyclopean image, which the human visual system recreates from the stereoscopic stimuli, performing the task of 3D quality assessment on synthesized views can deliver better performance. I verified the performance of this FR framework on the LIVE 3D Image Quality Database and the results indicate that applying the proposed framework improves the performance of FR 2D quality assessment algorithms when applied to stereo 3D images. Further, I proposed a no-reference (NR) 3D quality assessment (QA) algorithm based on natural scene statistics in both the spatial and the depth domain. Experiments indicate that the proposed NR algorithm outperforms all 2D FR QA algorithms and most 3D FR QA models in predicting 3D quality of stereo images. Finally, a fourth subjective study was conducted to understand depth quality when stereo content is free from visual discomfort. The result suggests that human perception of depth quality is correlated

with the content of the stereo image and the stereoacuity function of human visual system.

Table of Contents

Chapter 1 Introduction 1
1.1 Motivation 1
1.2 Contribution 2
1.2.1 Exploring the knowledge of distorted stereo 3D images 2
1.2.2 Quality assessment algorithms and database for stereo 3D images 3
1.2.3 Depth quality of stereoscopic 3D images 6
1.3 Goal 8
Chapter 2 Background 11
2.1 Human visual system and depth perception 11
2.2 Stereoscopic 3D viewing 14
2.3 Human perception of distorted stereoscopic 3D images 16
2.4 Quality assessment 17
2.4.1 2D quality assessment 17
2.4.2 Stereoscopic 3D quality assessment 18
Chapter 3 Study of Subjective Agreement of Stereoscopic Video Quality 23
3.1 Introduction 23
3.2 Subjective study 23
3.2.1 Stimuli 23
3.2.2 Display 25
3.2.3 Study design 25
3.2.4 Obtaining subjective scores 27
3.3 Data analysis and discussion 28
3.3.1 Quality assessment metrics 28
3.3.2 Inter-metric analysis 34
3.3.3 Discussion 35
3.4 Conclusion 39
Chapter 4 Masking in Distorted 3D Stereo Images 40
4.1 Introduction 40
4.2 Stereo image source 41
4.3 Display 42
4.4 Observers 43
4.5 Study One 43
4.5.1 Stimuli 43
4.5.2 Procedure 45

4.5.3 Binocular suppression 47
4.5.4 Contrast and depth masking 51
4.6 Study Two 59
4.6.1 Stimuli 59
4.6.2 Procedure 61
4.6.3 Binocular suppression 61
4.6.4 Contrast and depth masking 65
4.7 Conclusion 68
Chapter 5 Full Reference Quality Assessment of Stereopairs Accounting for Rivalry 70
5.1 Introduction 70
5.2 Binocular rivalry/suppression 71
5.3 A framework for stereo quality assessment 74
5.3.1 Stereo matching algorithms 76
5.3.2 Gabor filter bank 78
5.3.3 Cyclopean image 80
5.4 Experiment 83
5.4.1 Stereoscopic image quality dataset 83
5.4.2 Results 89
5.5 Conclusion 103
Chapter 6 No-Reference Stereoscopic Quality Assessment 106
6.1 Introduction 106
6.2 The proposed NR 3D IQA model 107
6.2.1 2D feature extraction 109
6.2.2 3D feature extraction 111
6.2.3 Quality estimation 120
6.3 Results 121
6.3.1 LIVE 3D Image Quality Database 121
6.3.2 Classification accuracy 126
6.3.3 Performance 127
6.4 Conclusion 135
Chapter 7 Quality of Depth 137
7.1 Introduction 137
7.2 Presentation model 140
7.2.1 Foreground/ background dominance 141
7.2.2 Maximizing depth resolution 143

7.2.3 Optimizing the presentation 146
7.3 Human Study 148
7.3.1 Study design 148
7.3.2 Display 149
7.3.3 Observers 150
7.3.4 Stimuli 150
7.4 Results and analysis 151
7.5 Conclusion 154
Chapter 8 Conclusion and Future Works 156
Bibliography 160

List of Figures

Fig. 1 Comparison of natural viewing and stereo 3D viewing 15
Fig. 2 Standard deviations of subjective ratings. 29
Fig. 3 Mean ratings vs. individual ratings..... 30
Fig. 4 Means and standard deviations of ranked correlations..... 30
Fig. 5 Means and standard deviations of Pearson correlations 31
Fig. 6 Ratings of depth quality from two distinct subjects 33
Fig. 7 Mean ratings of viewing comfort. 33
Fig. 8 Mean ratings of 3D quality and the predicted ratings from these two linear regression models 37
Fig. 9 Example of stereo image pair 41
Fig. 10 Range map associated with the stereo image pair in Fig. 9..... 42
Fig. 11 Illustration of 4-Mirror stereo rig (top view)..... 43
Fig. 12 Image with local white noise distortion. The boundary was blended using a Gaussian blending window. When the image was presented, the subject was requested to point out the distortion by clicking the mouse cursor on the distortion. 46
Fig. 13 When the rating bar showed up, the subject was requested to render a subjective 3D quality opinion on the entire image..... 46
Fig. 14 Plot of PC (top), DMOS (middle), and Times (bottom) in finding distortions..... 49
Fig. 15 The analysis flow of discussing contrast and depth masking on binocular distorted stereo images..... 54
Fig. 16 The results of Welch's t test of white noise (top left), blur (top right), JPEG (bottom left) and JP2K (bottom right) compression distortion on the local contrast. Red dot indicates mean and blue bar represents standard deviation. 55
Fig. 17 The results of Welch's *t* test of white noise (top left), blur (top right), JPEG (bottom left) and JP2K (bottom right) compression distortion on the range. Red dot indicates mean and blue bar represents standard deviation..... 56
Fig. 18 The results of *t* test on subjects' selection on left and right views. . 58
Fig. 19 The PC and MS of dichoptic distorted 3D image and binocular distorted 3D image. The distortion type is White Noise..... 63

Fig. 20 The PC and MS of dichoptic distorted 3D image and binocular distorted 3D image. The distortion type is JPEG compression distortion 64
Fig. 21 The PC and MS of dichoptic distorted 3D image and binocular distorted 3D image. The distortion type is JP2K compression distortion 65
Fig. 22 The PC and MS of binocular distorted images. The blue solid lines show results for stimuli where the distortion is placed at high ratio areas, and the red dotted lines show results when the distortion is placed at low ratio areas. 66
Fig. 23 The PC and MS of dichoptic distorted images. Blue solid lines show the results for stimuli where the distortion is placed at high ratio areas, and the red dotted lines show results when the distortion is placed at low ratio areas. 68
Fig. 24 Illustration of binocular rivalry: Two different patterns are presented to the left eye (an arrow) and the right eye (a star). The blue line indicates that the stimulus is perceived by a human observer inside that time interval. 72
Fig. 25 Illustration of binocular suppression: Two different patterns are presented to the left eye (an arrow) and the right eye (a star). An observer only sees the arrow when s/he experiences binocular suppression. 73
Fig. 26 The proposed framework for 3D QA 76
Fig. 27 A stereo image distorted by white noise (free-fused) and the cyclopean image created by the proposed algorithm. 83
Fig. 28 The eight stereo images used for the database 85
Fig. 29 A stereo image (free-fuse the left and right images) and the ground truth disparity maps. 86
Fig. 30 SROCC values using MS-SSIM, broken down by distortion type. 92
Fig. 31 Plot of predicted objective scores versus DMOS and prediction errors. Top Left: Prediction by MS-SSIM cyclopean framework. Top Right: Prediction errors of MS-SSIM cyclopean framework. Bottom Left: Predictions by MS-SSIM 2D baseline. Bottom Right: Prediction errors of MS-SSIM 2D baseline 93

Fig. 32 Depth estimation using SSIM-based stereo algorithm on noised distorted stereo pairs. Free-fuse the noisy stereo image to see a 3D image. 99
Fig. 33 The flowchart of the proposed 3D NR QA model 109
Fig. 34. A stereopair with ground truth disparity and estimated disparity. Top left: Right view of the stereo image. Top right: Left view of the stereo image. Bottom left: Ground truth disparity. Bottom right: Estimated disparity. 114
Fig. 35. Top left: Histogram of ground truth disparity map. Top right: Histogram of the estimated disparity map. Bottom left: Histogram of the local-normalized ground truth disparity map GGD fit overlaid. Bottom right: Histogram of the estimated disparity map with GGD fit overlaid. Bottom left: Ground truth disparity. Bottom right: Estimated disparity. 115
Fig. 36. Disparity distributions of a distorted stereopair. 116
Fig. 37. Top left: Left view of a stereopair. Top right: Histogram of the uncertainty map and the best log-normal fit. Bottom left: The estimated disparity map. Bottom right: The uncertainty map produced by the stereo matching algorithm. 118
Fig. 38. Plot of modelled uncertainty distributions of distorted stereopair. 119
Fig. 39. A stereo image (free-fuse the left and right images) and ground truth disparity maps. 122
Fig. 40. Left: DMOS of LIVE 3D Image Quality Database Phase I. Right: DMOS of LIVE 3D Image Quality Database Phase II. 126
Fig. 41. Left: DMOS of Phase II symmetric distorted stimuli. Right: DMOS of Phase II asymmetric distorted stimuli 126
Fig. 42 Left : Illustration of crossed and uncrossed disparity. Right: The parallel camera configuration. 138
Fig. 43 Zone of comfortable stereo viewing 140
Fig. 44. Stereoacuity function, *s(d)* with 100 ms stimuli from 144
Fig. 45 A stereo image 146
Fig. 46 Shifting the stereo image inside zone of comfortable viewing 146
Fig. 47. Flowchart of the proposed algorithm. 148
Fig. 48 The GUI of DSCQS in our study 149

Fig. 49 Performance of different 3D image presentation strategies. Error bars represent the standard errors. 153
Fig. 50 Top Left: a foreground dominant image with a positive skewed disparity distribution. Top Bottom: The histogram of disparities of the rank one stereo representation. Top Right: a background dominant image with a negative skewed disparity distribution. Top Right: The histogram of disparities of the rank one stereo representation. 154

List of Tables

Table 1 The QP values for the left view and right views of the stereoscopic video 25
Table 2 SROCC between subjective quality metrics 35
Table 3 SROCC of PSNR and MS-SSIM against spatial quality and overall 3D quality 38
Table 4 The results of ANOVA on the locations of distortions. 48
Table 5 The results of ANOVA of each distortion type on percent correct. 50
Table 6 The results of ANOVA of each distortion type on time spent to find the distortion. 51
Table 7 The results of ANOVA of each distortion type on DMOS 51
Table 8 An example of a Chi-square test setup. The number inside brackets is the expected value. 58
Table 9 The results of Chi-square test on contrast 59
Table 10 The results of Chi-square test on range 59
Table 11 Range of parameter values for distortion simulation 87
Table 12 SROCC scores obtained by averaging left and right QA scores (center column) and using the 3D “cyclopean” model (right column) 90
Table 13 LCC scores obtained by averaging left and right QA scores (center column) and using the 3D “cyclopean” model (right column) 90
Table 14 RMSE values obtained by averaging left and right QA scores (center column) and using the 3D “cyclopean” model (right column) 91
Table 15 SROCC scores relative to human subjective scores. Obtained using averaged left-right QA scores (2D Baseline) and the Cyclopean model on symmetric and asymmetric distorted stereopairs 95
Table 16 LCC scores relative to human subjective scores. Obtained using averaged left-right QA scores (2D Baseline) and the Cyclopean model on symmetric and asymmetric distorted stereopairs 96
Table 17 Fitting errors measured by RMSE. Obtained using averaged left-right QA scores (2D Baseline) and the Cyclopean model on symmetric and asymmetric distorted stereopairs 97
Table 18 Mean bad pixel rate value on 360 distorted stereopairs with standard deviation (inside the bracket) for three stereo algorithms. 98
Table 19 SROCC, LCC, RMSE relative to human subjective scores attained by cyclopean model using different disparity maps. 100

Table 20 SROCC, LCC, and RMSE relative to human subjective scores attained by 3D QA models using SSIM-based stereo algorithm. 102
Table 21 Range of parameter values for distortion simulation 124
Table 22 Comparison of 2D IQA algorithms: SROCC against DMOS on the LIVE Phase I 3D IQA dataset. Italicized algorithms are NR IQA algorithms, all others are FR IQA algorithms. 128
Table 23 Comparison of 3D IQA models: SROCC against DMOS of the Phase I dataset. Italicized algorithms are NR IQA algorithms, all others are RR or FR IQA algorithms. 130
Table 24 Comparison of 2D IQA algorithms: SROCC against DMOS on the LIVE Phase II 3D IQA dataset. Italicized algorithms are NR IQA algorithms, others are FR IQA algorithms. 131
Table 25 Comparison of 3D IQA algorithms: SROCC against DMOS on the LIVE Phase II 3D IQA dataset. Italicized algorithms are NR IQA algorithms, others are FR IQA algorithms. 132
Table 26 Break down of performance on symmetrically and asymmetrically distorted stimuli in the Phase II dataset. Italicized algorithms are NR IQA models, others are RR or FR IQA algorithms. 134
Table 27 Test across datasets: SROCC against DMOS of the Phase I dataset 135

CHAPTER 1 INTRODUCTION

1.1 Motivation

Stereoscopic vision was first systematically studied by Wheatstone [1] in the early 1800's and the production of 3D films can be dated back to 1903 [2]. Since then, numerous 3D films have been produced, culminating in the breakout success of *Avatar* in 2009, which went on to become the highest-grossing film of all time. The success of *Avatar* has since greatly inspired further efforts in 3D film production and improved technologies and methods for 3D content capture and display. According to the Motion Picture Association of America (MPAA), half of the moviegoers see at least one 3D movie in 2011, while those under 25 years old saw more than twice that number [3]. To meet his demand, the number of 3D movies has been increasing by at least 50% annually over the past few years [4].

The wave of 3D has not been limited to the movie theatre. In 2011, mobile phones supporting 3D capture and viewing were made available, and the number of 3D films released on home-viewing media such as DVDs tripled since 2008. Broadcast of 3D content over the internet becomes commonplace [5]. With the release of 3D phones and 3D broadcast services, it is reasonable to believe that the amount of 3D content that is

delivered by wireless and wireline will follow the trend of consumer video and increase exponentially over the next few years.

Understanding how to monitor the integrity of 2D and 3D visual signals throughout computer networks has become a critical question. Being able to provide visual quality assurance via the ability to automatically assess the quality of visual media delivered to the client is both demanding and increasingly urgent. Thus, the development of objective visual quality assessment models of images and videos has been a busy and fruitful area of work [4]. However, while great advances have been mode on modeling regular (non-stereoscopic) image and video quality [6, 7], progress on the question of 3D image quality has been limited [8].

1.2 Contribution

The following is an overview of the contributions presented in this dissertation.

1.2.1 Exploring the knowledge of distorted stereo 3D images

To understand human perception of quality of distorted stereoscopic images, several human studies have been conducted in the past two decades, but none of these studies analyzed depth masking. Further, these studies proposed a variety of binocular masking effects. To thoroughly analyze the effect of masking on perception of stereoscopic quality, I designed and conducted two human studies. These studies were

designed to infer possible binocular and depth masking effects when viewing stereoscopic 3D images.

Different from previous human studies on distorted stereoscopic images, my studies were conducted on stereoscopic images which have ground truth depth information obtained from a high-precision laser range scanner and the stereoscopic images were distorted locally. Since high-resolution statistics of local contrast and depth are available, possible masking effects were unearthed by conducting a series of statistical analyses. Correlations between the visibility of local distortions and local statistics were studied. The results suggest that binocular masking effect is observed in those stereo images that are distorted by blur, JPEG compression, or JPEG2K compression. However, no binocular masking effect was observed for stereo images distorted by white noise. The results also indicated that no depth masking effect was observed for distorted stereo images. In contrast, a facilitation effect was observed against disparity variation for blur and JPEG2K distorted stereo images. The studies and analyses are described in Chapter 4.

1.2.2 Quality assessment algorithms and database for stereo 3D images

While there has been tremendous activity in the area of 2D quality assessment, the field of automatically assessing the quality of 3D images remains ill-explored. Further, most of the existing 3D models for QA are ad-hoc in nature and lack a grounded model of human visual perception to support the design choices. Some of these algorithms are simple extensions of previously proposed 2D quality assessment approaches. Thus, it is not a surprise that most of these 3D quality assessment algorithms struggle to show better performance in predicting the 3D quality of stereoscopic 3D images compared to high performance 2D quality assessment algorithms.

In this dissertation, I first propose a full reference (FR) 3D quality assessment framework. This framework allows for easy extension of 2D FR algorithms, providing a plug-and-play approach to the development of 3D FR QA algorithms. This framework attempts to model the fact that binocular masking alters perceived 3D quality, as observed from my previous study. The proposed framework uses the linear model proposed by Levelt [9] to synthesize an intermediate view from a stereo image pair, called the *cyclopean view*. Levelt's work models the binocular masking effect with local statistics of a stereo image. Thus, the synthesized cyclopean view is visually close to the

true Cyclopean image which a human subject recreates in his head while viewing a stereo image pair on a stereoscopic 3D display.

To verify my framework, I conducted another human study to construct a 3D image quality database with both symmetrically and asymmetrically distorted stereo stimuli annotated with human subjective ratings on the perceived 3D quality. My framework was verified against the database using different 2D FR quality assessment algorithms. The experimental results indicate that my proposed framework when coupled with 2D FR algorithms predicts stereoscopic distortions with greater accuracy, especially for asymmetrically distorted images. For symmetrically distorted images, my proposed framework does not drastically improve performance. This result is a further verification of the observations from my human studies. Since only asymmetrically distorted images demonstrate binocular masking, and only binocular masking seems to have an effect on stereoscopically viewing 3D quality, modeling binocular masking algorithmically produces gains only in the case of such asymmetrically distorted stereopairs.

Based on this proposed framework, I further developed a no-reference (NR) quality assessment algorithm. To the best of my knowledge, there exists only one other 3D QA algorithm in literature, which, as we shall see, does not perform well. The

proposed NR algorithm utilizes natural scene statistics in both spatial and depth domains to extract 2D and 3D features. A support vector machine is used for training and later for predicting the quality of stereo images. This NR algorithm is verified against the database which I constructed and the phase I of LIVE Image Quality Database (only symmetrically distorted stereo image are available in this database). Experiments show that my proposed NR 3D QA algorithm outperforms all 2D quality assessment algorithms and most FR 3D quality assessment algorithms. Using the multi-scale structural similarity index (MS-SSIM) in my previously proposed 3D QA framework produces similar correlations as that of my proposed NR 3D QA algorithm.

1.2.3 Depth quality of stereoscopic 3D images

Previous studies on visual quality of stereoscopic images demonstrated that both image quality and depth quality affect the overall 3D quality of the presentation. However, each of these studies differs in their observations of the influence of spatial distortions on depth quality. Tam, *et al* [10] showed that perceived depth quality is correlated with spatial image quality, but the authors of [11, 12] claim that spatial distortion has no effect on perceived depth quality. To verify this effect, I conducted a human study to infer the coherence of subjective human ratings on spatial image quality,

depth quality, visual comfort, and perceived 3D quality. The analysis shows that ratings of depth quality are much more diverse than those of spatial quality and perceived 3D quality. The ratings of depth quality may be classified into two groups. The first of these groups are those ratings that correlate with spatial image quality, while those of the second group demonstrate no such correlation, and remain constant with varying spatial qualities. The results of my study demonstrate that different human populations rate depth quality differently, and hence, opinion on depth quality should be gauged before modeling the depth and/or overall 3D quality of stereoscopic presentations. Further, depth quality is not limited to the depth quality of the distorted stereo image.

For a pristine stereo image, perceived depth quality may be changed due to the arrangement of display. In our natural 3D environment, the accommodation of our eyes changes with an increase/decrease of the vergent distance from our eyes to the focused object. However, while viewing stereoscopic 3D presentations on 2D displays, the accommodation of our eyes remains fixed on the display screen, while the vergence varies as we focus on different perceived depth planes. This is referred to as the accommodation-vergence conflict.

To infer if depth quality is also affected by the accommodation-vergence conflict, I conducted another human study. The results of this study suggest that perceived depth quality of a distortion-free stereo image is also affected by the stereoacuity function of human vision system and the prior knowledge that we have about the natural 3D world. To the best of my knowledge, this is the first time that someone has studied depth quality in this fashion and demonstrated that depth quality is correlated with the stereoacuity function and the prior knowledge we have about the natural 3D world.

1.3 Goal

This dissertation aims at advancing our knowledge of 3D quality of distorted stereoscopic images and understanding the effect that depth quality has on the perception of otherwise distortion-free stereo images. Towards this end, a human study was first conducted to infer the interaction between image quality, depth quality, visual comfort, and 3D viewing quality. The results indicate that image quality and depth quality both affect overall 3D viewing quality, and that image quality is the dominant factor in determining overall perception of stereoscopic stimuli. Two more human studies were conducted to understand the effect that masking has on the perception of distorted

stereoscopic 3D images. My analysis indicates that binocular rivalry is an important effect which influences the 3D viewing quality of a distorted stereopair. Further, no depth masking effect was observed. Based on my discovery of the (lack of) masking effect, a full-reference (FR) 3D quality assessment (QA) framework and a no reference (NR) QA algorithm were proposed. To verify the performance of these 3D QA algorithms, a fourth human study was conducted on asymmetrically distorted stereo pairs which resulted in the construction of phase II of the LIVE 3D Image Quality Database which hitherto consisted only of symmetrically distorted stereo pairs. The proposed 3D QA algorithms were tested on this database and demonstrated to outperform present-day 2D FR QA algorithms and most NR 3D QA algorithms. Finally, I conducted a fifth human study to infer the effect that depth quality of distortion-free stereo images has on overall 3D perception and proposed another algorithm to automatically assess this effect.

The rest of this dissertation is organized as follows. Chapter 2 reviews the literature and justifies the direction of the proposed research. Chapter 3 summarizes a human study which analyzes the interaction between image quality, depth quality, visual comfort, and 3D viewing quality. Chapter 4 describes two subjective studies which investigate the masking effect of distorted stereo images. Chapter 5 describes the

construction of a full reference 3D quality assessment model based on the findings from the studies described in Chapter 4 and also details another human study on asymmetrically distorted stereo images. Chapter 6 details a no reference quality assessment model based on binocular rivalry and natural scene statistics. Chapter 7 describes a fifth human study on depth quality and details an algorithm to predict depth quality. Conclusion and future works of this dissertation are described in Chapter 8.

CHAPTER 2 BACKGROUND

In this chapter, I first give an introduction of the human visual system and some background on stereoscopic 3D displays. I then review previous work on the human perception of distorted stereoscopic 3D images, 2D quality assessment algorithms, and 3D quality assessment algorithms.

2.1 Human visual system and depth perception

Light from the visual world, when incident on the photoreceptors of our eyes is converted into neuron responses by these receptors, which are then processed by the higher level processing mechanisms within the visual cortex, to interpret and understand the visual stimulus [13].

Light reflected from a scene enters each eye through the pupil; then, the image of the scene is focused at the retina, with the aid of an adjustable lens through a process called *accommodation*. The spatial resolution of the incident stimulus is not equally distributed on the retina. The highest spatial resolution is in the central zone of the retina, called the *fovea.* Therefore, eye movements are required by our visual system to produce a high spatial resolution of the observed scene. This eye movement is called *visual fixation*. *Saccades* and *vergence* are two types of fixational eye movements. A saccade is

the process of changing fixations horizontally and vergence is the process of changing fixations in depth. These two eye movements are triggered by different stimuli. The stimulus that triggers vergence is retinal disparity information, and the stimulus that triggers accommodation is retinal image blur. In natural conditions, accommodation and vergence are intrinsically linked [14].

The human binocular visual system receives two slightly different views of a visual scene, and ideally (without occlusion) the corresponding points in these two views are apart from each other by various horizontal distances. However, instead of two distinct images, the human visual system perceives a single image, called the *cyclopean view* [15], and infers stereoscopic depth from these two stimuli [16]. The distances between corresponding points in these two views are used by our brain to interpret depth information.

Objects that are binocularly fixated are on the same relative coordinates in the left view and the right view and have zero retinal disparity. A curved line that connects all points that have zero retinal disparity is called the *horopter*. Human subjects have the same perceived distance of the fixation point with those points which are located at the horopter. Objects that are more distant from the point of fixation are said to produce

uncrossed disparities, objects that are closer than the point of fixation are said to produce *crossed disparities*. The human visual system uses these disparities to extract the relative depth of objects in a visual scene.

A small region around the horopter, called *Panum's fusional area (Panum's area)*, is a region within which objects can be fused binocularly even though these objects have non-zero retinal disparities. Objects that are located outside Panum's area result in double images. The size of Panum's area is not constant over the retina, and its size also depends on the spatial and temporal properties of the fixation target [17, 18].

In addition, when we fixate at an object, the image of the fixated object falls on the retina. Theoretically, objects which are closer to or farther from the accommodation distance will be seen as blurry images. However, the visual system is tolerant of a small amount of blur. Thus, objects that lie within a small region around the accommodation point can be perceived with high resolution (i..e., not blurred), and the size of this region is known as the *depth of field* (DOF).

Finally, disparity is not the only cue used by the human visual system to perceive depth. There are other cues that also contribute to human depth perception. For example, monocular cues include occlusion, relative size, texture gradient, geometry perspective,

lighting, shading, and motion parallax. However, it is not yet clear that how our brain integrates these cues and produces the final perceived depth, but recent research [19] has shown that depth cues may be integrated by a statistical inference model.

2.2 Stereoscopic 3D viewing

The main difference between stereoscopic 3D viewing on a display and natural 3D viewing is that the synchronization between accommodation and vergence only exists in natural 3D viewing. As mentioned in the previous section, changes in accommodation naturally induce changes in vergence and vice versa. Stereoscopic 3D viewing on a 2D display, however, can produce conflicts between these two processes. Fig. 1 shows an example of such a conflict. In Fig. 1, the left hand-side shows that the vergence point and focal (accommodation) point are always at the same distance in natural viewing. However, inconsistencies exist in stereo 3D viewing on displays because the vergence distance varies depending on the image content while the accommodation distance remains constant (on the display). This inconsistence is called *vergence-accommodation conflict*.

It is known that the vergence-accommodation conflict creates visual discomfort and visual fatigue [20-22], but the conflict is not the only factor that produces negative effects. For example, crosstalk produced by the display [23, 24], the magnitude of disparity variations and the presence of fast motion [25] may also create visual discomfort. Because this dissertation focuses on the quality (spatial image quality, depth quality, and perceived 3D quality) of stereoscopic 3D images, I take precautions to minimize the visual discomfort for all human studies conducted in this dissertation. Thus, a comprehensive review of visual discomfort is not performed in this dissertation.

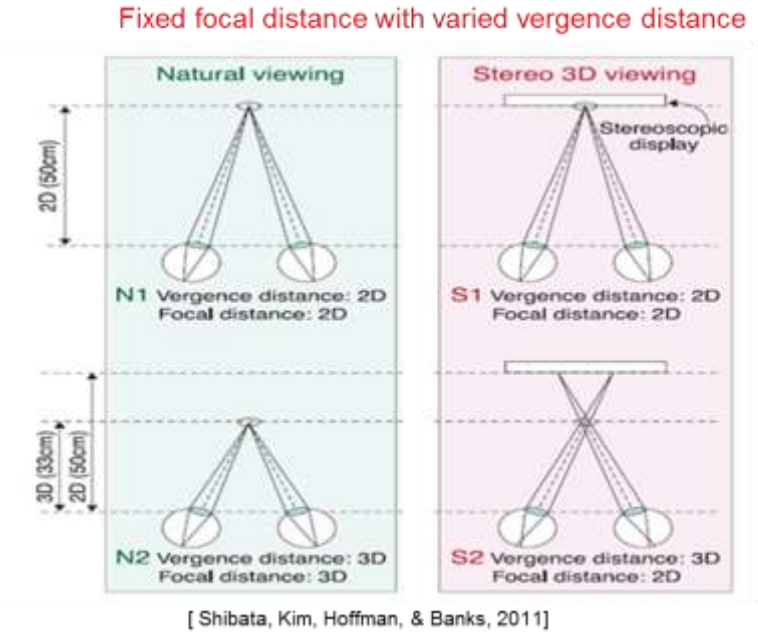

Fig. 1 Comparison of natural viewing and stereo 3D viewing

2.3 Human perception of distorted stereoscopic 3D images

Viewing stereoscopic 3D images produces a vastly different perception than that produced when viewing 2D images. In general, depth quality and visual comfort are the two extra factors that need to be considered when a human subject is asked to give ratings to stereoscopically viewed 3D images. Seuntiëns [16] conducted human studies to discuss the interactions between spatial image quality, depth quality, visual discomfort, and overall 3D viewing experience. He proposed the term *3D visual experience,* to be a combination of spatial image quality, depth quality, and visual comfort, to describe the overall stereo 3D viewing experience. He further demonstrated that without considering visual comfort, 3D visual experience can be predicted from a linear combination of spatial image quality and depth quality, and that spatial image quality is more important than depth quality in predicting 3D visual experience. Because my dissertation focuses on the quality of stereo images with an assumption that visual discomfort is minimized, I use the term "perceived 3D quality" to represent the ratings given by human subjects when viewing a stereo image stereoscopically. While there exists research that discusses the reliability of human ratings [26] given to 2D images and shows that across different

human subjects there is a general agreement on spatial image quality, equivalent research is absent for ratings given to depth quality and perceived 3D quality. Moreover, research on the possible masking effects of distorted stereo images is very limited. The authors of [12, 27] claimed that the quality of JPEG compressed images is approximately the average of the quality of the left and the right view. They further claimed that JPEG encoding has no effect on depth quality. However, Tam, et al. [10] claimed that depth quality is correlated with spatial image quality. This dispute is discussed in Chapter 3 of this dissertation.

Other than this dispute, Meegan, *et al.*[28] claimed that the perceived 3D quality of asymmetric MPEG-2 distorted stereo images is approximately the average of the spatial image quality of the two views, but that the perception of asymmetric blur distorted stereo images is dominated by the higher quality view. Similar findings are also mentioned in the JPEG distorted stereo images [27]. However, to the best of my knowledge, depth masking effects have never been considered while assessing stereoscopic quality.

2.4 Quality assessment

2.4.1 2D quality assessment

2D image quality assessment has seen quite a bit of research activity [6, 7]. These efforts can be categorized into three sets based on the availability of reference content. The first set is full-reference (FR) quality assessment (QA) algorithms. An algorithm is classified as a FR QA algorithm if the reference content is required to perform quality assessment. Some well-known FR QA models are widely used now, such as the peak signal-to-noise ratio (PSNR) and the Structural SIMilarity (SSIM) index [29, 30].

The second category is called reduced-reference (RR) QA algorithms. Models that only require reduced information, such as features or coefficients extracted from the reference source, are classified as RR QA models [31]. The last category is no-reference (NR) QA, where the reference is not needed while predicting the quality of an unknown image. NR QA models can be subdivided into single distortion models and multi-distortions models. Single distortion models can only deal with a single type of expected distortion such as blur [32], or compression artifacts [33]. Multi-distortions models [34], on the other hand, can deal with multiple pre-defined distortion types. Recently, research [35-38] on multi-distortions models are shown to be able to perform as well as FR QA models.

2.4.2 Stereoscopic 3D quality assessment

Other than the existing categories (FR, RR, and NR) for 2D QA algorithms, we further divide 3D QA algorithms into two categories. The first class (Class 1) [39-41] are 2D-based 3D QA models which do not utilize computed or otherwise measured depth/disparity information from the stereopairs. Among Class 1 models, the methods in [39, 40] conduct 2D FR QA on the left and right views independently, then combine (by various means) the two scores into predicted 3D quality scores. Gorley, *et al.* [41] compute quality scores on matched feature points delivered by SIFT [42] and RANSAC [43] applied to the two views.

The second class (Class 2) of models includes some kind of disparity information in the overall 3D QA process. Among these 3D QA models, Benoit, et al. [44] proposed a FR 3D QA algorithm which computes the quality scores between the left reference and left distorted view, the right reference and right distorted view, and the reference and distorted disparity map. The quality scores are computed by C4 [45] and SSIM [29], and different combinations are used to produce a final predicted scores from these three scores. Their results show that disparity information can improve the 3D QA algorithm based on SSIM (called 3D-SSIM); but the 2D C4 algorithm performs better than the 3D-SSIM algorithm. In addition, they also pointed out that the disparity estimation

algorithm can affect the performance of 3D QA algorithms. You, *et al.* [46] further extended the idea of predicting the 3D quality of a stereopair by applying 2D QA algorithms on the stereopair and its disparity map. They applied a large pool of FR 2D QA algorithms on stereopairs and the associated disparity maps, and concluded that applying SSIM on stereopairs and mean-absolute-difference (MAD) on their estimated disparity map can achieve the best performance in predicting the 3D quality of stereo images. In contrast to the results in Benoit, *et al.*, their SSIM-based 3D QA algorithm significantly outperformed all 2D FR QA algorithms on their dataset. Under the same framework, Zhu, *et al.* [47] proposed a 3D QA algorithm utilizing their own 2D quality assessment algorithm. Similarly, Yang, *et al.* [48] proposed a FR 3D QA algorithm based on the average PSNR of the stereopair and the absolute difference between the left and right view. Their algorithm did not need a stereo matching algorithm. None of these 3D QA algorithms are supported by research in human perception, and the techniques described therein are ad-hoc in nature.

Recently, some new perspectives have been explored in area of FR 3D QA. Maalouf, *et al.* [49] proposed to perform the task of 3D QA on the Cyclopean image, which is defined as the average of the left view and the disparity-compensated right view.

However, studies [9, 50] have shown that Cyclopean image is not simply the average of the left and the disparity-compensated right view. Bensalma, *et al.* [51] proposed a 3D QA algorithm based on measuring the difference of binocular energy between the reference and the tested stereopair. Their algorithm considers the potential influence of binocular effects on the perceived 3D quality.

Compared to 2D QA algorithms, there are a very limited number of RR or NR 3D QA algorithms. A RR 3D QA algorithm was proposed by Hewage, *et al.* [52]. In their algorithm, edge information of the depth map is transmitted, and they compute the PSNR between the reference and tested edge maps to predict the 3D quality of the tested stereo videos. Akhter, *et al.* [53] proposed a NR 3D QA algorithm which extracted features from stereopairs and estimated disparity map. Then a logistic regression model is used to predict 3D quality scores from these features[1].

The above review of 3D QA algorithms demonstrates that most 3D QA algorithms are still full-reference in nature and struggle to justify why their design

[1] There are other NR 3D QA algorithms which are designed to deal with DIBR-based 3D images/videos, but this dissertation only discusses those algorithms which function on natural stereo content.

performs better than 2D FR QA algorithms in predicting the 3D quality of stereoscopic content.

CHAPTER 3 STUDY OF SUBJECTIVE AGREEMENT OF STEREOSCOPIC VIDEO QUALITY

3.1 Introduction

The objective of this chapter is to discuss the coherence of ratings given to spatial image quality, depth quality, visual comfort, and overall 3D viewing quality, and the interactions between these four ratings. To evaluate the performance of different quality assessment algorithms, image quality databases with human annotated subjective quality scores are generally used as the ground truth. The correlations between the predicted quality scores and the human annotated scores in the databases are the criteria used to evaluate the performance of the quality assessment algorithms. However, this verification is valid only if there is a general agreement amongst the ratings given by different subjects. Hence, this chapter first discusses the coherence of ratings given to these four measurements. In addition, Seuntiens proposed that overall 3D viewing experience can be predicted only from spatial image quality and depth quality for distorted stereo images. This chapter verifies if his model can be applied on distorted stereo videos.

3.2 Subjective study

3.2.1 Stimuli

Six uncompressed natural scene videos, including indoor and outdoor scenes, were chosen as source videos. Two of them (soccer, puppy) are from ETRI in Korea and the other four are from the EPFL stereo video database [54]. All videos were down-sampled to 720 x 480 resolution. Two of these videos are fifteen seconds long, while the rest are ten seconds long. All of the sequences have a frame rate of 25 frames per second.

H.264 compression was chosen as the distortion method and both symmetric and asymmetric coding scenarios were included. Each pristine sequence was used to create 9 distorted test sequences compressed with different quantization parameter (QP) values. The specific settings for the nine distorted videos associated with each original video are shown in Table 1.

Table 1 The QP values for the left view and right views of the stereoscopic video

Left view QP	Right view QP
25	Pristine
30	Pristine
35	Pristine
25	25
30	25
35	25
30	30
35	30
35	35

3.2.2 Display

An nVidia active 3D kit plus an Alienware OptX AW2310 full HD 3D monitor were used to display the 3D videos. The viewing distance from subjects to screen was fixed at 23 inches which is 3 times the screen height.

3.2.3 Study design

I adopted a single stimulus continuous quality scale (SSCQS) [55] protocol to obtain subjective quality ratings for all of the video sequences in the database. A training session was conducted at the beginning of the study to familiarize each of the subjects with the graphical user interface (GUI). The subjects were pre-screened to ensure normal stereovision. In addition, a pristine video and a “highly distorted” video were shown in

the training session to help observers normalize their ratings. The training content was different from the videos used in the study and the content was impaired by the same type of distortion. Repeated viewing of the same 3D video was allowed, since I found that subjects sometimes needed time to accommodate their eye convergence to a new 3D video.

The goal of this work is to understand subjects' ratings of 'spatial image quality' (SIG), 'depth quality' (DQ), 'visual comfort' (VC), and 'overall 3D quality (3DQ)'. However, in experiments preliminary to the study we found that it was difficult for subjects to rate these quality scores independently. Further, when being asked to give an overall 3D quality score for each stimulus, subjects tended to have trouble assigning relative 'weights' to SIQ, DQ, and VC. Hence, a matched-pair experimental design was used to conduct the study.

In the matched-pair study, the study is repeated using two groups of subjects to obtain matched measurements of subjective scores. In the first study, the subjects in group A were requested to give subjective scores on SIQ, DQ, and VC. In assigning SIQ, the subjects were requested to assign quality scores only based on the content quality they viewed without considering the quality of their 3D viewing experiences. In addition, the

subjects were asked to assign depth quality scores based only on the realism of 3D depth they viewed when viewing stereo 3D videos. The subjects were also asked to give a visual comfort score based on how comfortable they felt when viewing stereoscopic 3D videos. In the second study, the subjects in group B were requested to give an overall 3D quality score when viewing stereo 3D videos. Again, the task of rating videos was explained carefully in the training session prior to each subjects' participation. Instructions were given to observers that the scoring is based on overall 3D viewing experience.

In both study groups, 11 video sequences (a 3D pristine video, a 2D pristine video (right view), and nine distorted videos) were shown to the subjects for each pristine video. The 3D reference video was hidden to enable the calculation of DMOS scores of perceived spatial video quality and overall 3D video quality.

Subjects having similar backgrounds were recruited for the two groups. In group A, thirteen subjects (twelve males and one female) were recruited with ages ranging from 24 to 45. In group B, fourteen subjects (eleven males and three females) were recruited and their ages ranged from 24 to 50.

3.2.4 Obtaining subjective scores

Differential mean opinion scores (DMOS) were calculated by subtracting the ratings of each 3D reference video from each associated rating. Those scores were then normalized to Z-scores. I followed the suggestion from [55] to perform subject screening. Outliers were screened by removing any ratings that fell outside two standard deviations from the center of a Gaussian fit to the ratings' SROCC against the mean DMOS. Finally, the DMOS score for each video was computed as the mean of the rescaled Z-scores from the remaining subjects following subject rejection.

After the subject rejection process, only one subject was rejected in group A. No outlier was found in group B.

3.3 Data analysis and discussion

3.3.1 Quality assessment metrics

Following Seuntiens, et al. [27], I calculated the standard deviation of the normalized ratings (Z-scores scaled to 0~100) assigned to each video. Then, the average of these standard deviation values was calculated to show the degree of agreement of the ratings (see Fig. 2) shows that the ratings given to perceived spatial video quality have the least variation. However, it is difficult to claim any significant difference between the four subjective metrics from the table. Therefore, I decided to use the correlation between the ratings given by different subjects to discover whether their ratings were similar

across the four kinds of 'qualities.' I first calculated the correlation values between the mean scores and the ratings given by each subject. As shown in Fig. 3, I computed the correlation values between the individual ratings and the mean ratings. The average of these correlation values reflects the degree of agreement of ratings among the subjects. Fig. 4 shows the Spearman Ranked Order Correlation Coefficients (SROCC) and Fig. 5 is the Pearson Correlation Coefficients. The ratings of SIQ show the highest agreement while the ratings of DQ show the least.

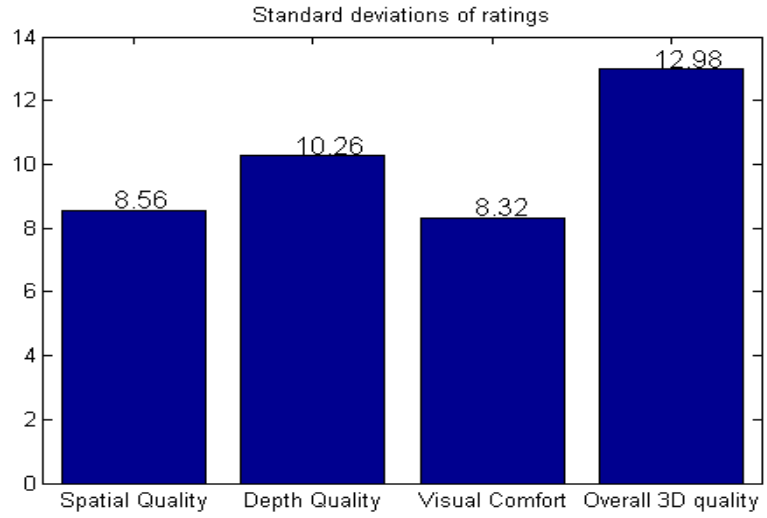

Fig. 2 Standard deviations of subjective ratings.

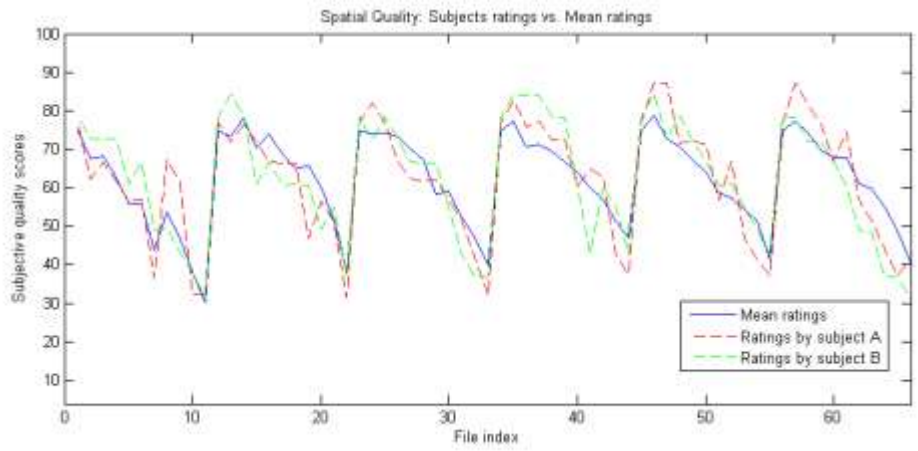

Fig. 3 Mean ratings vs. individual ratings

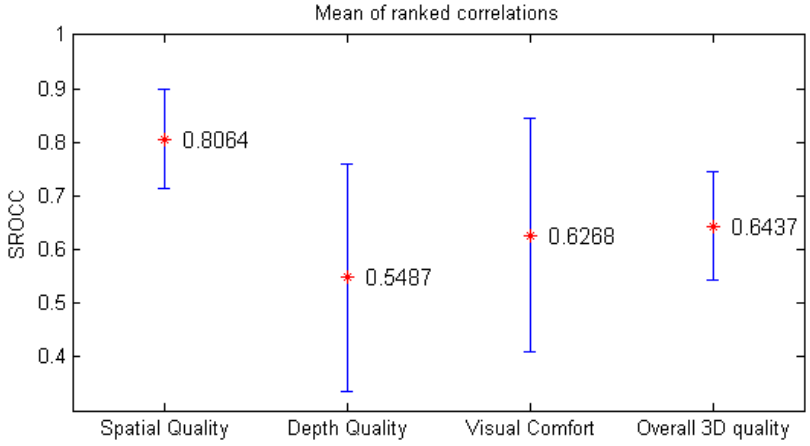

Fig. 4 Means and standard deviations of ranked correlations

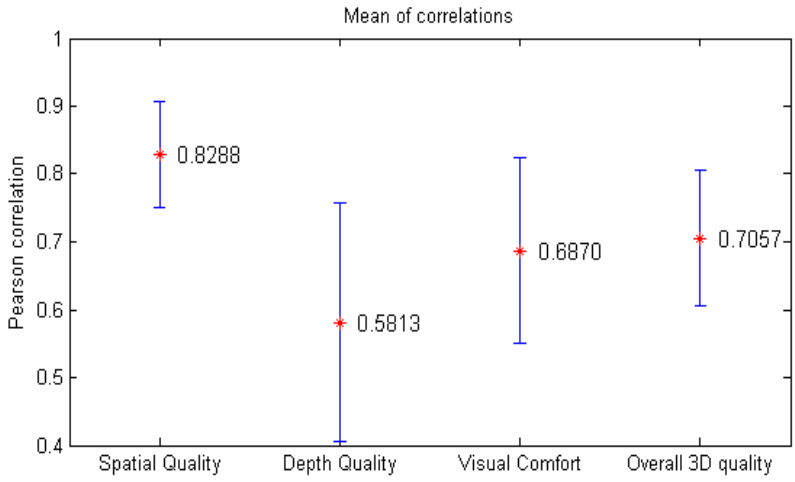

Fig. 5 Means and standard deviations of Pearson correlations

I further analyzed the low agreement ratings for DQ. The data shows that some subjects assigned a lower depth quality score when the video had lower spatial video quality, while others subjects thought that compression distortion did not affect perceived depth quality. Fig. 6 is an example which shows the rating of two subjects in our study. Subject A assigned a variety of depth quality scores while subjects B assigned very similar depth quality scores. Across multiple subjects, there are diverse options for interpreting depth quality. Discovering why different people have different opinions is worthy of further exploration.

The degree of agreement of ratings on overall 3D video quality is lower than that for spatial video quality and higher than that for depth quality. This observation may provide an insight on how to build a 3D video quality database.

The ratings of visual comfort assigned when viewing distorted stereoscopic 3D videos shows an average degree of agreement. While the underlying 3D geometric setting of the distorted videos is unaltered and carefully dealt to ensure that there is minimized accommodation-vergence conflict and crosstalk caused by the viewing setting, some discomfort in viewing a stereoscopic video may result either from the intrinsic geometry of the videos or from the compression distortion. Although subjects did not closely agree on visual comfort, my data shows that they were more comfortable when viewing the hidden 2D pristine video. As shown in Fig. 7, the subjective scores assigned when viewing 2D video were the highest comfort scores.

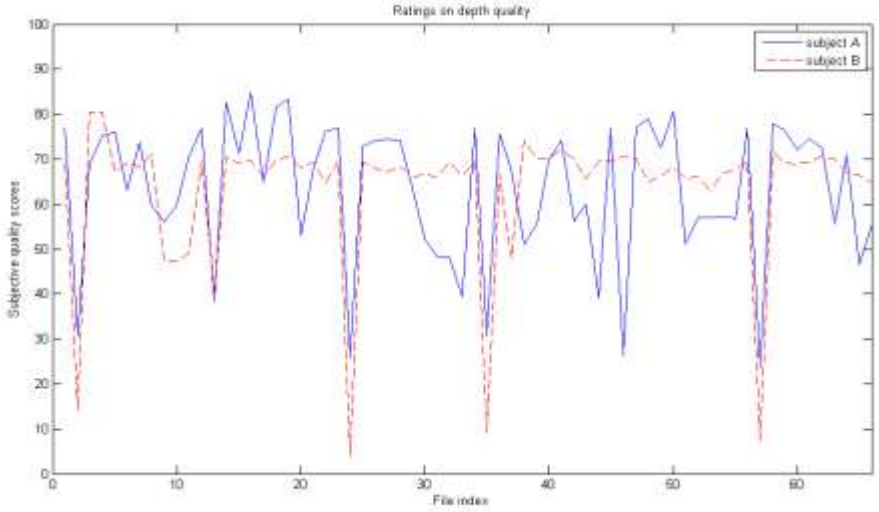

Fig. 6 Ratings of depth quality from two distinct subjects

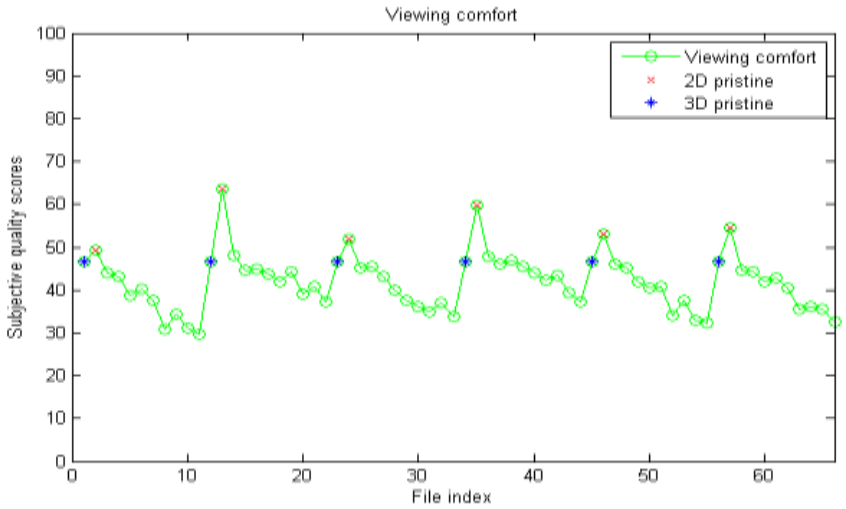

Fig. 7 Mean ratings of viewing comfort.

One possible explanation for the phenomena I have observed is that humans are more familiar with distortions in 2D videos than in 3D videos. After all, television was invented in the late 1930s and we have been living with distorted 2D videos for a long time. Whereas, for most people, stereoscopic 3D video viewing is still a new experience. Viewing stereoscopic 3D videos is more complex task than our daily stereo vision. In our daily stereo vision, our eyes verge and focus at the same time. However, when viewing stereoscopic 3D video, our eyes only change vergence while the focused point is fixed on the screen. So, most human subjects may simply be insufficiently experienced in viewing stereoscopic 3D video to reliably judge perceived depth quality. This may partly explain why the subjects have more diverse opinions on perceived depth quality and why they feel more comfortable viewing 2D videos. Lastly humans exhibit a wide range of stereoacuity and stereosense [56, 57], ranging from complete deficiency to better than normal. This ability would naturally affect a subject's impressions of both 3D distortions and comfort.

3.3.2 Inter-metric analysis

In this section, interactions between subjective quality metrics are discussed. Table 2 lists the SROCC scores between these subjective quality metrics. First, for spatial

quality, this subjective quality metric has high correlation with visual comfort and overall 3D quality. The results indicate that visual discomfort mainly results from coding artifacts since other variables are controlled in this study, and the overall 3D quality is more correlated to the spatial quality compared to depth quality, as was mentioned in previous work [58]. Second, for depth quality, this subjective measurement doesn't have a high correlation with spatial quality and visual comfort, but it is correlated with the overall 3D quality. Finally, visual comfort is most correlated to spatial quality and overall 3D quality is most correlated to the spatial quality.

Table 2 SROCC between subjective quality metrics

	Spatial Image Quality	**Depth Quality**	**Visual Comfort**	**Overall 3D Quality**
SIQ	1	0.520	0.891	0.844
DQ	0.520	1	0.429	0.685
VC	0.891	0.429	1	0.765
3DQ	0.844	0.685	0.765	1

3.3.3 Discussion

Seuntiëns [58] proposed that the 3D visual experience can be predicted from combining spatial quality and depth quality. From my results, since visual comfort is highly correlated with spatial quality, overall 3D quality should be able to be predicted

only from spatial quality and depth quality. A linear regression is performed to verify this model with the data. The predicting model is as follows:

$$\overline{Y} = a \cdot SQ + b \cdot DQ + c \cdot VC + d \quad (1)$$

where $\overline{Y}$ is 3D viewing experience, SQ is spatial quality, DQ is depth quality, VC is visual comfort and d is a constant. Following linear regression, the SROCC between $\overline{Y}$ and overall 3D quality is 0.905, which is higher than using only spatial quality to predict overall 3D quality. The regression coefficients have value a = 0.65, b = 0.32, c = 0.35 and d = -17. However, a simpler model using only SQ and DQ:

$$\overline{Y}' = a \cdot SQ + b \cdot DQ + d \quad (2)$$

can achieve the same performance : the SROCC between $\overline{Y}'$ and overall 3D quality is 0.90 and the regression coefficients are a = 0.80, b = 0.64 and d = -6.8. Fig. 8 shows the mean ratings of 3D quality and the predicted ratings from these two linear regression models. From, this figure, one can see that the predicted result from SQ, DQ, and VC is almost identical to the predicted result from SQ and DQ. Thus, one would conclude that that the 3D viewing experience can be predicted using a single linear model from spatial image quality and depth quality for stereoscopic videos.

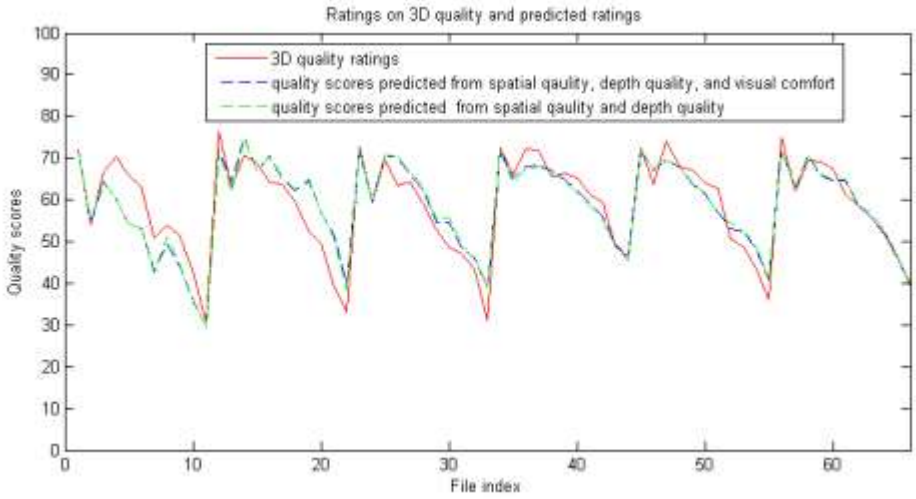

Fig. 8 Mean ratings of 3D quality and the predicted ratings from these two linear regression models

However, the overall 3D quality has significantly lower agreement (0.643) amongst subjects compared to the agreement (0.864) of spatial quality and this finding strongly suggests that we should use two independent quality assessments metrics: spatial quality metric and depth quality metric, in evaluating the quality of 3D content to provide more reliable results across different people. For applications that require QA metrics, such as 3D content encoding and 3D content broadcasting, the geometry setting of the content won't be altered during the encoding or transmission. Only distortions caused either by insufficient bit-rate or packet loss will lower the content quality. Hence, using

the subjective spatial quality scores to evaluate the quality assessment metrics used in these applications will provide more reliable results across subjects.

Table 3 shows the SROCC number of two quality assessment metrics, PSNR and MS-SSIM, evaluated by two different subjective quality scores: spatial image quality and overall 3D quality. The quality scores of the 3D content is simply the average of the predicted quality scores from both views. In Table 3, MS-SSIM has a significantly better performance if the QA metrics are evaluated against spatial image quality, and its performance is statistically indistinguishable from the performance of PSNR against overall 3D quality. Since previous work already pointed out that the MS-SSIM outperforms PSNR in evaluating the 2D quality, this result suggests that without properly modeling overall 3D quality, 2D quality assessment algorithm will perform poorly in predicting the overall 3D quality of stereoscopic videos.

Table 3 SROCC of PSNR and MS-SSIM against spatial quality and overall 3D quality

	PSNR	**MS-SSIM**
Spatial Quality	0.790	0.820
Overall 3D Quality	0.769	0.675

3.4 Conclusion

The analyses in this chapter show that human subjects have coherent opinions on spatial image quality and diverse opinions on depth quality. In addition, compared to predicting overall 3D viewing quality from only spatial image quality, considering both spatial image quality and depth quality can more precisely predict the overall 3D viewing quality. From these observations, the design of a stereo 3D quality assessment algorithm can be divided into two parts - one is designed to predict the spatial image quality and the other one is designed to predict the depth quality. Because it is not clear why more diverse ratings are observed on depth quality, I first focus on predicting the spatial image quality of stereoscopic images in the next three chapters and study the depth quality of distortion-free stereo 3D images in the chapter 7.

CHAPTER 4 MASKING IN DISTORTED 3D STEREO IMAGES

4.1 Introduction

From the previous chapter, one infers that designing a 3D quality assessment algorithm which can precisely predict the spatial image quality of stereoscopic 3D image is an important step towards an ultimate 3D quality assessment algorithm. Later on, this quality assessment algorithm can be combined with a depth quality prediction model to predict the overall 3D quality. However, the spatial image quality of a stereo image-pair is not simply the average of the qualities of the left and right view – there may be some binocular masking and depth masking effects that need to be modeled.

Towards the development of an advanced 3D quality assessment algorithm, this chapter describes study designs, analytic methods and observations to explore masking effects in distorted stereo 3D images. Two studies were conducted toward this goal. These two studies share the same pristine stereo image set and viewing environment, but have different participants, study designs, and analytic methods. After applying varied analytic methods, two observations were drawn from the studies. First, the spatial image quality of a stereo image is dominated by the high quality view for blur, JPEG, and JP2k distorted stereo images, which is related to *binocular suppression*. Second, no depth

masking is observed when viewing distorted stereoscopic presentations, but distortions located at high range variation regions are easier to find, implying a *facilitation effect*. The details of these studies are elaborated in the following sections.

4.2 Stereo image source

The stereo images used for the studies were captured by members of the LIVE lab. They captured co-registered stereo images and range data with a high-performance range scanner (RIEGL VZ-400) with a Nikon D700 digital camera mounted on the top. The stereo images pairs were shot with a 65 mm interocular distance. Off-line corrections were applied later to deal with translations occurring during capture. The sizes of the images are 640 by 360. Fig. 9 is one image pair used in the study, and Fig. 10 shows the ground truth depth map of that image pair. The eight pairs of stereo images to be used in this study were taken on the campus of The University of Texas and a park nearby.

Fig. 9 Example of stereo image pair

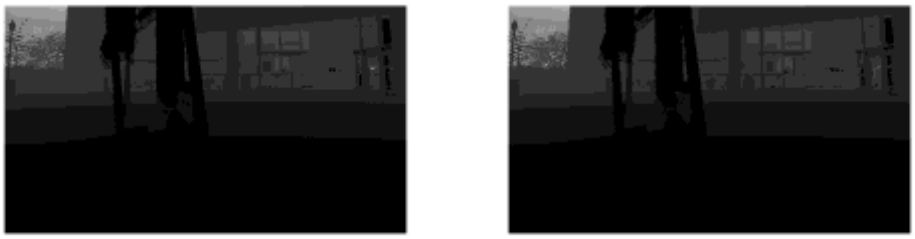

Fig. 10 Range map associated with the stereo image pair in Fig. 9

4.3 Display

A 42” Sharp HD television with a 4-mirror stereo rig was chosen as the display for this study, and the viewing distance was 30”. Although stereoscopic 3D televisions are available now, the 4-mirror stereo rig was used in order to avoid potential crosstalk or reduced luminance problems. No subject reported discomfort in this viewing environment. Fig. 11 shows the setup for a 4-mirror stereo rig and TV.

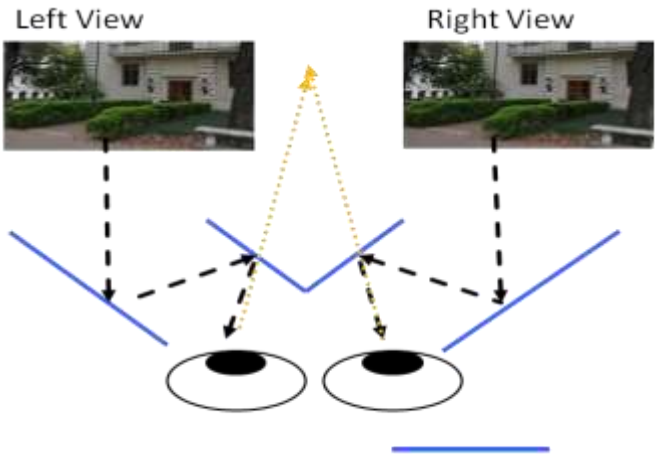

Fig. 11 Illustration of 4-Mirror stereo rig (top view)

4.4 Observers

In both the studies in this chapter, observers were screened before the study for visual acuity using the Snellen test and stereo depth perception using the Randot. Their ages ranged from 20~35 years old. In the first study, nineteen naïve observers (four females and fifteen males) were recruited from The University of Texas student population. In the second study, eighteen naïve observers (five females and fourteen males) participated in the study.

4.5 Study One

4.5.1 Stimuli

Eight stereo image pairs were used to create all stimuli. In order to explore the relationship between the statistics (texture and range) of the 3D data and the perception of

stereoscopic distortions, each pristine stereo pair was distorted within a local area only (128 x 128 square window). To create a locally distorted image, the pristine and global distorted images were blended together within a local patch using a 2D Gaussian weighting function with the standard deviation set as 34 pixels. Fig. 12 is an example of an image with a local blending distortion.

The variables that were controlled to create different stimuli were the distortion type, severity of distortion, and the position of the distortion in each view. Four different distortion types were used: white noise, blur, JPEG compression distortion, and JPEG2000 compression distortions. The degree of severity of the distortion on each image was randomly chosen within a predefined range, from just noticeable to pretty obvious.

Distorted image pairs having random degrees of distortion severity were created in this study. Varying the degree of severity allowed us to probe the distortion conspicuity as a function of both severity and image or range content, possibly revealing insights regarding masking effects under 3D stereoscopic viewing. The locations of the local distortions in the left view and right view were defined in two ways, both random. In the first case, the local distortions were inserted at the same position in both views.

This is defined as ‘’binocular distortion” in this report. In the second case, the distortion was inserted into both images randomly. This is called as “dichoptic distortion” in this report. In total, 136 stimuli were created for the study, including 8 pristine stereo image pairs.

4.5.2 Procedure

We followed the recommendation for a single stimulus continuous quality scale (SSCQS) [55] to decide the time spent by observers locating the local distortion and to supply a subjective quality rating for each stereo image pair. For each stimulus, the subject was first asked to point out the distortion using a mouse cursor. The time subjects spent on the task was recorded. Then the subject was requested to give a subjective 3D image quality rating. Each subject was tested separately. Training sessions were conducted individually before the beginning of each study to verify they were comfortable with our 3D display and to help familiarize them with the user interface used in the task(s) they would complete in the study. The training content was different from the images in the study and was impaired using the same distortion.

Fig. 12 Image with local white noise distortion. The boundary was blended using a Gaussian blending window. When the image was presented, the subject was requested to point out the distortion by clicking the mouse cursor on the distortion.

Fig. 13 When the rating bar showed up, the subject was requested to render a subjective 3D quality opinion on the entire image.

4.5.3 Binocular suppression

This section discusses the effect of binocular suppression on the conspicuity of each distortion type. Fig. 14 shows the Percentage Correct (PC), DMOS and the median Time Spent (TS) by all subjects for both methods of randomly determining the locations of the distortions in both views. The blue solid lines represent the results when the distortions were placed at the same location in both views (binocular distortion) and the red dotted line shows the results when the distortions were inserted at two different random locations in both views (dichoptic distortion). The horizontal axis represents the file index. The dichoptic and binocular distorted stimuli located in the same file index means that the same parameters were employed to distort the two images. In the task of identifying a local distortion on a dichoptic distorted image, locating either one of the local distorted patches is considered as success. We hypothesize that the PC, the DMOS, the TS provide statistical clues regarding the effects of binocular suppression and content masking on distortion conspicuity. If fewer subjects locate a distortion, a higher DMOS is rated, or more time is spent completing a task. This *process* on average provides evidence that distortion was less visible due to masking or suppression.

As shown in Fig. 14, a significant difference between the blue lines (distortions were inserted into the same location in both view) and the red lines (distortions were inserted into two random locations) is apparent. Further, an analysis of variance (ANOVA) was applied to verify the significance and the results were given in Table 4. We can see a significant difference in behavior from the test results from the PC ($p=0.012$), the TS ($p<0.001$) and the DMOS ($p<0.001$). The results suggested that binocular suppression [9] plays an important role in the perception of stereoscopically viewed 3D distortions.

Table 4 The results of ANOVA on the locations of distortions.

Variables	**F ratio**	**p**
PC	6.49	0.012
TS	20.1	<0.001
DMOS	16.39	<0.001

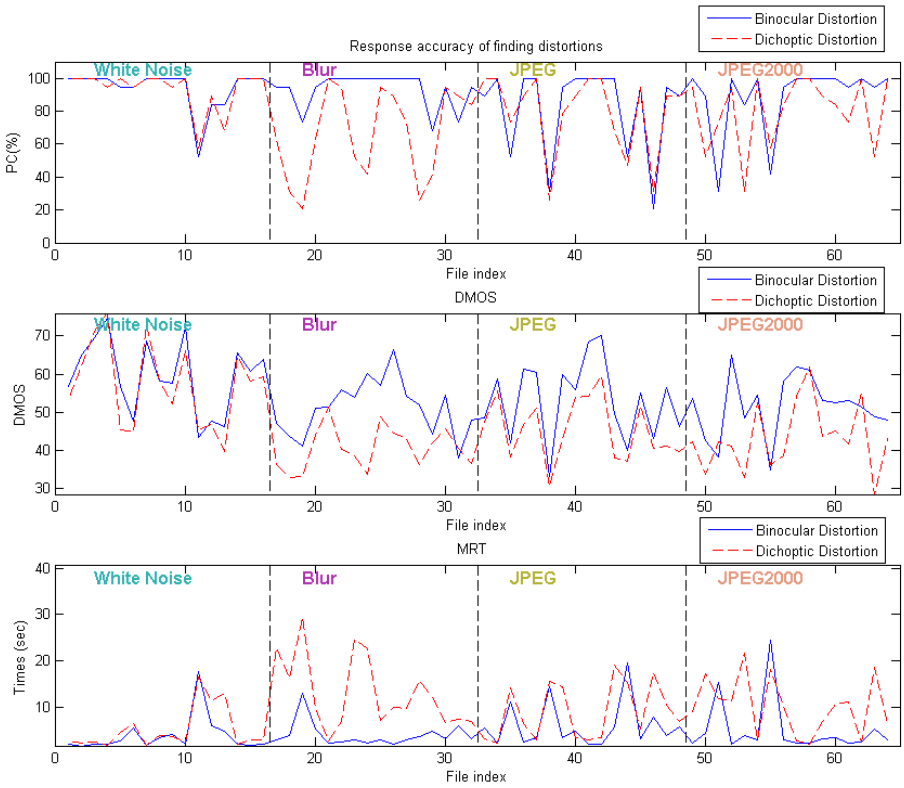

Fig. 14 Plot of PC (top), DMOS (middle), and Times (bottom) in finding distortions

To further examine if binocular suppression is independent from the distortion types, I examined binocular suppression across distortion types. Table 5, Table 6 and Table 7 show the results of ANOVA on PC, TS and DMOS of each distortion type. As shown in Table 5, my observers' behavior differed significantly whilst viewing dichoptic and binocular blur distorted stereo images. The difference of subjects' behaviors indicates that binocular suppression was observed in blur distorted stereo images. Table 6 shows

that binocular suppression was observed in the blur and JP2K distorted stereo images. Table 7 further indicated that blur, JP2K, and JPEG distorted stereo images induced binocular suppression masking.

We did not observe any significant binocular suppression effect for white noise distorted stereo images, whereas binocular suppression appears to have played a significant role in affecting blur, JPEG and JP2K distortion conspicuity. This accords with [28], where the authors pointed out that binocular suppression affects blur distortion, but not MPEG-2 distortion. The results from this study further indicate that binocular suppression also occurs in JPEG and JP2K distorted images which may provide useful pointers for designing asymmetric codecs or quality assessment algorithms.

Table 5 The results of ANOVA of each distortion type on percent correct.

Distortion Type	**F ratio**	**P**
White Noise	0.02	0.882
Blur	13.11	**0.001**
JPEG	0.08	0.775
JP2K	1.39	0.249

Table 6 The results of ANOVA of each distortion type on time spent to find the distortion.

Distortion Type	F ratio	p
White Noise	0.65	0.428
Blur	19.59	**<0.001**
JPEG	2.2	0.149
JP2K	5.43	**0.027**

Table 7 The results of ANOVA of each distortion type on DMOS

Distortion Type	F ratio	p
White Noise	0.42	0.520
Blur	21.5	**<0.001**
JPEG	5.15	**0.031**
JP2K	7.35	**0.011**

4.5.4 Contrast and depth masking

Binocularly Distorted Images

I used an analytic method to analyze the data that was collected without binocular rivalry. The flowchart is shown in Fig. 15. The first step is to divide all test stereopairs of each distortion type into two groups; a High percent correct group (High) and a Low percent correct group (Low). All stimuli were classified into these two groups according to a threshold on the percentage correct, which was set at 85%. Then, a Welch's *t* test was

conducted on the subjects' PC, TS, the range activity within local patches, and the luminance activity within local patches inside each group. Range activity was defined as the weighted mean of the gradient values of the range map inside the patches in both views. Luminance activity was defined as the weighted mean of the gradient values of the image inside the patches in both views. Our assumption is that if there is a contrast masking or range masking effect while the subjects were viewing the stereoscopic 3D image, the subjects' performance on a test should be correlated with the luminance or the range activity of the local patch.

The analysis of luminance and range activity for four types of distortion are shown in Fig. 16 and Fig. 17, the results of Welch's *t* test are also shown and the significance tests are marked. From the results, it appears that luminance activity affects the visibility of white noise and JP2K distortion. For white noise (top left plot in Fig. 16, the result indicates that there is contrast masking when viewing stereoscopic 3D because significantly higher contrast values occur in the low performance group. Regarding luminance activity for JP2K distortion (bottom right plot in Fig. 16), distortions in lower contrast areas tend to be less visible. With respect to range activity, a significant effect on blur distortion was the only effect observed (top right plot in Fig. 17). However, from our

understanding of distortions, JP2K compression distortion is basically blur and ringing distortion. In addition, from the bottom left and the bottom right plots in Fig. 16 and Fig. 17, both blur and JP2K distorted stereo images tend to have higher range activity and higher luminance activities in the high performance group. This suggests that both blur and JP2K distortion are correlated with contrast and range activity in stereoscopically viewed 3D images. To be explicit, the experimental results suggest distortions are more conspicuous in regions with higher luminance or range activities for blur and JP2K distorted stereo 3D images.

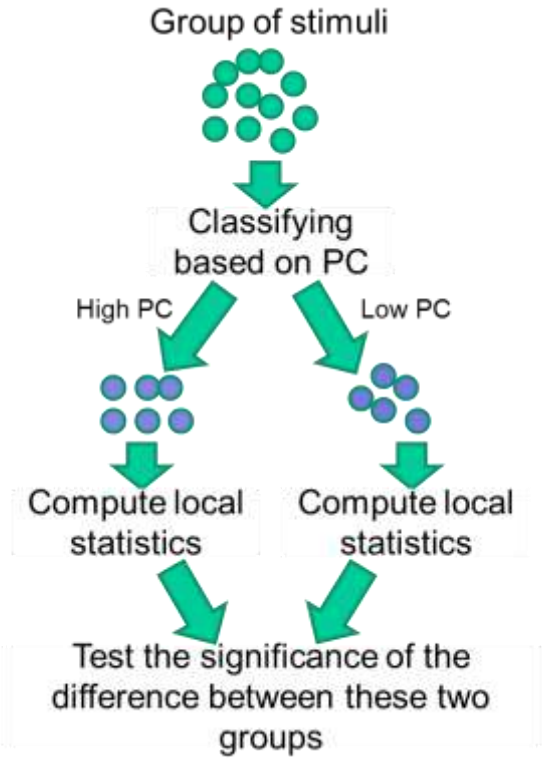

Fig. 15 The analysis flow of discussing contrast and depth masking on binocular distorted stereo images

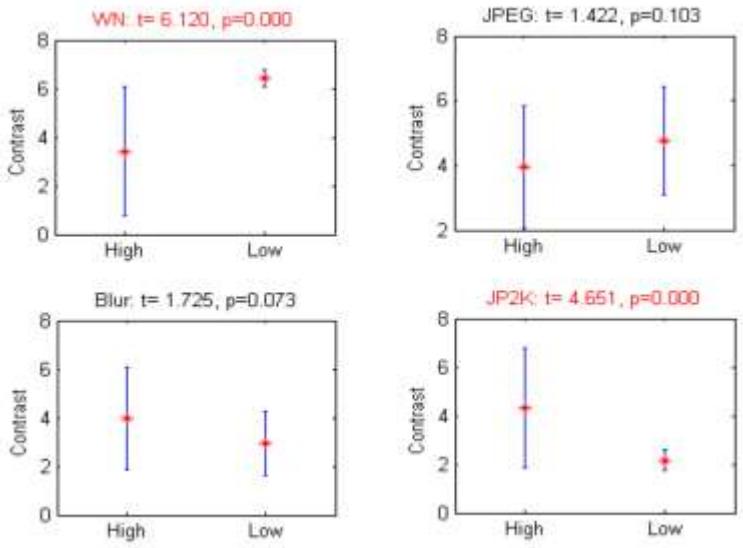

Fig. 16 The results of Welch's t test of white noise (top left), blur (top right), JPEG (bottom left) and JP2K (bottom right) compression distortion on the local contrast. Red dot indicates mean and blue bar represents standard deviation.

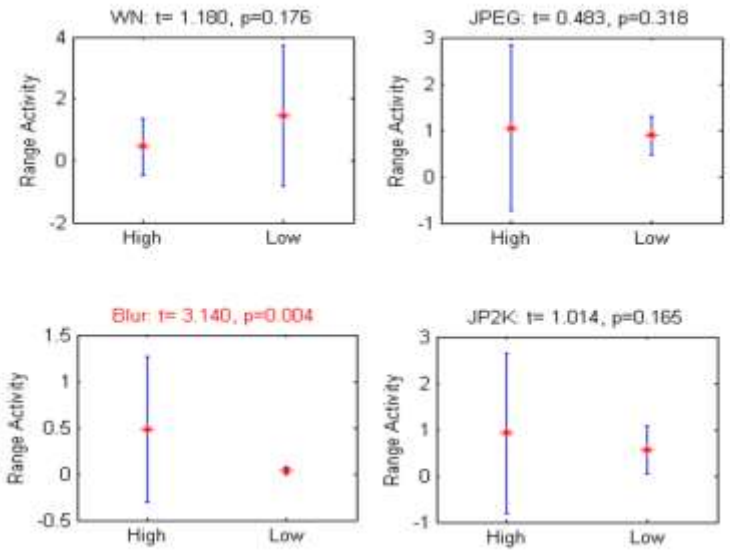

Fig. 17 The results of Welch's *t* test of white noise (top left), blur (top right), JPEG (bottom left) and JP2K (bottom right) compression distortion on the range. Red dot indicates mean and blue bar represents standard deviation.

Dichoptic Distorted Stereo Images

For the data that were collected with binocular suppression, a chi-square test was used to verify whether luminance or range activity had a significant impact on the distortion conspicuity even with binocular masking. In this data, different subjects chose different patches in a single test and the luminance or range activity inside these two patches (left and right) could vary. Hence, the analysis method in the previous section to

analyze the correlation between local image statistics and subjects' performance cannot be applied. A different analysis method was used.

The null hypothesis of the chi-square test indicates that the contrast or range activity in left or right view is not related to the subjects' selections. Namely, the subjects' selections on left view or right view are decided by chance. The null hypothesis was verified by a *t*-test on the overall selections on left and right views. The result is shown in Fig. 18 and the null hypothesis ($p=0.751 > 0.05$) was not rejected. Therefore, one can conclude that the subjects' choices on left view and right view are decided by chance.

Next, a chi-square test was performed to see if contrast or range activity influences the subjects' selections. Table 8 is an example of the chi-square test setup. Table 9 shows the results of all tests on contrast masking. From the table, one can see that the contrast value of local patches is correlated with the visibility of blur ($p=0.003$), JPEG ($p=0.004$), and JP2K ($p=0.03$) distortions. As for the influence of range activity, Table 10 shows that the range activity of local patches is correlated with the visibility of blur ($p<0.001$) and JP2K ($p=0.016$) distortions. Thus, both contrast and range activities have significant influence on the perception of stereoscopically viewed blur and JPEG distortions even with binocular masking. However, it is not clear what kind of correlation

between contrast and range activities and the visibility of dichoptic distorted stereo images occurs from the data in this study.

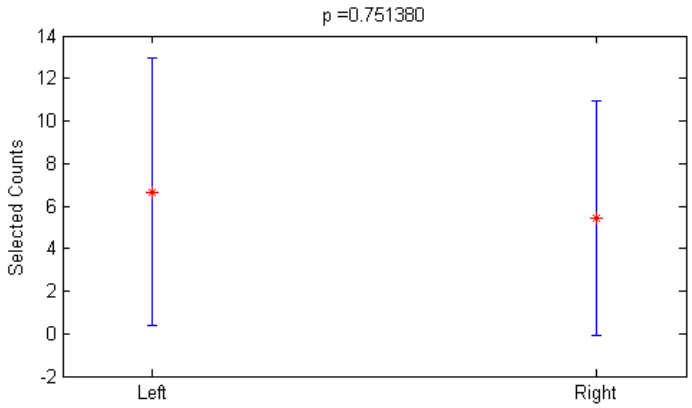

Fig. 18 The results of *t* test on subjects' selection on left and right views.

Table 8 An example of a Chi-square test setup. The number inside brackets is the expected value.

	Subjects Selection		
Contrast on blur	Left	Right	Total
Left > Right	113 (107.72)	33(38.28)	146
Left < Right	8(13.28)	10 (4.72)	18
Total	121	43	164

Table 9 The results of Chi-square test on contrast

Chi-square test on Contrast				
Distortion type	WN	Blur	JPEG	JP2K
Chi value	2.551	8.995	8.400	4.733
p	0.110	**0.003**	**0.004**	**0.030**

Table 10 The results of Chi-square test on range

Chi-square test on Range				
Distortion type	WN	Blur	JPEG	JP2K
Chi value	1.576	18.389	1.532	5.770
p	0.209	**0.000**	0.216	**0.016**

4.6 Study Two

The second study was conducted for two reasons. First, the result in the second study can be used to verify the findings in the first study. Second, while the first study provided significant observations on binocular masking effects, more observations on contrast and range masking of white noise and JPEG compression are needed.

4.6.1 Stimuli

In the first study, the position of a local distortion was randomly decided. This random sampling may fail to sample the area which has both high range variation and high luminance activity. In addition, if there is a masking effect observed in a local

patch with high range and luminance activities, it may be difficult to decide whether the masking is related to luminance activity or to range activity. Therefore, a measurement ('ratio'), defined as range activity divided by luminance activity, is used to create more specific stimuli. Based on this ratio, stimuli where distortions are preferentially inserted into high ratio areas (higher range with lower luminance variation) and low ratio areas were created. In high ratio areas, the assumption is that the visibility of the distortion is more correlated with range activity than the luminance activities. On the other hand, the luminance activities may contribute more to the masking effect in the low ratio areas. The detailed procedure of creating stimuli is described below.

Eight pristine stereo image pairs were used to create all distorted stimuli. For each pristine stereo pair, a distortion was created within a local area (128 x 128 square patch) by the same Gaussian blending method used in the first study. The variables controlled to create the different stimuli were distortion type, display type (binocular distorted 3D images, dichoptic distorted 3D images, and distorted 2D images) and the local statistics (high or low ratio areas). Because the blur distortion is very similar to JPEG2000 compression distortion, only white noise, JPEG compression distortion, and JPEG2000 compression distortions were used. This enabled us to reduce the session durations to

reduce subject fatigue. For each distortion type, a predefined distortion parameter was chosen to create all locally distorted stimuli. The severity of each distortion was fixed to reduce the number of variables. Finally, 48 stimuli were created for each distortion type yielding 144 stimuli in total.

4.6.2 Procedure

The same GUI and SSCQS method mentioned in the first study was used. However, a few modifications were made to fit the study design. First, the stimuli were subdivided into three groups according to distortion type. The same group of subjects participated in the study on three different days. They viewed one distortion type each day and a gap of at least one day separated two consecutive sessions. This design further avoided subject fatigue. With this design, all observers were able to finish each session within a reasonable time. The maximum time that was used by a subject to finish a session was 37.4 minutes and the average time to finish a session was about 20 minutes (the time spent in training sessions is not included). A short training session was given to each subject before conducting each session. The content shown in each training session was different than the source images used in the study.

4.6.3 Binocular suppression

By analyzing the observers' performance in finding local distortions, binocular suppression was again probed. Fig. 19, Fig. 20 and Fig. 21show the Percentage Correct (PC), the median Time Spent (TS) by all subjects for binocular distorted images (blue solid line) and dichoptic distorted images (red dotted line). The *p* values shown in these figures are the ANOVA tested results. Looking Fig. 21, there is no significant difference for these two types of distorted images. Hence, it appears that there is no binocular suppression effect in white noise distorted stereo images. However, significantly different behavior can be observed from Fig. 20 and Fig. 21, which indicate that there is a binocular suppression effect that caused the subjects to behave differently, since the other variables were controlled.

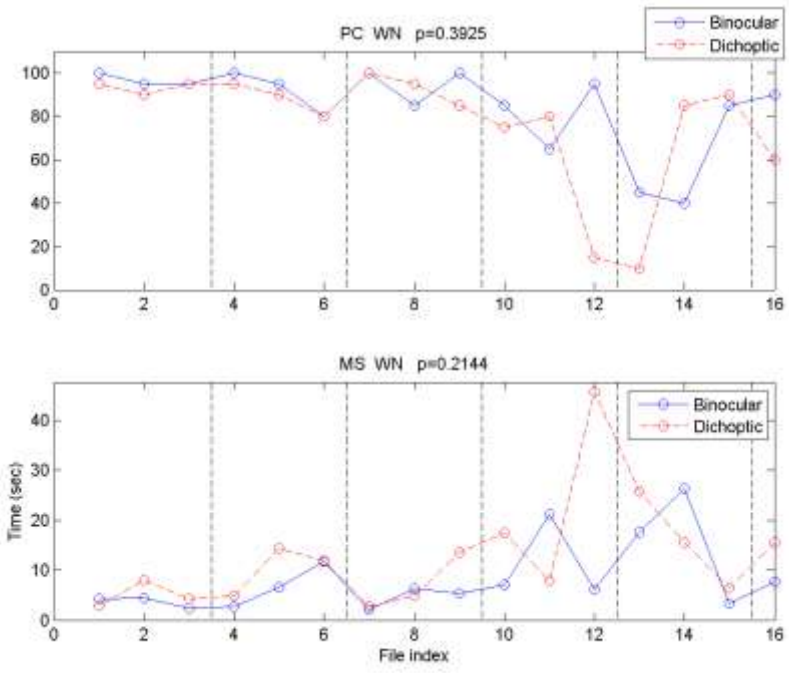

Fig. 19 The PC and MS of dichoptic distorted 3D image and binocular distorted 3D image. The distortion type is White Noise

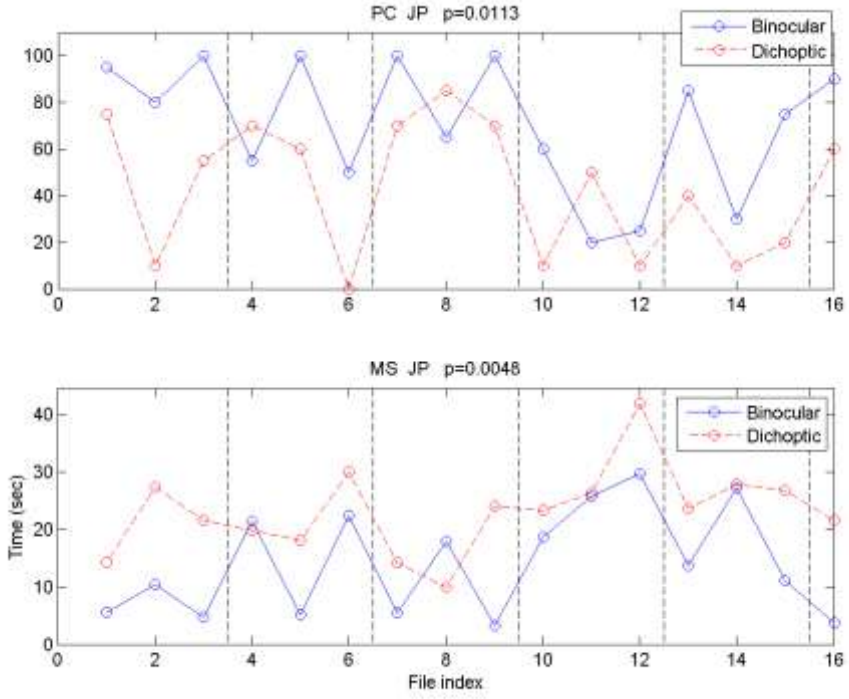

Fig. 20 The PC and MS of dichoptic distorted 3D image and binocular distorted 3D image. The distortion type is JPEG compression distortion

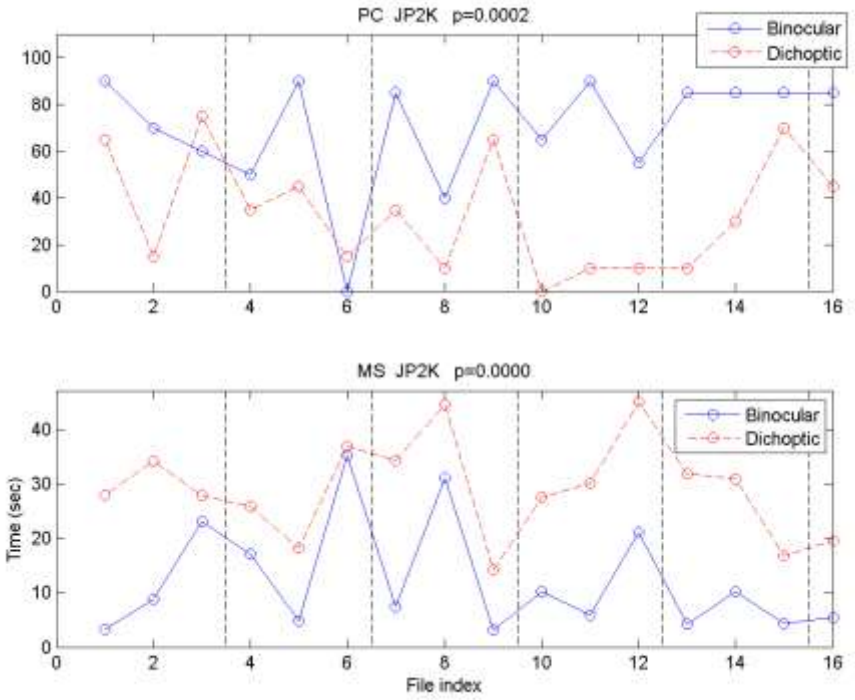

Fig. 21 The PC and MS of dichoptic distorted 3D image and binocular distorted 3D image. The distortion type is JP2K compression distortion

4.6.4 Contrast and depth masking

Binocularly Distorted Images

The distortions in the stimuli were placed according to their local statistics. Fig. 22 shows the results for three types of distortions. In Fig. 22, the results of the high ratio group are plotted with blue solid lines while the results of the low ratio group are plotted with red dotted lines. An ANOVA analysis on the high and the low groups was done and the *p* values are shown in these figures. Based on the results shown in Fig. 22, significant

differences were observed in JP2K and JPEG compression distortions. Hence, we conclude that there is actually an anti-masking, or *facilitation effect* for visibility of JP2K and JP compression distorted images. For white noise distortion, there was no significant result observed. It is worth noting that JP2K and JP are both *nonlinear* distortions, unlike additive noise.

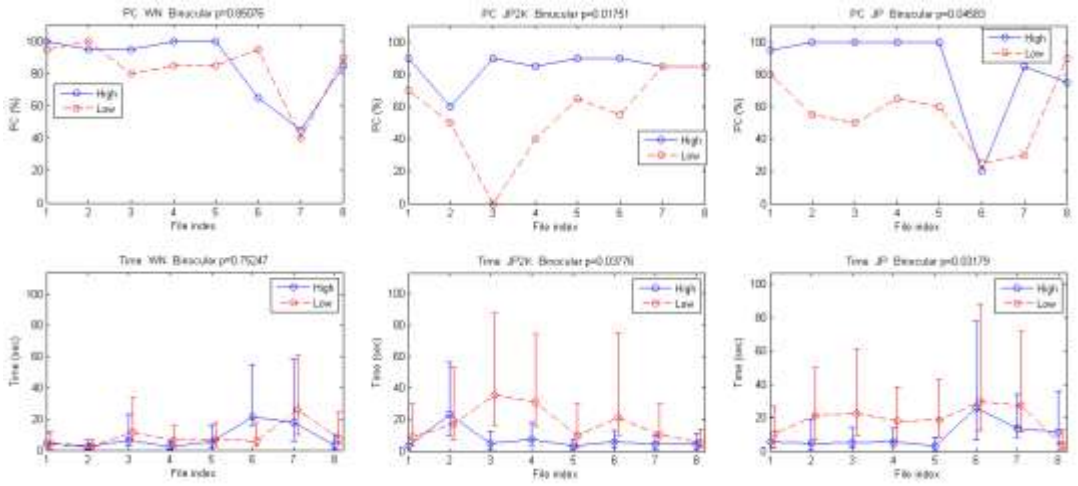

Fig. 22 The PC and MS of binocular distorted images. The blue solid lines show results for stimuli where the distortion is placed at high ratio areas, and the red dotted lines show results when the distortion is placed at low ratio areas.

Dichoptic Distorted Images

From our results in previous sections, we claim that there is binocular suppression, which reduces the visibility of a distortion, and facilitation effect, which

make a distortion more visible, for JPEG and JP2K distorted images. This section discusses what was observed when these two effects co-exist in dichoptic distorted images.

A similar analysis as that performed for the binocular case was also conducted on the dichoptic distorted stereo images, and the results are shown in Fig. 23. The results of the high ratio group are plotted with blue solid lines and the results of the low ratio group are plotted with red dotted lines. As seen in Fig. 23, there is a significant difference observed in JP2K distorted images and there is no significant difference for JPEG distorted images. This result matches our findings in section 4.5.4, where our analysis pointed out that there is some masking or facilitation effects on dichoptic JP2K/blur distorted images. Due to the design of the first study, we were not able to determine what kind of effects is observed in dichoptic distorted images. However, the design of the study two provides deeper insights into effects on dichoptic distorted images. Fig. 23 indicates that the effect is a facilitation effect.

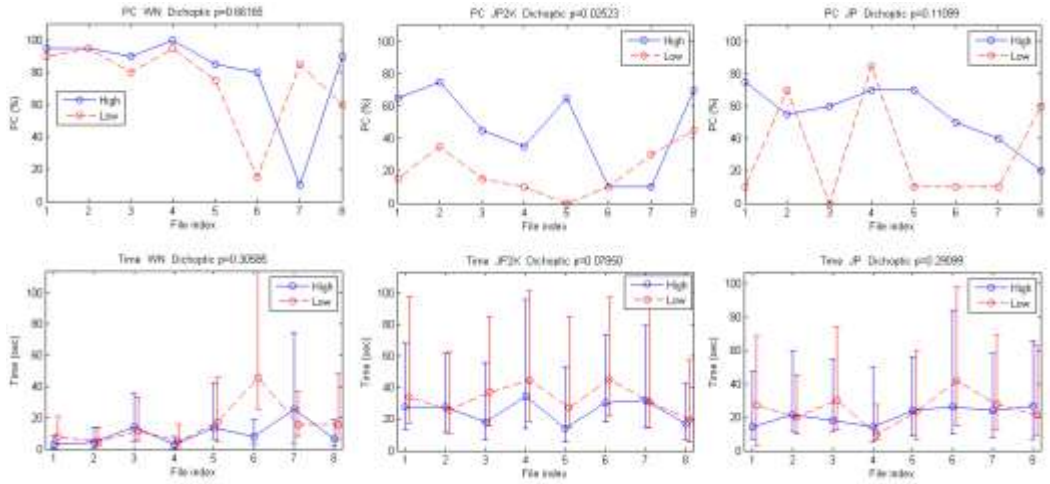

Fig. 23 The PC and MS of dichoptic distorted images. Blue solid lines show the results for stimuli where the distortion is placed at high ratio areas, and the red dotted lines show results when the distortion is placed at low ratio areas.

4.7 Conclusion

I discovered that there is binocular masking for blur, JPEG, and JP2k distorted images. Further, there is a facilitation effect corresponding to depth activity for blur, JP2K and JPEG binocular distorted images. For dichoptic distorted images, a facilitation effect was only observed for JP2K/blur distorted images. Finally, we observed contrast masking for white noise distorted images in the first study; however there was no contrast masking observed in the second study.

The observations found in this chapter strongly suggest that binocular masking effect (binocular rivalry) should be modeled to predict the spatial image quality for stereoscopically viewing. In the next chapter, a 3D quality assessment framework will be introduced. This framework is based on earlier work in modeling the binocular rivalry proposed by Levelt [9].

CHAPTER 5 FULL REFERENCE QUALITY ASSESSMENT OF STEREOPAIRS ACCOUNTING FOR RIVALRY

5.1 Introduction

This chapter deals with the design of a full reference (FR) stereoscopic 3D image quality assessment (QA) framework. As mentioned in chapter 2, research on stereo 3D QA algorithms has been conducted for years, but most of these 3D QA algorithms are still based on ad-hoc techniques. From the discussion in chapter 3 and chapter 4, one would conclude that the task of predicting the perceived 3D quality of a 3D stereopair involves predicting spatial image quality and depth quality. Between these two factors, spatial image quality is clearly affected by binocular masking effect, and it is not yet clear how spatial distortion affects the perceived quality of depth. As a first step towards the creation of a high performance 3D QA algorithm, I first design a 3D QA algorithm that accounts for the binocular masking effect.

This chapter proposed a framework that allows for easy extension of 2D FR algorithms, providing a plug-and-play approach to the development of 3D FR QA algorithms. This framework attempts to model the fact that binocular masking alters perceived 3D quality, as described in chapter 4. The proposed framework uses the linear model proposed by Levelt [9] to synthesize an intermediate view from the stereo image

pair, called the *cyclopean view*. Levelt's work models the binocular masking effect with local statistics of a stereo image pair. Thus, the synthesized cyclopean view is visually close to the true Cyclopean image a human subject recreates in his brain, while viewing a stereo image pair on a stereoscopic display. Finally, experimental results show that higher performance can be achieved in predicting the perceived 3D quality of stereo 3D image by applying existing 2D QA algorithms on the synthesized cyclopean view.

5.2 Binocular rivalry/suppression

Binocular rivalry is a perceptual effect that occurs when the two eyes view mismatched images at the same retinal location(s). Here, 'mismatch' means that the stimuli received by the two eyes are sufficiently different from each other to cause match failures or to otherwise affect stereoperception. Failures of binocular matching trigger binocular rivalry, which is experienced in various ways, i.e., a sense of failed fusion or a bi-stable alternation between the left and right eye images. Fig. 24 shows an example of binocular rivalry when mismatched stimuli are present. In Fig. 24, in the interval (t_0, t_1), the observer saw the stimulus from the left eye (the arrow). Then, the stimulus from the right eye (the star) dominated until time t_2, after which the observer again saw an arrow.

This fluctuation continues when an observer is experiencing a binocular rivalry. The fluctuation period may vary from a fraction of a second to several seconds, and it may depend on the color, shape, and texture of the stimuli. Binocular suppression [59] is a special case of binocular rivalry. When binocular suppression is experienced, no rivalrous fluctuations occur between the two images when viewing the mismatched stereo stimulus. Instead, only one of the images is seen while the other is hidden from conscious awareness. Fig. 25 shows an example of binocular suppression.

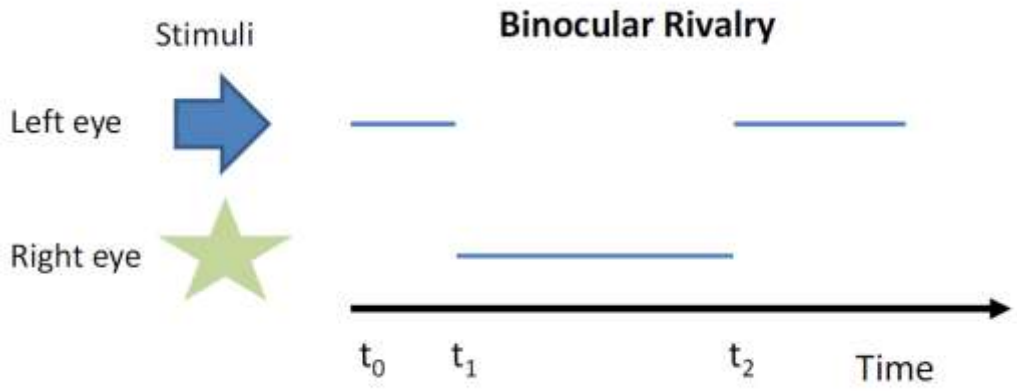

Fig. 24 Illustration of binocular rivalry: Two different patterns are presented to the left eye (an arrow) and the right eye (a star). The blue line indicates that the stimulus is perceived by a human observer inside that time interval.

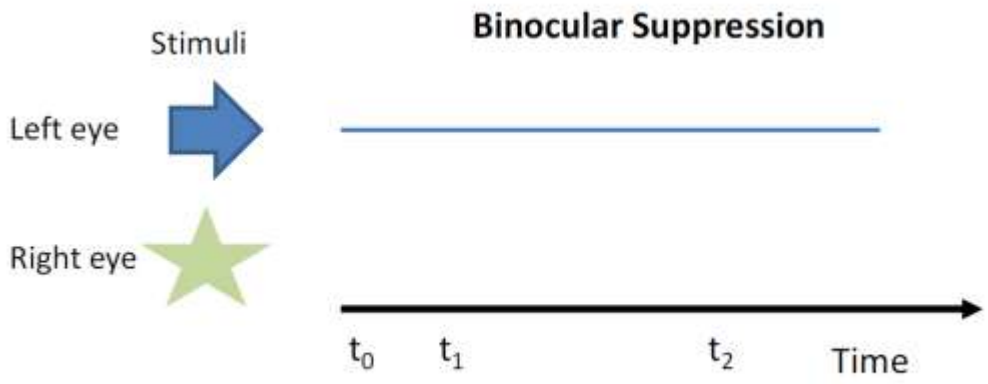

Fig. 25 Illustration of binocular suppression: Two different patterns are presented to the left eye (an arrow) and the right eye (a star). An observer only sees the arrow when s/he experiences binocular suppression.

Numerous studies have been conducted towards understanding binocular rivalry/suppression. Currently, three different models are prevalent: early suppression, late (high-level) suppression, and a hybrid model including both early and late processes. The early suppression model [9, 59-62] suggests that binocular rivalry is the result of competition between the eyes. This model views rivalry as an early visual process involving reciprocal inhibition between the monocular channels.

Research that supports a high level suppression model [63, 64], on the other hand, argues that there is very little correlation between neural activity and perceptual alternations in area V1 of visual cortex. Moreover, there are some early cortical neurons

whose activity is anti-correlated with binocular perception; this means that these neurons fire more when acting on suppressed stimuli. Hence, it has been claimed that the rivalry model should be high-level. For example, Alais and Blake [65] showed that grouping information may contribute to binocular rivalry. Finally, since both early and late models are supported by some evidence, more recent research [11] suggests that a hybrid model may be the best explanation. However, these ideas have not previously been applied towards understanding how binocular rivalry might be related to distortion type.

Another important finding of binocular rivalry/suppression is that it is a nearly independent local process. A series of papers [65-67] discuss whether the binocular rivalry zones function independently, and their findings indicate that binocular rivalry is composed of local processes. In addition, evidence [68, 69] supports the contention that the size of this local process is governed by the size of receptive fields in early visual cortex. The discussions in this section provide basic concepts that are used in the 3D QA framework which is introduced in the next section.

5.3 A framework for stereo quality assessment

The logical goal of a 3D stereoscopic QA model is to estimate the quality of the true cyclopean image formed within an observer's mind when presented with a stereo

image are presented with a stereopair. Of course, simulating the true *cyclopean* image [15] associated with a given stereopair is a daunting task, since it would require accounting for the display geometry, the presumed fixation, vergence, and accommodation. This task is already herculean, and is compounded by the fact that it is still unclear how a cyclopean image is formed! Towards a limited approximation of this goal, however, we seek to synthesize an internal image having a quality level that is close to the quality of the true *cyclopean* image. By way of notation, henceforth we still use the term "cyclopean" image to represent the synthesized image and *cyclopean* image to mean the one formed in the observer's mind. By performing 3D quality assessment on the "cyclopean" image we hope to produce accurate estimates of 3D quality perceived on the *cyclopean* image.

The concept underneath the model framework is shown in Fig. 26. Given a stereo image, an estimated disparity map is generated by a stereo algorithm, while Gabor filter responses are generated on the stereo images using a bandpass filter bank. A “cyclopean” image is synthesized from the stereo image pair, the estimated disparity map, and the Gabor filter responses. A “cyclopean” image is created from the reference stereopair and another "cyclopean" image is calculated from the test stereopair. Finally, full

reference 2D QA models are applied to the two "cyclopean" images to predict 3D quality scores.

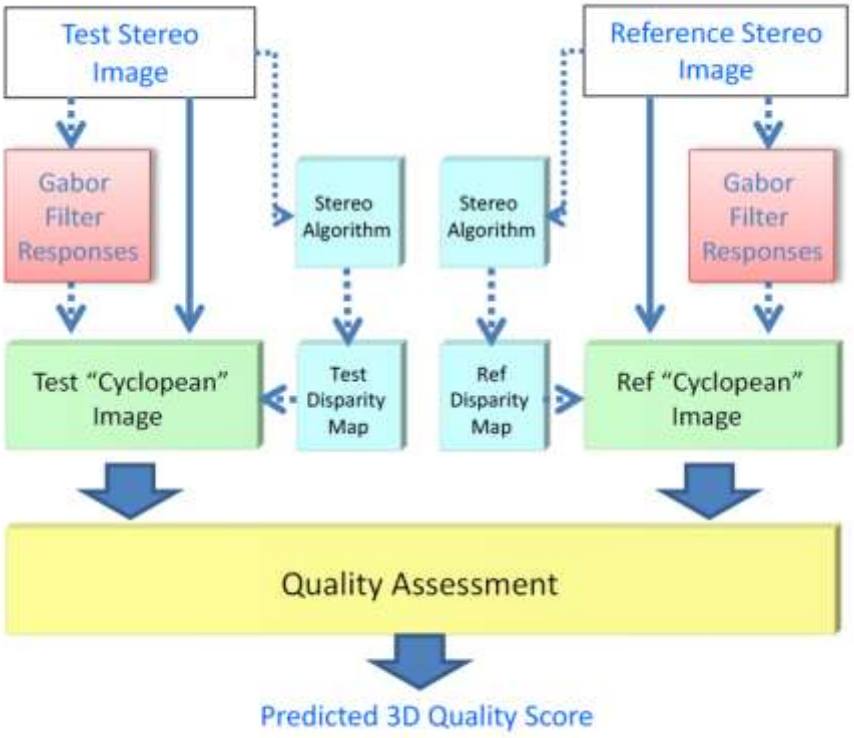

Fig. 26 The proposed framework for 3D QA

5.3.1 Stereo matching algorithms

Research on stereo algorithm design has been a topic of intense inquiry for decades. However, there is no consensus on the type of stereo matching algorithm that should be used in 3D QA other than it be of low complexity. Further, there is scarce

literature on the performance of stereo algorithms operating under different distortion regimens. Therefore, we deploy a variety of these efficient stereo depth-finding algorithms differing considerably in their operational constants along with the framework we described above to assess perceived 3D quality.

In order gain insights into the influence of stereo algorithms on the performance of 3D QA models, three stereo algorithms were selected based on their complexity and performance. In general, better stereo algorithms (based on results on the Middlebury database [70]) have higher computational complexity, and we balanced this tradeoff in the choice of stereo matching models.

The first algorithm has the lowest complexity. It uses a very simple sum-of-absolute difference (SAD) luminance matching functional without a smoothness constraint. The disparity value of a pixel in a stereopair is uniquely computed by minimizing the SAD between this pixel and its horizontal shifted pixels in the other view with ties broken by selecting the lower disparity solution. The second algorithm [71] has the highest complexity among the three models. This segmentation-based stereo algorithm delivers highly competitive results on the Middlebury database [70]. The third is a SSIM based stereo algorithm that uses SSIM scores to choose the best matches. The

disparity map of a stereopair is generated by maximizing the SSIM scores between the stereopiar along the horizontal direction, again resolving ties by a minimum disparity criterion.

5.3.2 Gabor filter bank

As discussed earlier, when the two images of a stereopair present different degrees or characteristics of distortion, the subjective quality of the stereoscopically viewed 3D image generally cannot be predicted from the average quality of the two individual images. Binocular rivalry is a reasonable explanation for this observation. Levelt [9] conducted a series of experiments that clearly demonstrated that binocular rivalry/suppression was strongly governed by low-level sensory factors and the Cyclopean image could be modeled by a linear model from stereo stimuli. He used the term stimulus strength, and noted that stimuli that were higher in contrast, or had more contours, tend to dominate the rivalry. Inspired by his result, we use the energy of Gabor filter bank responses on the left and right images to model stimulus strength and to simulate rivalrous selection of "cyclopean" image quality.

The Gabor filter bank extracts features from the luminance and chrominance channels. These filters closely model frequency-orientation decompositions in primary

visual cortex and capture energy in a highly localized manner in both space and frequency [72]. The complex 2-D Gabor filter is defined as

$$G(x, y, \sigma_x, \sigma_y, \zeta_x, \zeta_y, \theta) = \frac{1}{2\pi\sigma_x\sigma_y} e^{-\frac{1}{2}\left[\left(\frac{R_1}{\sigma_x}\right)^2 + \left(\frac{R_2}{\sigma_y}\right)^2\right]} e^{i(x\zeta_x + y\zeta_y)} \quad (3)$$

where $R_1 = x\cos\theta + y\sin\theta$ and $R_2 = -x\sin\theta + y\cos\theta$, σ_x and σ_y are the standard deviations of an elliptical Gaussian envelope along x and y axes, ζ_x and ζ_x are the spatial frequencies of the complex sinusoidal grating, and θ is the orientation. The design of the Gabor filter was based on the work conducted by Su, et al. [73]. The local energy is estimated by summing Gabor filter magnitude responses over four orientations (horizontal, both diagonals, and vertical (90 degrees) at a spatial frequency of 3.67 cycles/degree, under the viewing model described in Section 5.4.3.

Regarding the choice of the spatial center frequency, Tyler [74] pointed out that the depth signal in human vision is analyzed over a much smaller band-width than is the luminance channel. In addition, Schor, *et al.* [17] found that the stereoacuity of human vision normally falls off quickly when seeing stimuli dominated by spatial frequencies lower than 2.4 cycles/degree. Based on their findings, using filters having spatial center

frequencies in the range from 2.4 to 4 cycles/degree should produce responses to which a human observer would be most sensitive.

5.3.3 Cyclopean image

A linear model was proposed by Levelt [9] to explain the experience of *binocular rivalry* in perceived Cyclopean when a stereo stimulus is presented. The model he proposed is:

$$w_l E_l + w_r E_r = C \tag{4}$$

where E_l and E_r are the stimuli to the left and the right eye respectively, w_l and w_r are weighting coefficients for the left and the right eye that are used to describe the process of binocular rivalry, where $w_l + w_r = 1$,and C is the Cyclopean image.

Given that a foveally presented monocular stimulus generally does not disappear spontaneously, he hypothesized that the duration of a period of dominance period of an eye does not depend on the strength of the stimulus presented to that eye, but rather on the stimulus strength presented to the other eye. Therefore he concludes that the experience of binocular rivalry is not correlated to the absolute stimulus strength of each view, but is instead related to the relative stimulus strengths of two views. He also proposed a model whereby the weighting coefficients are positively correlated with the

stimulus strengths, which we embody in a biologically plausible model whereby the local energies of the responses of a bank of Gabor filters are used to weight the left and right image stimuli. Since binocular rivalry is a local phenomena, broading Levelt's model in this manner is a natural way to simulate a synthesized cyclopean image. In our model, as in Levelts; the stereo views used to synthesize to the cyclopean view are disparity-compensated. Thus the localized linear model that we use to synthesize a cyclopean image is:

$$CI(x,y) = W_L(x,y) \times I_L(x,y) + W_R((x+d),y) \times I_R((x+d),y) \tag{5}$$

where CI is the simulated "cyclopean" image, I_L and I_R are the left and right images respectively, d is a disparity value which matches a local stimulus of the left view to the corresponding stimulus of the right view. The weightings W_L and W_R are computed from the normalized Gabor filter magnitude responses:

$$\begin{aligned} W_L(x,y) &= \frac{GE_L(x,y)}{GE_L(x,y) + GE_R((x+d),y)} \\ W_R(x+d,y) &= \frac{GE_R((x+d),y)}{GE_L(x,y) + GE_R((x+d),y)} \end{aligned} \tag{6}$$

where GE_L and GE_R are the summation of convolution responses of the left and right images to filters of the from (3). Because of the normalization in (6), increasing the Gabor energy on the right view suppresses the dominance of the left view when there is

binocular rivalry. Finally, the task of 3D QA is performed by applying a full reference 2D QA algorithm on the reference ``cyclopean'' image and the test “cyclopean” image.

Fig. 27 shows an example of a synthesized cyclopean image. The sterepair in Fig. 27 are locally distorted by a white noise patch located at different location. Since the white noise distortion produces an elevated stimulus strength, the synthesized cyclopean image is dominated by white noise while approximates the experience when stereoscopically viewing the stereopair.

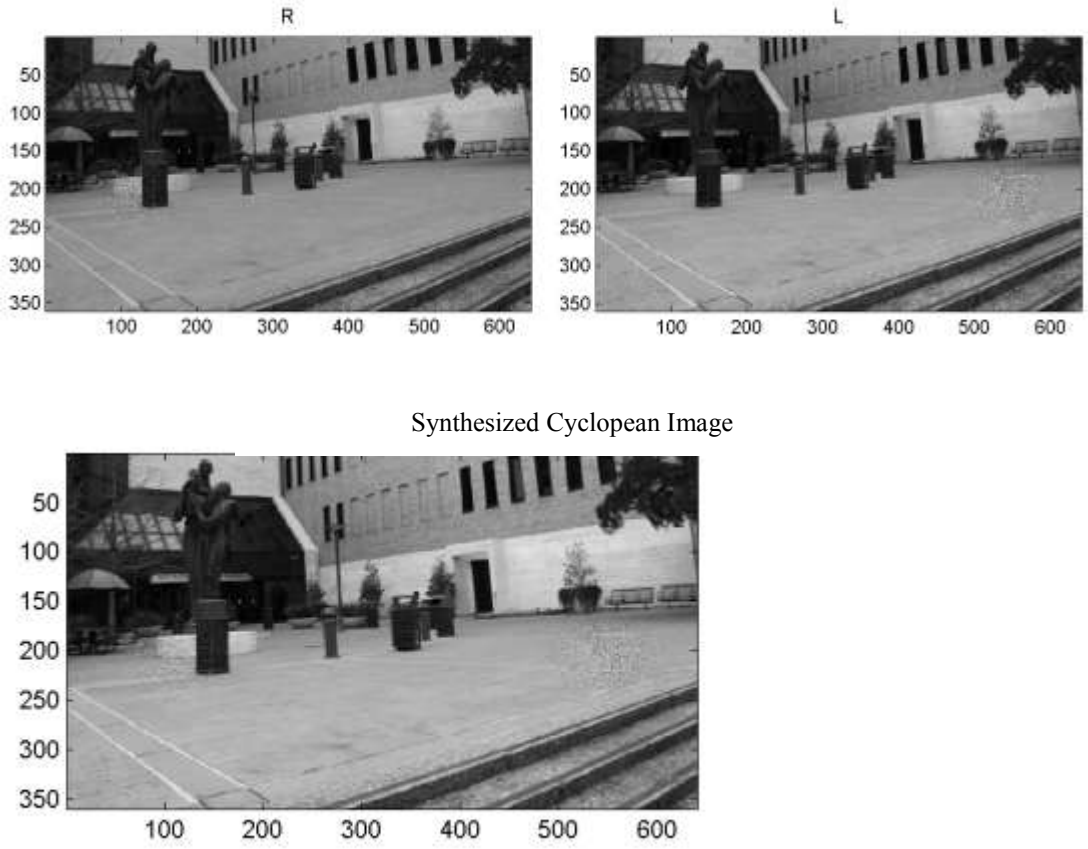

Fig. 27 A stereo image distorted by white noise (free-fused) and the cyclopean image created by the proposed algorithm.

5.4 Experiment

A human study was conducted to construct a subjective data set to be used in assessing algorithms of the proposed 3D QA framework. This section describes the human study and experiments performed using it.

5.4.1 Stereoscopic image quality dataset

A stereoscopic image quality dataset annotated with associated subjective quality ratings was constructed using the outcomes of a human study. The details of the dataset and human study are described in the following.

5.4.1.1 Source image

The stereo images used for the study were captured by members of the LIVE lab. They captured co-registered stereo images and range data with a high-performance range scanner (RIEGL VZ-400[73]) with a Nikon D700 digital camera mounted on the top. The stereo images pairs were shot with a 65 mm camera base distances. Off-line correction was later applied to deal with translations occurring during capture. The sizes of the images are 640 by 360 pixels. The eight pristine images are shown in Fig. 28 while Fig. 29 shows the ground truth depth map of one of them. The eight pairs of stereo images to be used in this study were taken on the campus of The University of Texas at Austin and a nearby park. The ground truth depth map of each stereopair was transformed to a ground truth disparity map based on the captured model described above.

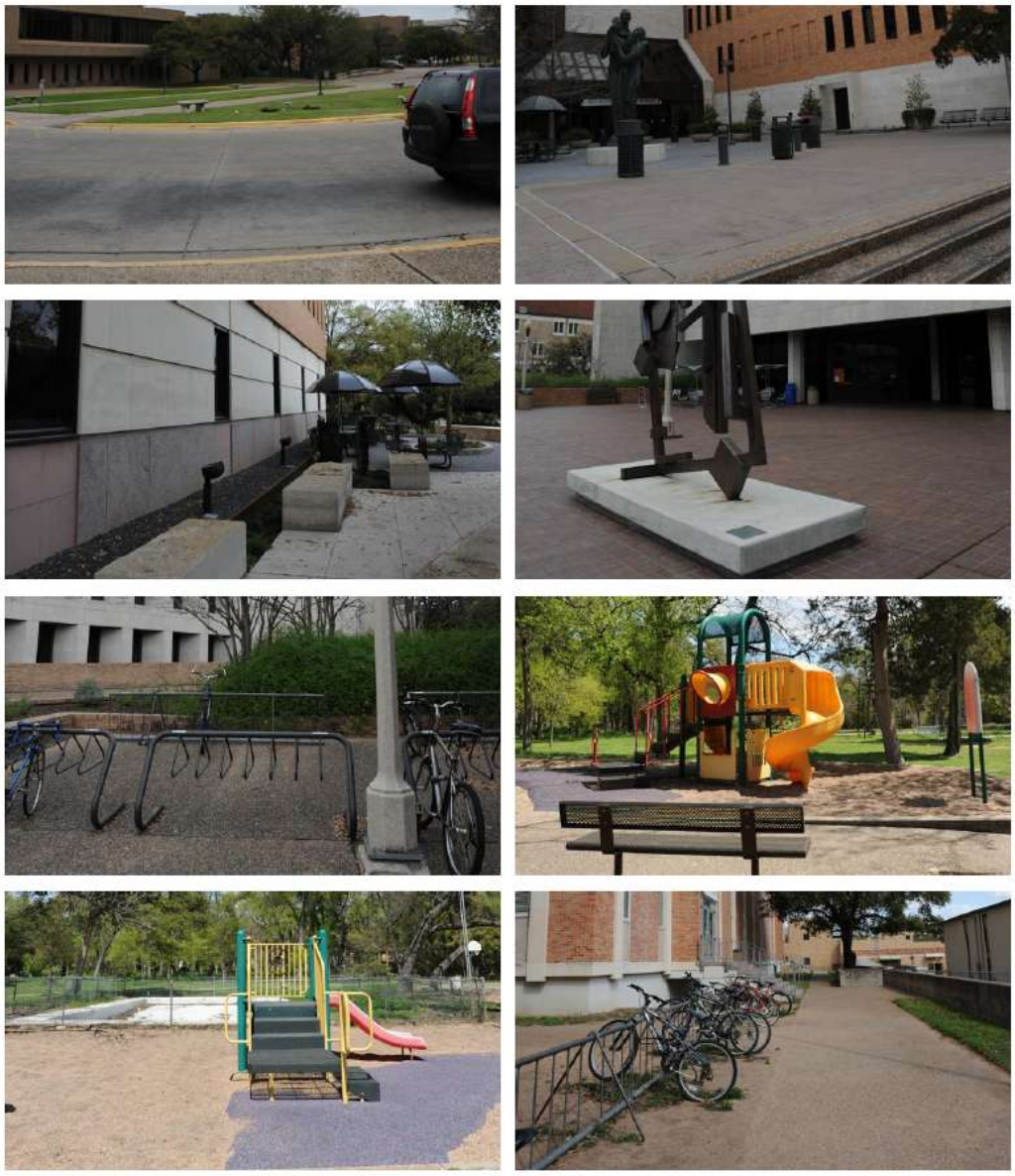

Fig. 28 The eight stereo images used for the database

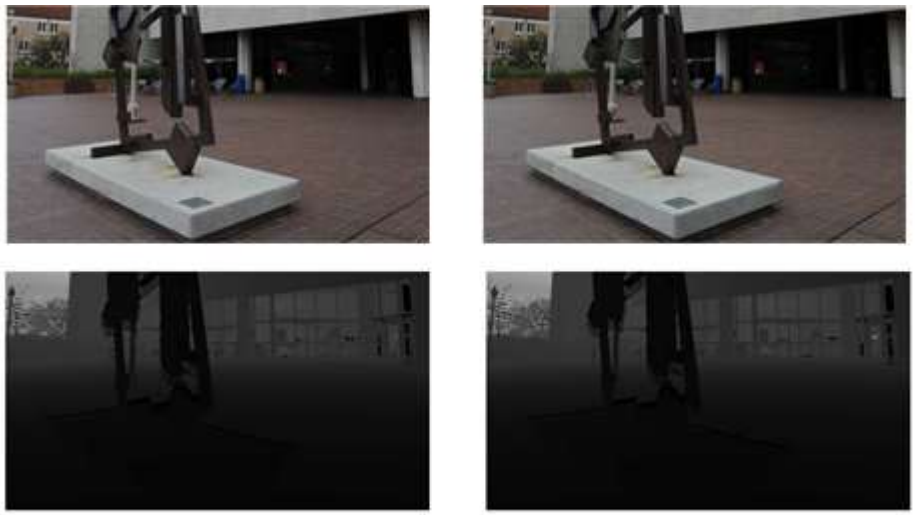

Fig. 29 A stereo image (free-fuse the left and right images) and the ground truth disparity maps.

5.4.1.2 Participants

Six females and twenty-seven males participated in the experiment all aged between 22 and 42 years. A Randot stereo test was used to pre-screen participants for normal stereo vision. Each subject reported normal or corrected normal vision and no acuity or color test was deemed necessary.

5.4.1.3 Display setting

The study was conducted using a Panasonic 58" 3D TV with active shutter glasses. The viewing distance was set at 116 inches, which is four times the screen height.

5.4.1.4 Stimuli

Both symmetric and asymmetric distortions were generated. The distortions that were simulated include compression using the JPEG and JPEG2000 compression standards, additive white Gaussian noise, Gaussian blur and a fast-fading model based on the Rayleigh fading channel. The degradation of stimuli was varied by control parameters within pre-defined ranges; the control parameters are reported in Table 11. The ranges of control parameters were decided beforehand to ensure that the distortions varied from almost invisible to severely distort with good overall perceptual separation. For each distortion type, every reference stereopair was distorted to create three symmetric distorted stereopairs and six asymmetric distorted stereopairs. Thus, a total of 360 distorted stereopairs were created.

Table 11 Range of parameter values for distortion simulation

Distortion	Control Parameter	Range
WN	Variance of Gaussian	[0.001 0.5]
Blur	Variance of Gaussian	[0.5 30]
JP2K	Bit-rate	[0.04 0.5]
JPEG	Quality parameter	[8 50]
FF	Channel signal-to-noise ratio	[15 30]

5.4.1.5 Procedure

We followed the recommendation for a single stimulus continuous quality scale (SSCQS) to collect the "3D subjective image quality" of each distorted stereoscopic image. The instructions given to each participant was: Give an overall rating based on your viewing experience when viewing the stereoscopic stimuli. The obtained ratings were obtained on a continuous scale labelled by equally spaced adjective terms: bad, poor, fair, good, and excellent, i.e. a Liekart scale.

The experiment was divided into 2 sessions; each held to less than 30 minutes to minimize subject fatigue. A training session using six stimuli was conducted before the beginning of each study to verify that the participants were comfortable with the 3D display and to help familiarize them with the user interface used in the task. The training content was different from the images in the study and was impaired using the same distortion. Questions about the experiment were answered during the training session and a short post-interview was conducted to determine whether the participant experienced visual discomfort during the experiment. Only two participants reported any visual discomfort.

5.4.1.6 Subjective quality scores

Differential opinion scores (DOS) were obtaining by subtracting the ratings that the subject gave each reference stimuli from the ratings that the subjective gave to the corresponding test distorted stimuli. The remaining subjective scores were then normalized to Z-scores, and then averaged across subjects to produce difference mean opinion scores (DMOS).

5.4.2 Results

5.4.2.1 Performance using ground truth disparity map

I studied four widely-used full-reference 2D QA metrics (PSNR, SSIM [29], VIF [33], and MS-SSIM [30]) as candidate 2D QA methods to be used in the 3D QA framework. We used Spearman's rank ordered correlation coefficient (SROCC), the linear (Pearson's) correlation coefficient (LCC) and the root-mean-squared error (RMSE) to measure the performance of 3D QA models. LCC and RMSE were computed after logistic regression through a non-linearity which is described in [75] . Higher SROCC and LCC values indicate good correlation with human perception, while lower values of RMSE indicate better performance.

I begin the performance analysis by using ground truth depth, which minimizes the effects of flaws in the stereo matching algorithms. The performance numbers are

shown in Table 12, Table 13 and Table 14. Also included are the performance values arrived at using the same 2D FR QA algorithms, simply applied to the left and right views and the QA scores averaged, called "2D Baseline". The cyclopean model QA framework does significantly better than the 2D baseline QA algorithms on the mixed data set containing both symmetric and asymmetric distorted data.

Table 12 SROCC scores obtained by averaging left and right QA scores (center column) and using the 3D "cyclopean" model (right column)

	2D Baseline	**Cyclopean Model**
PSNR	0.672	0.762
SSIM	0.796	0.856
MS-SSIM	0.78	**0.901**
VIF	0.822	0.864

Table 13 LCC scores obtained by averaging left and right QA scores (center column) and using the 3D "cyclopean" model (right column)

	2D Baseline	**Cyclopean Model**
PSNR	0.687	0.783
SSIM	0.804	0.867
MS-SSIM	0.784	**0.908**
VIF	0.844	0.872

Table 14 RMSE values obtained by averaging left and right QA scores (center column) and using the 3D "cyclopean" model (right column)

	2D Baseline	**Cyclopean Model**
PSNR	17.67	15.09
SSIM	14.43	12.11
MS-SSIM	15.09	**10.2**
VIF	13.03	11.89

It is clear from Table 12-14 that MS-SSIM delivers the best performance among the four 2D QA algorithms when embedded in the cyclopean model. Fig. 30 breaks down the performance of the cyclopean model using MS-SSIM. Clearly, the QA performance is improved on blur and JP2K as might be expected, since strong binocular rivalry exists in asymmetric blur and JP2K distorted stereo images. The improvement in QA performance for FF distorted images is also significant for similar reasons. For stereo images distorted by white noise, there is no significant difference between the performance of averaged 2D QA and the "cyclopean" mode since binocular rivalry does not occur in white noise distorted stereo images [76]. For JPEG compression distorted stereo images, the performance numbers of the averaged 2D QA and the "cyclopean" model are very close. These results strongly suggest that binocular rivalry is an important ingredient in subjective stereoscopic QA, and our "cyclopean" framework successfully captures and

utilizes binocular rivalry to predict subjective 3D quality. Fig. 31 plots the predicted quality scores using MS-SSIM (after logistic regression) versus DMOS. Predicted scores from the proposed cyclopean framework are shown on top-left, while the bottom-left plot shows the scores from the 2D baseline. Clearly, the predicted scores attained using the cyclopean framework are better than the scores predicted by the 2D baseline. Moreover, the prediction errors which are measured by Root Mean Square Error (RMSE) of the cyclopean framework are lower than the prediction errors of the 2D baseline.

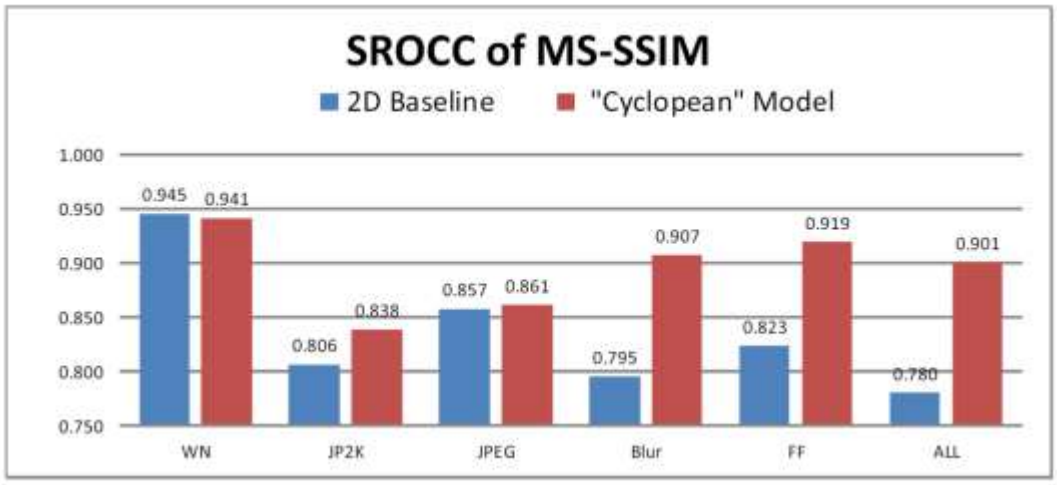

Fig. 30 SROCC values using MS-SSIM, broken down by distortion type.

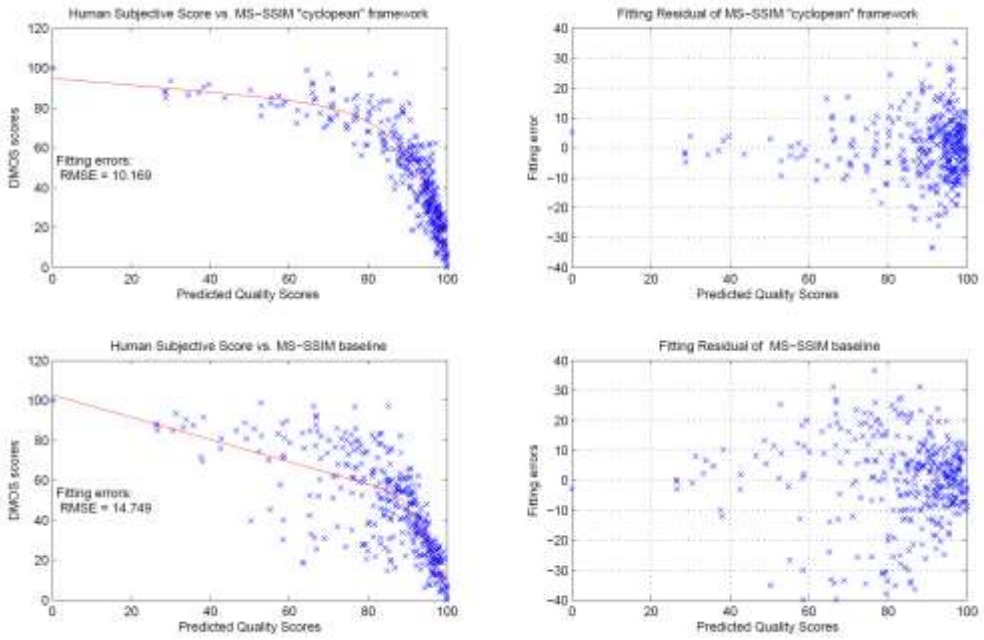

Fig. 31 Plot of predicted objective scores versus DMOS and prediction errors. Top Left: Prediction by MS-SSIM cyclopean framework. Top Right: Prediction errors of MS-SSIM cyclopean framework. Bottom Left: Predictions by MS-SSIM 2D baseline. Bottom Right: Prediction errors of MS-SSIM 2D baseline

To obtain deeper insights into how the performance of the cyclopean 3D QA model is improved by accounting for binocular rivalry, its performance on the separated symmetric and asymmetric distorted stereopairs is reported in Table 15, Table 16 and Table 17. The performance numbers in these three tables indicate that the cyclopean model did not boost performance on symmetric distorted stereoscopic images. However,

performance was greatly enhanced on the asymmetric distorted stereopairs. Furthermore, these performance numbers indicate that the task of predicting the quality of asymmetric distorted stereopairs is more difficult than that of predicting the quality of symmetric distorted data.

Table 15 SROCC scores relative to human subjective scores. Obtained using averaged left-right QA scores (2D Baseline) and the Cyclopean model on symmetric and asymmetric distorted stereopairs

	Symmetric		**Asymmetric**	
	2D Baseline	Cyclopean Model	2D Baseline	Cyclopean Model
PSNR	0.781	0.819	0.596	0.698
SSIM	0.826	0.85	0.742	0.827
MS-SSIM	0.912	**0.929**	0.687	**0.854**
VIF	0.916	0.902	0.737	0.804

Table 16 LCC scores relative to human subjective scores. Obtained using averaged left-right QA scores (2D Baseline) and the Cyclopean model on symmetric and asymmetric distorted stereopairs

	Symmetric		**Asymmetric**	
	2D Baseline	Cyclopean Model	2D Baseline	Cyclopean Model
PSNR	0.791	0.825	0.625	0.737
SSIM	0.845	0.882	0.767	0.850
MS-SSIM	0.924	**0.937**	0.709	**0.879**
VIF	0.924	0.906	0.772	0.822

Table 17 Fitting errors measured by RMSE. Obtained using averaged left-right QA scores (2D Baseline) and the Cyclopean model on symmetric and asymmetric distorted stereopairs

	Symmetric		**Asymmetric**	
	2D Baseline	Cyclopean Model	2D Baseline	Cyclopean Model
PSNR	16.42	15.15	16.83	14.58
SSIM	14.35	12.65	13.85	11.37
MS-SSIM	10.23	**9.37**	15.20	**10.29**
VIF	10.23	11.35	13.69	12.27

5.4.2.2 Influence of stereo matching algorithms

The preceding discussion describing the stereoscopic cyclopean QA model assumed that highly accurate ground truth depth values are available. Next, we study stereoscopic QA performance when estimated depth is used as computed by stereo algorithms.

Currently, stereo matching algorithms are generally tested on undistorted stereo images and compared using a simple measure (bad-pixel rate) [70]. However, we believe that such metrics provide little or no information regarding perceived 3D image quality.

Indeed, there have been no studies conducted to determine the degree to which the quality of an estimated disparity map is correlated with subjective judgements of depth. It is likewise unclear whether distortions of stereopairs affects perceived depth quality [10, 27]

The bad-pixel rates of the three selected stereo algorithms against ground truth are reported in Table 18. Clearly, all perform equally poorly when applied to distorted images. This lack of robustness is not unexpected owing to the ill-posedness of the stereo problem, and since none of these (or any other) stereo algorithms has been designed to excel in the presence of distortions. In addition, the ground truth maps that we used were obtained using a high-resolution laser range scanner. The ground truth maps have relatively fine disparity resolution over both smooth and depth-textured regions.

Table 18 Mean bad pixel rate value on 360 distorted stereopairs with standard deviation (inside the bracket) for three stereo algorithms.

	SAD	SSIM	Klaus
Bad-pixel rate	79.8% (9.24)	79.52 % (10.7)	78.04% (11.83)

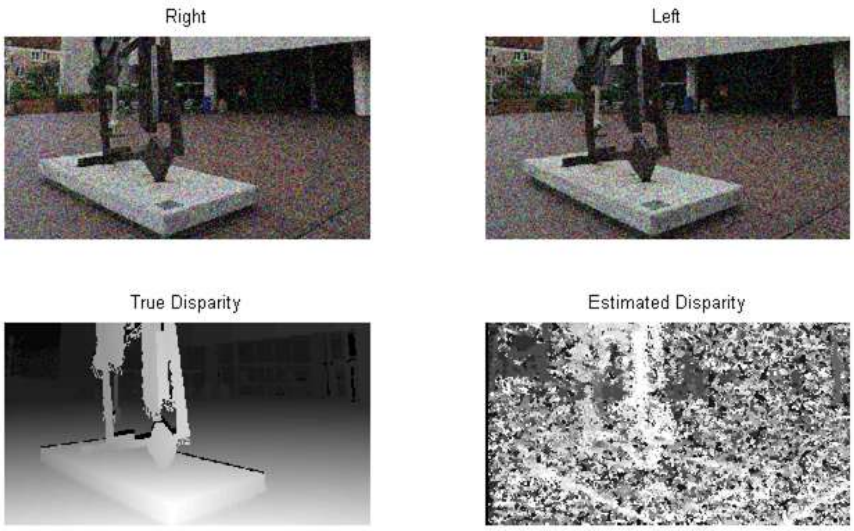

Fig. 32 Depth estimation using SSIM-based stereo algorithm on noised distorted stereo pairs. Free-fuse the noisy stereo image to see a 3D image.

Next, we discuss the influence of poor disparity estimation performance on 3D stereo QA. The performance of the cyclopean model using ground truth disparity, estimated disparity, and no disparity information are reported in Table 19. The table shows that there is no significant difference in the performance attained using the ground truth and estimated disparities, although the performance of the very simple SAD-based stereo algorithm is slightly lower than the other two stereo algorithms. All three significantly outperform the no-disparity case indicating that estimated disparities provide useful information when predicting the quality of the stereo 3D images. Based on

the results, it is clear that we should not use the bad-pixel rate to evaluate stereo algorithms when the objective is to design a 3D quality assessment algorithm. An explanation for this high bad-pixel rate, but decent results in performing a 3D QA task is that the depth signal in human vision has much lower bandwidths than the luminance spatial channel [77-79]. Therefore, we only need a low-resolution disparity map for the task of 3D quality assessment.

Table 19 SROCC, LCC, RMSE relative to human subjective scores attained by cyclopean model using different disparity maps.

Stereo Algorithm	**SROCC**	**LCC**	**RMSE**
Ground Truth	0.901	0.907	10.2
SAD	0.876	0.885	11.29
SSIM	0.893	0.901	10.58
Klaus	0.890	0.896	10.80
No depth information	0.817	0.824	13.73

5.4.2.3 Comparison with existing 3D QA models

Gorley and Holliman [41] proposed a PSNR-based 3D stereo QA model that does not include depth. Benoit, *et al.*[44] proposed a SSIM-based stereo QA model operating on both stereopairs and disparity maps. You, *et al.*[46] applied a variety of 2D QA models on stereopairs and disparity map and tried a number of ways to combine the predicted quality scores from stereopairs and disparity maps into predicted quality scores. Their best result is also SSIM-based. Hewage and Martini [52] proposed a PSNR-based reduced-reference stereo quality model utilizing disparity. In our simulations, since some of these algorithms require estimated disparity maps from both reference and test stereopairs, we used the SSIM-based stereo algorithm to create disparity maps.

Table 20 SROCC, LCC, and RMSE relative to human subjective scores attained by 3D QA models using SSIM-based stereo algorithm.

Algorithm	SROCC	LCC	RMSE
Cyclopean MS-SSIM	**0.893**	**0.901**	**10.58**
2D Baseline MS-SSIM	0.780	0.784	15.09
Benoit	0.728	0.745	16.2
You	0.784	0.797	14.66
Hewage	0.496	0.55	20.29
Gorley	0.158	0.511	20.88

Table 20 shows the performances of these 3D QA algorithms as compared with the "cyclopean" model. The "cyclopean" model using MS-SSIM delivers the highest performance, followed by the model proposed by You *et al.* which yields no significant difference relative to the performance of left-right averaged 2D QA using MS-SSIM. The performances of the other three algorithms are lower than this 2D baseline. This is another powerful demonstration of the importance of accounting for binocular rivalry when conducting stereoscopic QA.

5.5 Conclusion

We presented a new framework for conducting automatic objective 3D QA that delivers highly competitive performance, with a clear advantage when left-right distortion asymmetries are present. The design of the framework is motivated by studies on the perception of distorted stereoscopic images, and recent theories of binocular rivalry. The cyclopean 3D QA model that we derived was tested on the Phase II of LIVE 3D Image Quality Database, and found to significantly outperform conventional 2D QA models and well-known 3D QA models. The impact of the stereo algorithm used to conduct 3D QA was also discussed. We also found that a low-complexity SSIM-based stereo algorithm performs quite well for estimating disparity in the "cyclopean" algorithm in the sense that a high level of 3D QA performance is maintained.

An important contribution of this work is the demonstration that accounting for binocular rivalry can greatly improve the performance of 3D QA models. Indeed, most of the advantage conveyed by the cyclopean model was observed on asymmetric distorted stereopairs. The framework can, therefore, ostensibly be used to evaluate the quality of stereo content that has been compressed using a mixed resolution coding technique [76,

80]. Compressed stereo content that is transmitted over the wireless Internet may be subjected to other asymmetric distortions as well.

To further advance the performance of current 3D QA models, we think that the effect of depth masking and depth quality needs to be further studied and addressed. Regarding depth masking, our prior work [50] revealed no depth masking effect when viewing distorted stereopairs. However, we do not regard the results of our prior study to be universal and there remain other distortions to be studied. Furthermore, while we did not find depth masking of distortions, we did find evidence of facilitation which may prove relevant to 3D QA.

Regarding the role of computed disparity, prior models utilizing disparity maps derived from reference and test stereopairs have generally failed to deliver better QA performance than 2D QA models on the individual stereo images. Of course, the disparity cue is not the only one used by the human visual system to perceive depth. For example, monocular cues such as occlusion, relative size, texture gradient, perspective distortion, lighting, shading, and motion parallax [13] all affect the perception of depth. It is not yet clear how the brain integrates all these cues to produce an overall sensation of depth [19].

The influence of distortions on perceived depth quality also remains an open question. While Seuntiens, *et al.* [27] claimed that JPEG encoding has no effect on perceived depth, other recent research suggested that perceived depth quality is affected by both blur and white noise distortion, although the influence of distortion on perceived depth is less than the influence on perceived image quality [81]. Another recent study showed that, when viewing stereoscopic videos compressed by an H.264/AVC encoder using a range of QP values, perceived depth quality remained constant for some subjects, but varied with perceived image quality for others [82]. Subject agreement on perceived depth quality was much lower than on perceived image quality. Clearly, more research is merited on how perceived depth quality is affected by different distortion types, and on what kinds of depth cues are most strongly correlated with the reduced quality of depth perception when viewing distorted stereoparis.

CHAPTER 6 NO-REFERENCE STEREOSCOPIC QUALITY ASSESSMENT

6.1 Introduction

As mentioned in Chapter 2, the field of 3D QA has seen lesser research activity than that of 2D QA. Amongst these 3D QA models, there exist only a few reduced reference (RR) and no-reference (NR) models. Further, none of these models demonstrated performance competitive with FR 2D or 3D models, or with human perception of 3D quality.

This chapter describes the design of a NR 3D QA model that outperforms all 2D FR QA models and most 3D FR QA models in predicting the perceived 3D quality of natural stereo images. This NR 3D IQA model deploys 2D and 3D features extracted from stereopairs to assess the perceptual quality. Both symmetric- and asymmetric-distorted stereopairs are handled by accounting for binocular rivalry using a classic linear rivalry model. The extracted natural scene statistics (NSS) features are used to train a support vector machine to predict the quality of a test stereopair. I verified the performance of the proposed NR IQA model on the LIVE 3D Image Quality Database,

which includes both symmetric- and asymmetric-distorted stereoscopic 3D images. The experimental results show that the proposed model significantly outperforms conventional 2D full-reference QA algorithms applied to stereopairs, as well as 3D full-reference IQA algorithms on asymmetrically distorted stereopairs.

6.2 The proposed NR 3D IQA model

A flowchart of the proposed model is shown in Fig. 33. Given a stereo imagepair, an estimated disparity map is generated by a SSIM-based stereo algorithm, while a set of Gabor filter responses are generated on the stereo images using a filterbank. A "cyclopean" image is then synthesized from the stereo image pair, the estimated disparity map, and the Gabor filter responses. 2D features are then extracted from the synthesized "cyclopean" image, while 3D features are independently extracted from the estimated disparity map and an uncertainty map that is also produced by the stereo matching algorithm. Finally, the extracted 2D and 3D features are fed into a quality estimation module which predicts the perceived 3D quality of each tested stereo imagepair.

The linear model proposed by Levelt [9] is used to synthesize the "cyclopean" image from a stereo image pair. First, a disparity map is estimated from a test stereo pair

using a very simple SSIM-based stereo matching algorithm. The algorithm operates by search for disparities yield the best SSIM match between left and right image patches, where ties are broken by selecting the lower disparity solution. This estimated disparity map is then used to create a disparity-compensated right view image. The Gabor filter responses are then extracted from the left view image and the disparity-compensated right view image. Finally, a "cyclopean" image is synthesized from the left and disparity-compensated right views and their Gabor filter responses. The details of this process can be found in Chapter 5. Since the contribution of this chapter is the method of selecting and extracting features and the way they are used to perform NR 3D QA, we only focus on explaining these later parts.

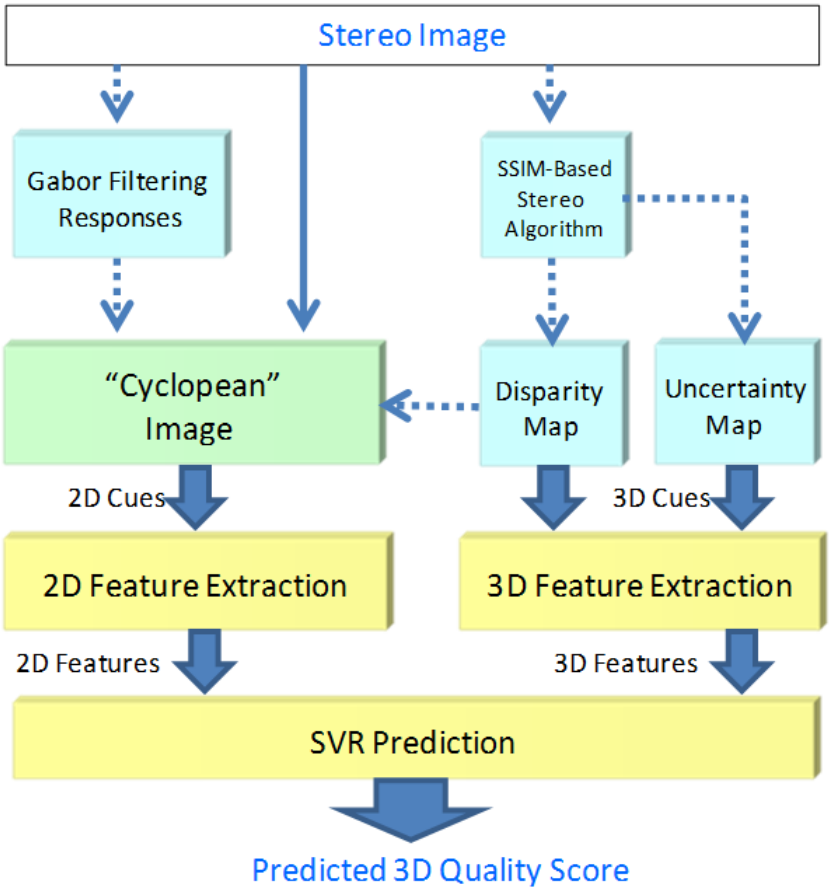

Fig. 33 The flowchart of the proposed 3D NR QA model

6.2.1 2D feature extraction

Research on natural scene statistics (NSS) has demonstrated that images of natural scenes belong to a small set of the space of all possible signals and that they obey predictable statistical laws [72] . Successful 2D NR QA algorithms [36-38, 83] based on the statistics of natural scenes (and the fact that human perception has adapted to these

statistics over the eons) have achieved comparable QA prediction performance as high performance FR QA models [30, 33]. Although images of real-world scenes may vary greatly in their luminance and color distributions, by pre-processing images in biologically relevant way, e.g., by processes of predictive coding [84] and divisive normalization [85] , yields transformed images obeying a regular parametric statistical model [86, 87]. Ruderman [87] showed that images processed via a simple local mean subtraction and divisive variance normalization produces nearly decorrelated luminance obeying a Gaussian-like distribution. This model closely mimics the classicalcenter surround model with adaptive gain control. Using these kind of NSS features, Mittal, *et al.* [36] developed a highly competitive 2D NR IQA model called BRISQUE.

We apply similar pre-processing on the synthesized Cyclopean image:

$$M(i,j) = \frac{(I(i,j) - \mu(i,j)}{\sigma(i,j) + C} \tag{7}$$

where *i, j* are spatial indices, μ and σ are the local sample mean and weighted standard deviation computed by a local window, and *C* is a constant that ensures stability. In our implementation, we use a 11x11 Gaussian weighting matrix to compute μ and σ and set C = 0.01.

Following [36] , we model the coefficients (7) of the possibly distorted cyclopean image as following a generalized Gaussian distribution (GGD):

$$f_x(x;\mu,\sigma^2,\gamma) = ae^{-[b|x-\mu|]^\gamma} \tag{8}$$

where μ , σ^2 and γ are the mean, variance, and shape-parameter of the distribution,

$$a = \frac{b\gamma}{2\Gamma(1/\gamma)} \tag{9}$$

$$b = \frac{1}{\sigma}\sqrt{\frac{\Gamma(3/\gamma)}{\Gamma(1/\gamma)}} \tag{10}$$

and $\Gamma(\cdot)$ is the gamma function:

$$\Gamma(x) = \int_0^\infty t^{x-1}e^{-t}dt,\ x > 0. \tag{11}$$

The parameters (σ and γ) are estimated using the method used in [88]. The skewness and the kurtosis of these coefficients are also estimated method.

6.2.2 3D feature extraction

Features based on natural scene statistics have been shown to be effective, robust tools for predicting the quality of natural images. The success of NSS-based features is built on the fact that pristine natural scenes trend to follow certain regular statistical laws. Following this philosophy when modeling 3D images, we also build into our QA model features derived from 3D natural scene statistic models. As compared to the extensive body of literature on 2D natural luminance statistics, studies on the statistics of disparity

and depth have been quite limited. One possible explanation for this dearth is that acquiring accurate disparity data is a much more difficult task than capturing 2D imaging data. Further, there is no known method to explicitly model the 3D experience of viewing a natural 3D scene.

There has also been only a small amount of work on modeling of natural 3D statistics. Huang [89] first studied the statistics of range images, using range data measured by a laser range-finder. They begin with the assumption that natural range maps follow the random collage model, i.e. that a range image can be partitioned into disjoint smooth surfaces separated by discontinuities. Yang and Purves [90] further studied the statistics of range data, which were acquired by a laser scanner, and found that their range data is quite rough and is anisotropic in nature. If a viewing model is defined, the range data can be transformed into a disparity data. To study the statistics of disparity, Hibbard [91] and Liu, *et al.* [92] model the fixation distance of a virtual subject. An essential difference between these two approaches is that ground truth range data (measured by a laser scanner) is analyzed in Liu's work, while Hibbard used a random collage sphere model to synthesize range data. Both groups found that the

distribution of disparity follows a Laplacian shape. In the following, we discuss how to incorporate 2D and 3D NSS models in the 3D QA problem.

To conduct the task of no-reference quality assessment on a stereo image pair, it is assumed that only the stereopair is available, without any reference data including ground truth disparity. Thus, the only accessible 3D feature is estimated disparity from a stereo matching algorithm. Here, we use a simple SSIM-based stereo matching algorithm to estimate a disparity map. Therefore, it is worth discussing the difference between ground truth disparity and estimated disparity. Fig. 34 shows a stereopair with ground truth disparity and estimated disparity map. This stereopair was captured using a parallel-camera set-up with a laser scanner that captures ground truth range data. The ground truth disparity map is directly converted from the range data since the capture model is known. In Fig. 34, one can clearly see that there are many estimated errors, especially towards the bottom sections of the image. The errors are produced by the complex, repetitive texture of the sidewalk, which the simple low-complexity stereo-algorithm doesn't handle well. Moreover, the slanting foreground surface plane is smoothly captured in the ground truth disparity map while the estimated map shows a ladder-like appearance map due to the integer pixel precision of the stereo algorithm. Fig.

35 depicts the histogram of a ground truth and an estimated disparity map both are before after local mean removal and divisive normalization as in (7). The top left and top right of Fig. 35 suggest that no known model distribution could be used to consistently fit them. However, following the normalization process, the ground truth disparity distribution takes a Gaussian-like shape, while the estimated disparity distribution is much more peaky and heavily tailed. Both are zero-mean symmetric distribution and can be modelled as following a GGD. As before, we take the following as features: GGD parameters, standard deviation, skewness, and kurtosis.

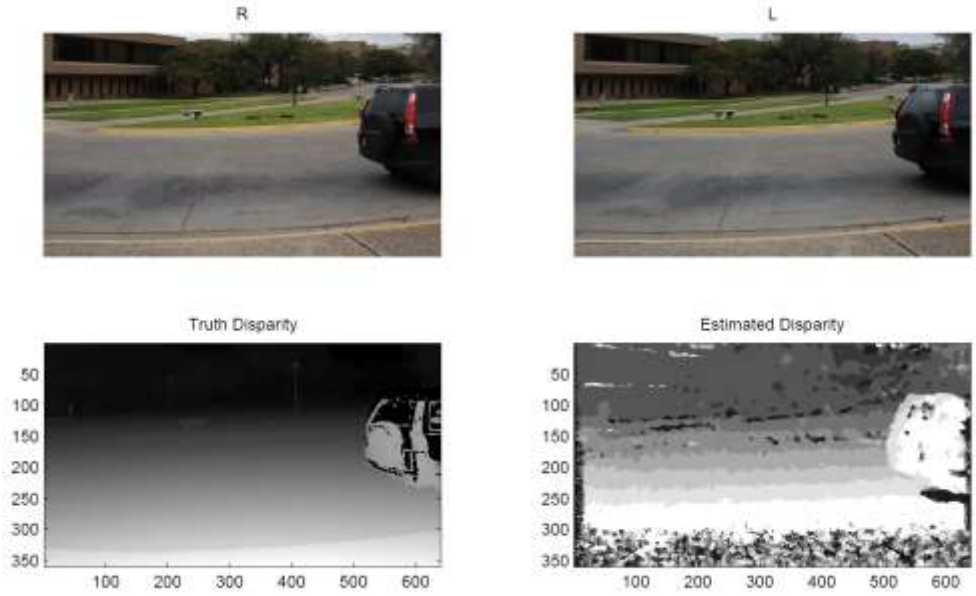

Fig. 34. A stereopair with ground truth disparity and estimated disparity. Top left: Right view of the stereo image. Top right: Left view of the stereo image.

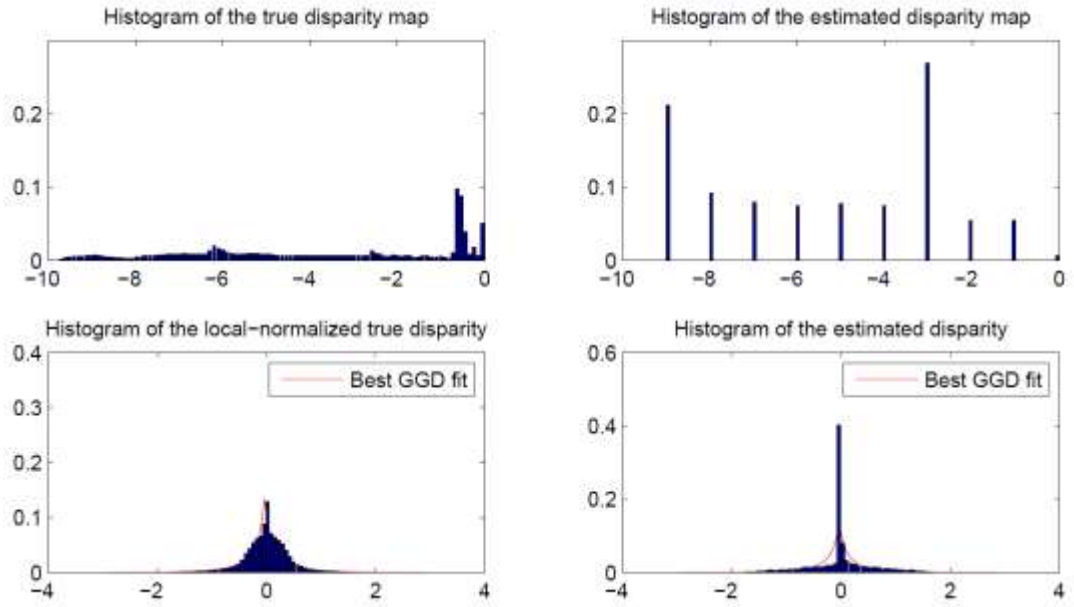

Fig. 35. Top left: Histogram of ground truth disparity map. Top right: Histogram of the estimated disparity map. Bottom left: Histogram of the local-normalized ground truth disparity map GGD fit overlaid. Bottom right: Histogram of the estimated disparity map with GGD fit overlaid. Bottom left: Ground truth disparity. Bottom right: Estimated disparity.

These 3D NSS features can be effectively used in the process of distinguishing a pristine stereo image pair from a distorted version of it. We used five common distortion types to impair the stereo image data: white noise, blur, JPEG compression, JPEG2000 (JP2K) compression, and a Fast-Fading (FF) model based on the Rayleigh fading channel. Fig. 36 shows histograms of the estimated disparities of stereo images distorted by these models. It shows clear differences in the shape (kurtosis) and spread. The

disparity estimated from the pristine stereopair has the highest kurtosis, while JP2K distortion produces the most Gaussian-like distribution.

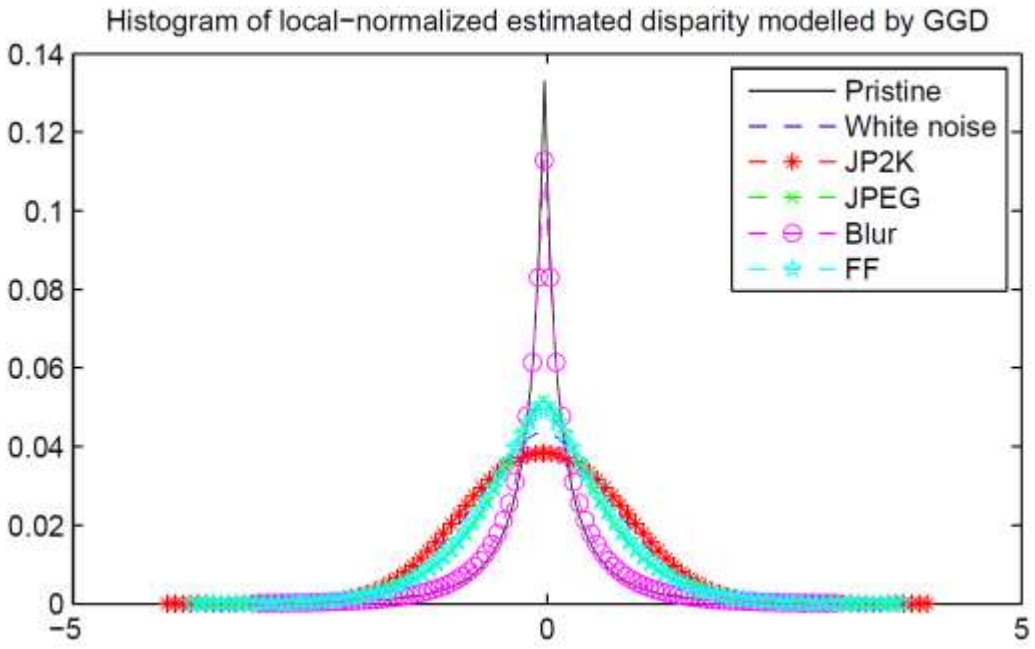

Fig. 36. Disparity distributions of a distorted stereopair.

Other than the estimated disparity, the uncertainty produced by the SSIM-based stereo matching algorithm is a useful feature for the task of 3D NR QA. The uncertainty is defined as

$$Uncertainty(l,r) = 1 - \frac{(2\mu_l\mu_r)(2\sigma_{lr} + C_2)}{(\mu_l^2 + \mu_r^2 + C_1)(\sigma_l^2 + \sigma_r^2 + C_2)} \tag{12}$$

where l is the left-view image and r is the disparity-compensated right-view image of a stereopair. The uncertainty reflects the degree of similarity (or lack thereof) between the

corresponding pixels of a stereopair. We have observed that the histograms of noise-free natural stereopairs captured using a paralleled-camera setting present a very positive skew distribution. This may be understood by observing that the stereo-matching algorithm generally finds good matches (low-uncertainty) at most places, while relatively rare occluded or ambiguous flat or textured areas may cause sparse errors in the results of the stereo matching algorithm (high-uncertainty), contributing weight to the tail of the uncertainty distribution. Fig. 37 demonstrates this observation. The bottom right plate of Fig. 37 shows that most regions of the image have a low uncertainty, while higher uncertainty values are observed around the sky and trees. To model this observation, we fit a log-normal distribution to the histogram of the uncertainty map. The probability density function of a log-normal distribution is defined as

$$f_X(x;\mu,\sigma) = \frac{1}{x\sigma\sqrt{2\pi}} \exp -\frac{(\ln x - \mu)^2}{2\sigma^2} \tag{13}$$

where μ is the location parameter and σ is the scale parameter. A maximum likelihood method is used to estimated μ and σ for a given histogram of uncertainties.

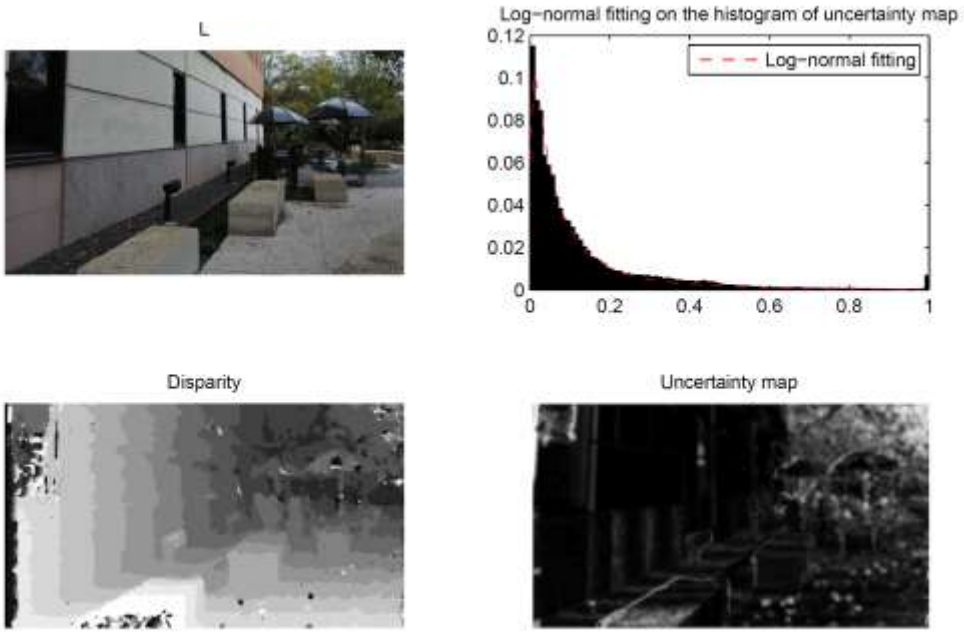

Fig. 37. Top left: Left view of a stereopair. Top right: Histogram of the uncertainty map and the best log-normal fit. Bottom left: The estimated disparity map. Bottom right: The uncertainty map produced by the stereo matching algorithm.

The histogram of uncertainty also varies when a stereopair is distorted. Fig. 38 shows the uncertainty distribution of stereopairs distorted by white noise, blur, FF, JPEG compression and JP2K compression. As depicted, in Fig. 38, the uncertainty distribution predictably changes with distortion type. For example, since a Gaussian blur distortion suppresses details in the stereopair, the uncertainties in the disparity estimation are reduced, yielding a more peaky distribution of uncertainties. White noise, JPEG,

JPEG2K, and FF distortion increase the uncertainty of stereo matching and reduce the peaky-ness of the uncertainty distribution.

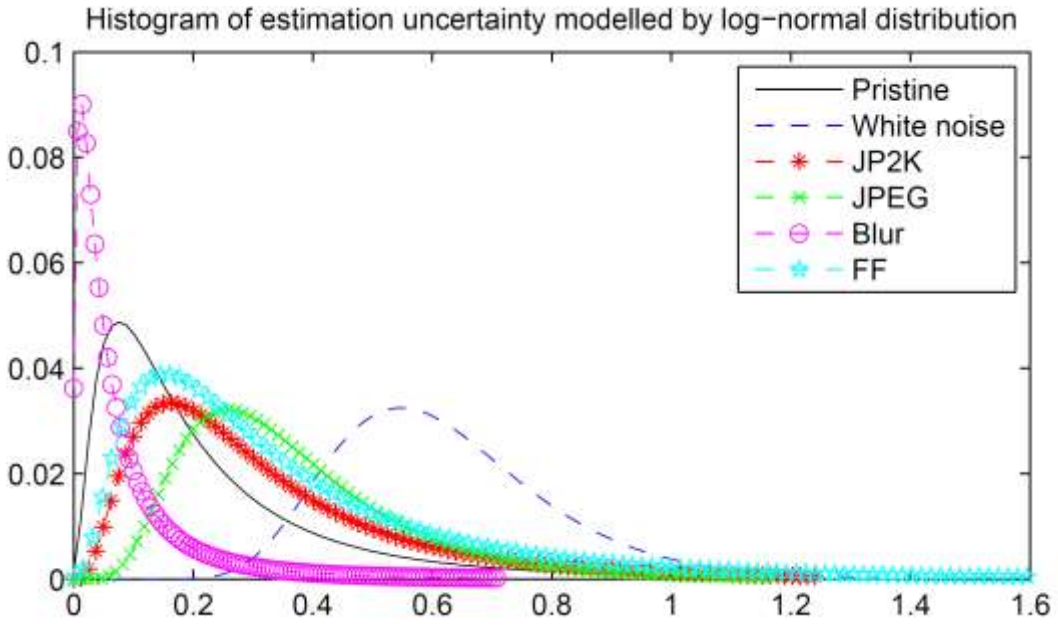

Fig. 38. Plot of modelled uncertainty distributions of distorted stereopair.

To summarize, the 3D features used for 3D NR QA prediction are the GGD fit parameters (μ, σ), the standard deviation, skewness, and kurtosis of the local-normalized estimated disparity map, and the best-fit log-normal parameters (μ, σ), skewness, and kurtosis of the uncertainty map.

6.2.3 Quality estimation

A two-stage QA framework is used to predict the quality of a test stereopair. This follows the framework introduced in [34] and elaborated in the 2D IQA DIIVINE index [38] . In their model, a probabilistic support vector classifier is applied first to decide the most likely distortion type afflicting the stereopair. A support vector regressor (SVR) is then used to assess the perceptual distortion severity. However, unlike DIIVINE, here, the classifier in our 3D NR IQA model is designed to decide whether a stereo pair is symmetrically or asymmetrically distorted without predicting the distortion type. This is important since asymmetrically distorted stereopairs may create binocularly rivalrous 3D experiences, and may yield different extracted 3D features than symmetrically distorted stereopairs. In the human study on distorted stereopairs that we conducted [50], we found that the perceived quality of a asymmetrically distorted stereopair is not accurately predicted by the simple average quality of the stereo views, although the quality of symmetrically distorted stereopairs might be accurately predicted in this manner. The same feature vector is used for classification and regression. After the classification process is complete, the predicted quality score is computed as the dot product of the distortion probability vector and the vector of symmetric/asymmetric quality scores.

6.3 Results

We utilized the LIVE 3D Image Quality Database to verify the performance of our proposed 3D NR IQA model. Although part of this database is publicly available [8] (Phase I consisting of symmetric distortions), a second phase has only recently been created.

6.3.1 LIVE 3D Image Quality Database

This database was constructed in two phases. Phase I contains symmetrically distorted stimuli while phase II has both symmetrically and asymmetrically distorted stimuli. Thus, phase I and phase II are actually different and complementary datasets. Phase I [8] has 20 pristine stereopairs and 365 distorted stereopairs, while phase II has 8 pristine sterepairs and 360 distorted stereopairs. The details of the dataset and of the human studies that were conducted on them to subjectively annotate the stimuli are described in the following.

6.3.1.1 Source Images

The pristine stereo images used in both phases are stereo images co-registered with range data measured by a high-performance range scanner (RIEGL VZ-400) obtained by a Nikon D700 digital camera. The stereo image pairs were shot using a 65

mm baseline. The sizes of the images are 640 x 360 pixels. Fig. 39 shows a stereopair and its associated a ground truth depth map. For further details on the data acquisition, see [8].

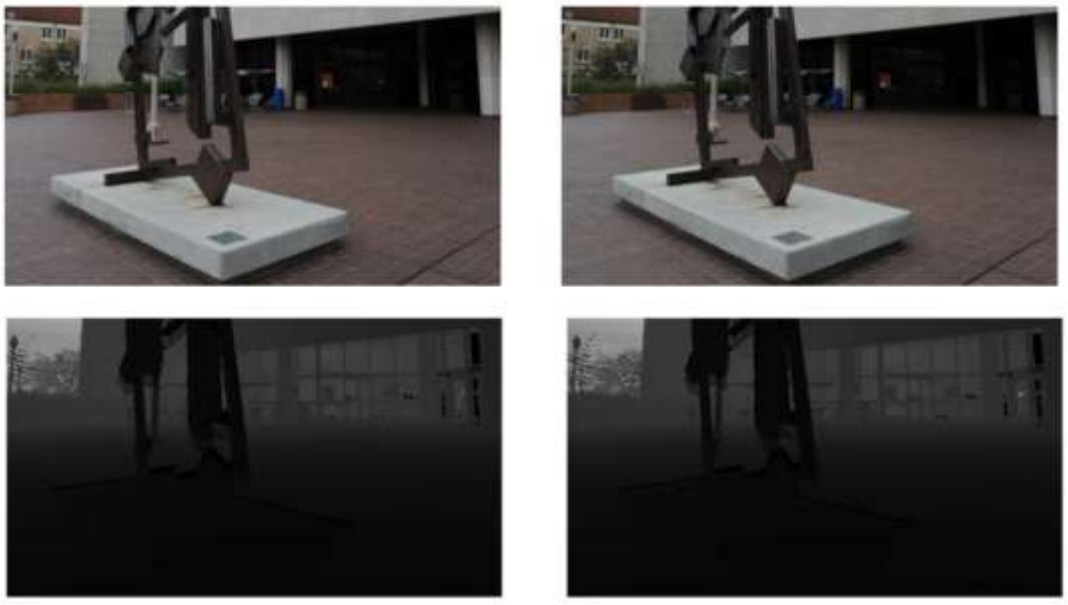

Fig. 39. A stereo image (free-fuse the left and right images) and ground truth disparity maps.

6.3.1.2 Participants

In both phases, each subject reported normal or corrected normal vision and no acuity or color test was deemed necessary. However, a Randot stereo test was used to pre-screen participants for normal stereo vision in phase II. Phase I utilized thirty-two participators with a male-majority population. In phase II, six females and twenty-seven males participated in the experiment, aged between 22 and 42 years.

6.3.1.3 Display Setting

Phase I was conducted with a iZ3D 22" 3D monitor with passive polarized 3D glasses, while phase II was conducted using a Panasonic 58" 3D TV with active shutter glasses. The viewing distance was four times the screen height in both cases.

6.3.1.4 Stimuli

Both phases used five types of distortions: compression using the JPEG and JPEG2000 compression standards, additive white Gaussian noise, Gaussian blur and a fast-fading model based on the Rayleigh fading channel. The degradation of stimuli was varied by controlling parameters within a pre-defined range, as reported in Table 21. The ranges of control parameters were decided beforehand to ensure that the distortions varied from almost invisible to severely distorted with a good overall perceptual separation between distortion levels throughout. Due to the different viewing environments, the ranges of distortions are also different in the two experimental phases.

The phase I dataset contains only symmetrically distorted stereo images (80 each for JP2K, JPEG, WN, and FF; 45 for Blur) while the phase II dataset has both symmetrically and asymmetrically distorted stereo images (72 images for each distortion type). A ‘symmetrically’ distorted stereopair implies that the same ‘amount’ of distortion was added to the left and right image, while the ‘asymmetrically’ distorted stereopair has

a different ‘amount’ of distortion in the two views. In the phase II dataset, for each distortion type, every reference stereopair was distorted to create three symmetric distorted stereopairs and six asymmetric distorted stereopairs.

Table 21 Range of parameter values for distortion simulation

Distortion	Control Parameter	Phase I Range	Phase II Range
WN	Variance of Gaussian	[0.01 0.15]	[0.001 0.5]
Blur	Variance of Gaussian	[0.01 20]	[0.5 30]
JP2K	Bit-rate	[0.05 3.15]	[0.04 0.5]
JPEG	Quality parameter	[10 50]	[8 50]
FF	Channel signal-to-noise ratio	[12 20]	[15 30]

6.3.1.5 Procedure

A single stimulus continuous quality scale (SSCQS) [55] study with hidden reference was conducted in both phases. Both studies used continuous scales labelled by equally spaced adjective terms: bad, poor, fair, good, and excellent, i.e. a Liekart scale. Both studies were divided into 2 sessions; each of less than 30 minutes to minimize subject fatigue. A training session was also conducted before the beginning of each study to help familiarize participants with the GUI.

6.3.1.6 Subjective Quality Scores

Difference opinion scores (DOS) were obtaining by subtracting the ratings that the subject gave each reference stimuli from the ratings that the subjective gave to the corresponding test distorted stimuli. The remaining subjective scores were then normalized to Z-scores, and averaged across subjects to produce difference mean opinion scores (DMOS). Fig. 40 shows the distribution of DMOS of the database. The DMOS distributions of phase I and phase II are quite different. In the phase I dataset, the DMOS given to WN and FF distorted stimuli varied from -10 to 60, the DMOS given to JP2K and Blur distorted stimuli have a range between -10 and 40, and the DMOS given to JPEG have a significantly narrower range from -10 to 20 indicating less perceptual distortion overall and smaller differences in perceived severity. Similarly, JP2K and Blur was generally less visible than WN and FF in the phase I dataset. However, in the phase II dataset, only the JPEG distorted stimuli were less visible than other distortion types, while WN, JP2K, Blur, and FF distortions were generally within a similar quality range. The DMOS scores of both symmetric and asymmetric stimuli are plotted in Fig. 41. From the plot, it is apparent that the different DMOS ranges were not caused by the symmetry (or lack of) of the distortion.

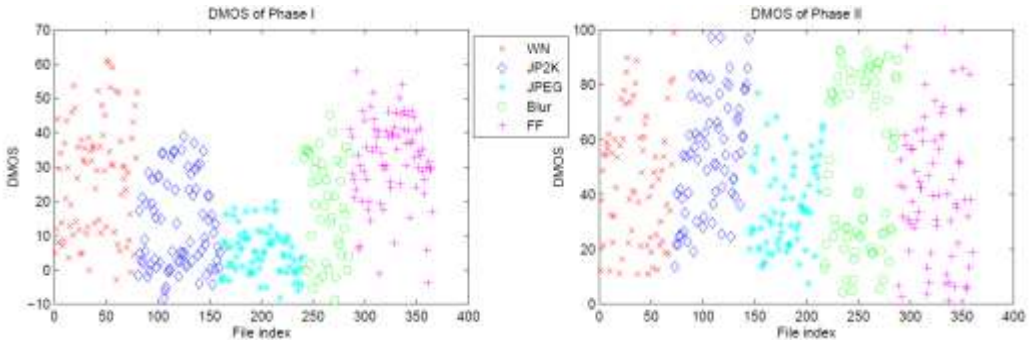

Fig. 40. Left: DMOS of LIVE 3D Image Quality Database Phase I. Right: DMOS of LIVE 3D Image Quality Database Phase II.

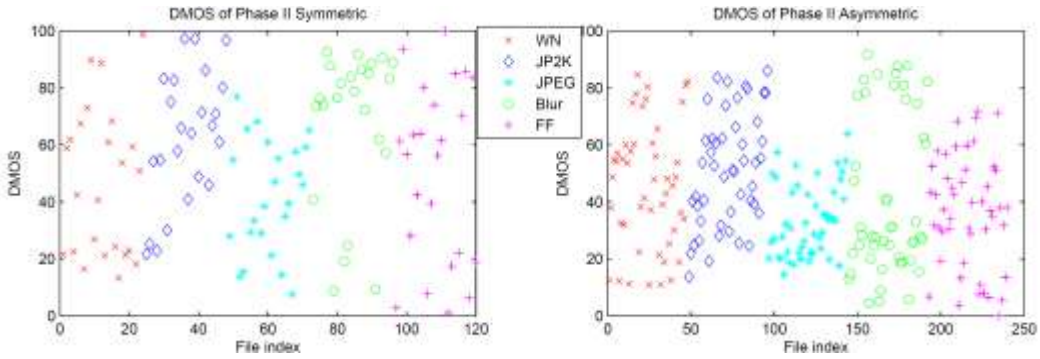

Fig. 41. Left: DMOS of Phase II symmetric distorted stimuli. Right: DMOS of Phase II asymmetric distorted stimuli

6.3.2 Classification accuracy

The first layer of our proposed 3D NR IQA model is the symmetric vs. asymmetric distortion classifier. We used the LIBSVM package [93] to perform classification. To assess the performance of the classifier, we performed 1000 iterations

of the train-test process. At each iteration, we randomly picked 80% of the dataset as training data and the remaining 20% to test. The mean classification was 82.07% with standard deviation 2.88.

6.3.3 Performance

6.3.3.1 Phase I dataset

Since the proposed algorithm requires training, 1000 iterations of the train-test process was used. At each iteration, the phase I dataset was randomly divided into 80% training and 20% test across 1000 iterations. The performance was measured by the Spearman's Rank Ordered Correlation Coefficient (SROCC) between the predicted scores and the DMOS. The median performance across the 1000 iterations is reported.

We compared the performance of our 3D NR IQA model with several 2D FR and NR IQA models: PSNR, SSIM [30], MS-SSIM [29], and BRISQUE [36]. SSIM and MS-SSIM are FR IQA algorithms, while BRISQUE is a high performance NR QA algorithm. For all 2D QA algorithms, the predicted quality of a stereopair is taken to be the average quality predicted from the left and right views. The SROCC numbers are shown in Table 22. Our proposed model performs as well as BRISQUE, but a little poorer than MS-SSIM. As we found in our prior human study, depth seems to have little influence on the perceived quality of distorted stereopairs, although other aspects of

binocular fusion, such as introduced rivalries, are important. Since there is little or no binocular rivalry from the distortion present in the stimuli in the Phase I dataset, we do not expect any significant improvement in performance under our 3D NR IQA model. In addition, the performance of our model is significantly lower on the JPEG distorted stimuli as compared to other distorted stimuli, but the same holds for all of the other QA algorithms. As shown in Fig. 40, the JPEG distorted stimuli represent a more difficult challenge because their qualities are less perceptually separated.

Table 22 Comparison of 2D IQA algorithms: SROCC against DMOS on the LIVE Phase I 3D IQA dataset. Italicized algorithms are NR IQA algorithms, all others are FR IQA algorithms.

	WN	**JP2K**	**JPEG**	**Blur**	**FF**	**All**
PSNR	0.932	0.799	0.121	0.902	0.587	0.834
SSIM	0.938	0.858	0.436	0.879	0.586	0.876
MS-SSIM	0.942	0.892	0.613	0.926	0.723	0.926
BRISQUE	0.940	0.812	0.569	0.860	0.784	0.901
Our Model	0.919	0.863	0.617	0.878	0.652	0.891

We also studied the relative performance of other 3D IQA algorithms. The FR 3D IQA algorithms compared include a PSNR-based 3D stereo IQA algorithm proposed by

Gorley and Holliman [41], a SSIM-based stereo IQA model proposed by Benoit, *et al.* [44], Cyclopean MS-SSIM [94] which considers the influence of binocular rivalry on perceived 3D quality, and a 3D QA algorithm proposed by You, *et al.*, who applied a variety of 2D FR IQA algorithms on steropairs and disparity maps to combine the predicted quality scores from stereopairs and disparity maps into predicted 3D quality score in a variety of ways. We report their best result (a SSIM-based algorithm) on our database. The RR 3D IQA algorithm proposed by Hewage, *et al.* [52] and the NR 3D IQA algorithm proposed by Akhter, *et al.* [53] are also included. We used a SSIM-based stereo-matching algorithm to generate disparity maps for these 3D IQA models. Their performances in terms of SROCC are reported in Table 23. This table shows that Cyclopean MS-SSIM has the best performance among all compared 3D IQA algorithms although its performance is not significantly different than the performance of 2D MS-SSIM. The results also show that our NR algorithm outperforms most of the 3D IQA algorithms, except for Cyclopean MS-SSIM and the FR models proposed by Benoit, *et al.* when dealing with symmetrically distorted stereopairs. The RR IQA model [52] performs slightly worse than 2D PSNR, while the NR IQA model proposed by Akhter, *et al* [53] performs significantly worse than 2D PSNR. As shown in the table, 2D MS-SSIM

showed the best performance on symmetrically distorted stereopairs, outperforming most 3D QA algorithms, except for Cyclopean MS-SSIM.

Table 23 Comparison of 3D IQA models: SROCC against DMOS of the Phase I dataset. Italicized algorithms are NR IQA algorithms, all others are RR or FR IQA algorithms.

	WN	**JP2K**	**JPEG**	**Blur**	**FF**	**All**
Benoit	0.930	0.910	0.603	0.931	0.699	0.899
You	0.940	0.860	0.439	0.882	0.588	0.878
Gorley	0.741	0.015	0.569	0.750	0.366	0.142
Cyclopean MS-SSIM	0.948	0.888	0.53	0.925	0.707	0.916
Hewage	0.940	0.856	0.500	0.690	0.545	0.814
Akhter	0.914	0.866	0.675	0.555	0.640	0.383
Our Model	0.919	0.863	0.617	0.878	0.652	0.891

6.3.3.2 Phase II dataset

Binocular rivalry is the main factor that affects the perceived 3D quality of asymmetrically distorted stereopairs. On the Phase II dataset, 1000 iterations of train-test process were again used. We report the median result of the 1000 runs. The same set of 2D and 3D IQA algorithms was tested on the phase II dataset. The results are reported in Table 24 and Table 25, respectively. As shown in Table 24, the performance of our model is significantly better than all of the 2D IQA models. Breaking down performance by distortion type, the improvement relative to different 2D QA models are observed for

all distortion types, except for WN. This observation is reasonable, since there is no binocular suppression observed in WN distorted stereopairs. The perceived quality of a WN distorted stereopair is about the average of the qualities of the left and right view.

Table 24 Comparison of 2D IQA algorithms: SROCC against DMOS on the LIVE Phase II 3D IQA dataset. Italicized algorithms are NR IQA algorithms, others are FR IQA algorithms.

	WN	**JP2K**	**JPEG**	**Blur**	**FF**	**All**
PSNR	0.924	0.619	0.525	0.686	0.719	0.672
SSIM	0.915	0.718	0.701	0.834	0.827	0.796
MS-SSIM	0.945	0.806	0.857	0.795	0.823	0.780
BRISQUE	0.940	0.773	0.569	0.865	0.931	0.770
Qur Model	0.933	**0.883**	0.883	**0.9**	0.9	**0.895**

Table 25 Comparison of 3D IQA algorithms: SROCC against DMOS on the LIVE Phase II 3D IQA dataset. Italicized algorithms are NR IQA algorithms, others are FR IQA algorithms.

	WN	**JP2K**	**JPEG**	**Blur**	**FF**	**All**
Benoit	0.919	0.755	0.861	0.459	0.762	0.728
You	0.902	0.888	0.777	0.808	0.888	0.785
Gorley	0.863	0.139	0.054	0.760	0.592	0.153
Cyclopean MS-SSIM	0.941	0.825	0.854	0.905	0.883	0.893
Hewage	0.874	0.581	0.705	0.014	0.674	0.496
Akhter	0.726	0.727	0.651	0.690	0.561	0.551
Our model	0.933	0.883	**0.883**	**0.9**	**0.9**	**0.895**

Table 25 shows the results against the mixed dataset of all 3D IQA algorithms. Our model delivers the best performance compared to most other models. The FR Cyclopean MS-SSIM yields an insignificant difference in performance. However, all of the others delivered significantly lower performance than these two models. Among individual distortion types, the proposed model performs either better than or at parity with the best for all distortion types. Compared with the other 3D NR IQA algorithms [53], our model performs significantly better on the entire dataset and for each distortion type.

We also studied the performance of the tested algorithms broken down by the way they are distorted (symmetrically or asymmetrically). Table 26 shows the performances of the 2D and 3D IQA algorithms. It is apparent that our model performs as well as 2D MS-SSIM, You's algorithm, and Cyclopean MS-SSIM on symmetrically distorted stereo 3D images. When dealing with asymmetrically distorted stereo 3D images, our model significantly outperforms all other 2D and 3D IQA algorithms, except Cyclopean MS-SSIM, which also models binocular rivalry.

Table 26 Break down of performance on symmetrically and asymmetrically distorted stimuli in the Phase II dataset. Italicized algorithms are NR IQA models, others are RR or FR IQA algorithms.

	Symmetric	**Asymmetric**
2D PSNR	0.781	0.596
2D SSIM	0.826	0.742
2D MS-SSIM	0.912	0.687
2D BRISQUE	0.850	0.690
Benoit	0.859	0.670
You	0.913	0.697
Gorley	0.382	0.047
Cyclopean MS-SSIM	0.930	0.854
Hewage	0.650	0.498
Akhter	0.426	0.526
Our model	0.923	0.848

Comparisons across the phase I and phase II datasets indicate that, the FR Cyclopean MS-SSIM model and our new 3D NR IQA model perform competitively on the symmetric dataset and outperform all other 2D and 3D IQA algorithms on the mixed dataset. The 3D FR QA model proposed by You, *et al.* performs as well as SSIM on both datasets. The SSIM-based 3D FR model proposed by Benoit performs as well as SSIM

on the symmetric dataset, but significantly worse than SSIM on the mixed dataset. The others 3D IQA models perform worse than PSNR on both datasets.

To further verify the performance of our model, we also report performance across datasets. Since only the phase II dataset included both symmetrically and asymmetrically distorted stereopairs, we trained our model on the phase II dataset and tested on the phase I dataset. The result is reported in Table 27. Across datasets, our model performs equally well on WN, JP2K, JPEG, and the blur distorted stereopairs, but the performance was lower on FF distorted stereopairs. The overall performance drops slightly due to the performance loss on the WN and FF distorted stimuli.

Table 27 Test across datasets: SROCC against DMOS of the Phase I dataset

	WN	JP2K	JPEG	Blur	FF	ALL
Train with Phase II dataset	0.826	0.849	0.626	0.882	0.423	0.865
1000 iterations on Phase I dataset	0.919	0.863	0.617	0.878	0.652	0.891

6.4 Conclusion

In this chapter, I proposed a no-reference stereoscopic 3D image quality assessment algorithm based on 2D and 3D natural scene statistics. The resulting algorithm utilizes statistical features previously proposed for 2D NR algorithms and binocular rivalry modelled by 3D FR IQA algorithms. When there is no binocular rivalry, our algorithm performs as well as the state-of-the-art 2D NR IQA algorithm. Compared with 3D IQA algorithms, our algorithm significantly outperforms 3D NR QA algorithms and delivers competitive performance relative to high performance 3D FR IQA algorithms.

In the future, this framework can be extended to predict the quality of depth-image-base-rendered (DIBR) 3D images. DIBR generated 3D images may have distortions caused by hole-filling algorithms, 3D warping algorithms, and errors from depth estimation. The challenge of IQA models for DIBR generated 3D images is not limited to visible distortions. Unnaturalness of synthesized 3D stereopairs may contribute to visual discomfort, which is more difficult to quantify than image quality. In the next chapter, I will discuss the perceived depth quality related to the 3D representation of distortion-free stereo images.

CHAPTER 7 QUALITY OF DEPTH

7.1 Introduction

Currently, most natural 3D images and videos are captured using a dual-camera configuration. Generally, there are two methodologies of camera settings: the parallel camera configuration and the toe-in camera configuration. Both have their own strengths and weaknesses, but the toe-in camera configuration requires more knowledge and effort in capturing stereo videos since the user needs to decide the vergence point in depth, which change during capture. Therefore, the parallel camera configuration is often used to capture stereo images or videos with consumer cameras [2] or smart phones [95]. However, post-processing is required to enable binocular fusion of the stereo content, and to avoid visual discomfort when viewing these images. Because the images captured by a parallel camera configuration only allows uncrossed disparity values, post-processing is needed to create crossed disparity values and to limit all disparity values to within a certain range. Fig. 42 illustrates that objects in front of the screen have disparities that are crossed, while objects behind the screen have uncrossed disparities. Fig. 42 also shows the parallel camera configuration. Since the focal plane is located at the point of infinity, only crossed disparities exist in stereo images captured in this way.

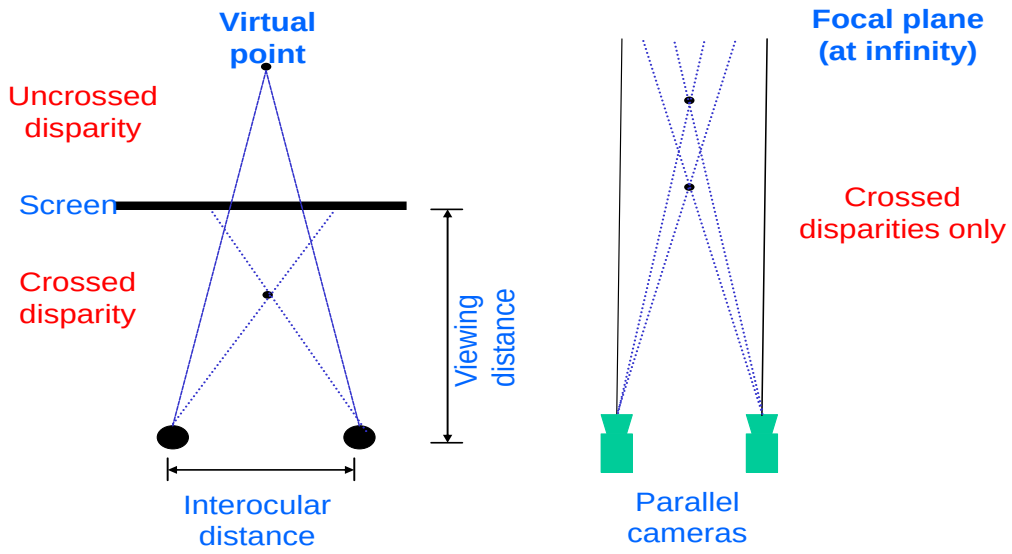

Fig. 42 Left : Illustration of crossed and uncrossed disparity. Right: The parallel camera configuration.

As mentioned in chapter 2, several factors may affect the quality of stereo viewing, but I focus on the visual discomfort caused by the vergence-accommodation conflict [96, 97]. The vergence accommodation conflict is illustrated in Fig. 43. In this figure, the observers' eyes focus on the screen, and the virtual point is the viewed object. Despite inconsistences between vergence and accommodation, human subjects can tolerate a certain degree of discrepancy and still see a fused stereo image instead of two overlapped images. To avoid seeing double images instead of a cyclopean image, the disparity values of a stereo image must fall within a certain range which depends on viewing distance, the depth of the object, the screen resolution, and individual factors.

Yeh and Silverstein [98] conducted a human study and claimed that human subjects can fuse stereo images if $|\alpha - \beta| \leq 4.93^{\circ}$ for crossed disparity and $|\alpha - \beta| \leq 1.57^{\circ}$ for uncrossed disparity.

However, observers may still feel discomfort even though they are able to view stereo 3D images. Hence, a theory "*zone of comfortable viewing*" has been proposed and several studies have been conducted [20-22, 99-102]. The zone of comfortable viewing is not tightly defined, but the concept can be understood from Fig. 43. The difference between the zone of comfortable viewing and the vergence-accommodation conflict is that the tolerable discrepancy is smaller in the forme. Wopking [103] claimed that human subjects will not experience any discomfort in viewing a stereo 3D image if $|\alpha - \beta| \leq 1^{\circ}$ (Fig. 43).

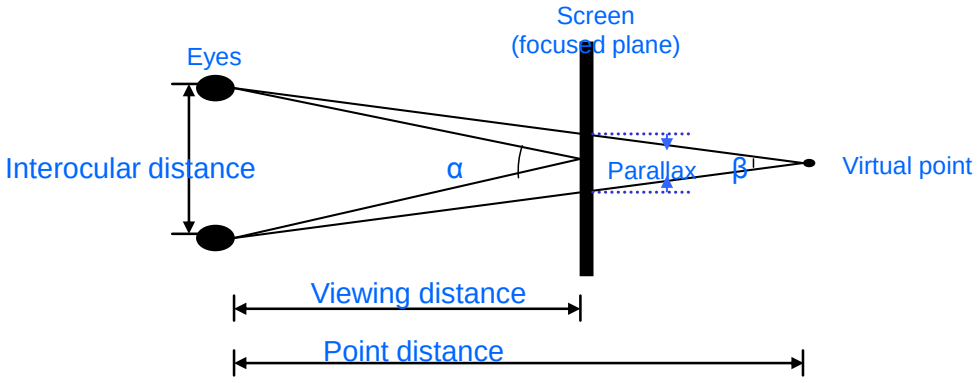

Fig. 43 Zone of comfortable stereo viewing

This chapter further discusses the 3D viewing experience (since there is no distortion, the 3D viewing experience is affected by the quality of depth) when a 3D image is displayed such that it is within - the "zone of comfortable viewing." To improve the stereo viewing experience, I propose post-processing techniques that not only cause the disparities in a 3D presentation to fall within the zone of comfortable viewing but also deliver an optimal 3D viewing experience. Specifically, we seek to compute the best 3D presentation that delivers the most pleasant 3D viewing experience.

7.2 Presentation model

Our approach analyzes the content of a given 3D image in order to deliver a more pleasant 3D presentation by avoiding conflicts in 3D viewing and optimizing stereoscopic depth resolution. Human depth perception is affected by both monocular

cues and binocular cues [16]. Conflicts between depth cues may create viewing discomfort or ambiguity in perceiving depths. Although it is not yet clear how the brain integrates these cues and produces a final sense of depths, it is rare to experience conflicting depth cues when viewing natural 3D images. Avoiding conflicts of depth cues by post-processing of the disparity values will help produce pleasant 3D viewing experiences.

In addition, when viewing a stereo 3D image on a stereo 3D display, the focused plane (accomodation of our eyes) is fixed on the screen. However, in our daily 3D vision, the accomodation of our eyes constantly changes as the vergence varies while scanning a 3D scene. Since the focal plane is fixed and the vergence plane may vary when viewing stereo 3D images on a display, the disconnect between accomodation and vergence may reduce the quality of depth percept.

7.2.1 Foreground/ background dominance

The human eyes have a very wide field of view [104], and thus images that we see in our daily life are likely to be mostly "background dominant". Hence, to avoid conflicts between depth cues, the composition of a stereo 3D image should be carefully

considered as an integral part of post-processing its disparity values. For example, if a stereo image is deemed to be "background dominant", then it should be disparity shifted so that it appears to be placed farther in the depth when displayed on 3D. Conversely, the presentation of a "foreground dominant" 3D image should be placed closer to the viewer.

To implement this idea, a foreground/background dominant classification process is needed. We have found two factors that can be used to sucessfully classify 3D images in this way: the skew of the disparity distribution, and Relative Dominant Depth (RDD) in the 3D image. The skew is computed as

$$skewness = \frac{\frac{1}{n}\sum_{i=1}^{n}(d_i - \bar{d})^3}{\left(\frac{1}{n}\sum_{i=1}^{n}(d_i - \bar{d})^2\right)^{\frac{3}{2}}} \tag{14}$$

where d_i is the disparity value of a pixel and $\bar{d}$ is the mean disparity of the 3D image. Then a 3D image is classified being "foreground dominant" if skewness > 1 $\text{skewness} > 1$ and "background dominant" if skewness < –1 $\text{skewness} < -1$. Images which have |skewness| < 1 $|\text{skewness}| < 1$ either have a non-normal disparity distribution or cannot be classified by skewness. In this case, the RDD is used to force a classification, where

$$RDD = \frac{(dominant\ disparity - minimum\ disparity)}{(maximum\ disparity - minimum\ disparity)} \tag{15}$$

and the $dominant\ disparity$ is the mode of the given disparity set. A 3D image is classified as "foreground dominant" if $|RDD| < \xi$ and "background dominant" if $|RDD| > \xi$, where $\xi = 0.25$ in this work.

7.2.2 Maximizing depth resolution

Field, et al. [105] showed that a minimum degree of variation in disparity that is needed for the human vision system to perceive different depths between objects. This is called the stereo threshold. Studies [16, 57, 106] have shown that the lowest thresholds are generally obtained at a zero pedestal disparity, and the threshold increases with increasing crossed or uncrossed pedestal disparity. The function which provides the stereo threshold at different disparities is called the stereoacuity function [57]. Fig. 44 shows the stereoacuity function of a female subject having normal stereo vision; her minimal threshold disparity is 24 arcsec at zero disparity. One can see clearly that human stereoacuity is most sensitive at the focus plane (the viewing screen in the 3D viewing of stereo images) and this observation indicates that human vision system has the highest depth resolution for objects around zero disparity.

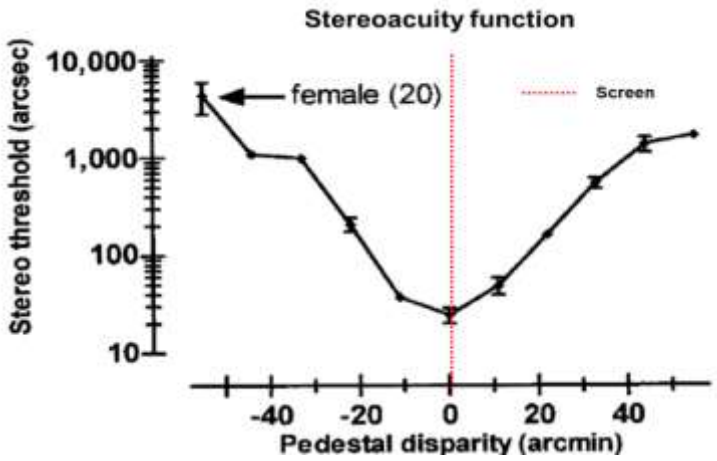

Fig. 44. Stereoacuity function, *s(d)* with 100 ms stimuli from [57]

Consider the case in which two objects have a relative disparity of 120 arcsec, but an average disparity of zero. (i.e., they have disparities of +1 and -1 arcmin, respectively) The subject, who has the stereoacuity function shown in Fig. 3, should see these two objects as laying in different depth planes. Now consider the case in which they have the same relative disparity, but one has a pedestal disparity of +40 arcmin and thus the other has a pesdestal disparity of +42 arcmin. In this case, the dispairty between them is below threshod, and no relative depth will be perceived. Hence, we claim that the subject can see depth with better resolution when the two objects are arranged around the zero pedestal disparity. To quantify the ability to resolve depth, we approximate the negative of the stereoacuity function, i.e. 1 – *s(d)*, in terms of the pixel disparity with a Gaussian

function, and call it the "depth resolution function" here. Then we solve the problem of optimizing the 3D presentation of a 3D image by maximizing the perceived depth resolution. This operation can be expressed by

$$opt\ shift = \underset{-255<i255}{\operatorname{argmax}}\ DRF \cdot Hist(i) \tag{16}$$

where the *DRF* is the depth resolution function using σ = 20 arcmin, which is chosen to give the best fit to the stereoacuity function. *Hist(0)* is the histogram of the disparity of a 3D image without being post-processed and *Hist(i)* is the histogram of the disparity of a 3D image shifted by *i*. Based on this operation, the shift value that yields the maximum product (i.e. *opt shift*) is deemed to provide a best 3D viewing experience in depth. The process of the shifting is illustrated in Fig. 45 and Fig. 46. Fig. 45 is a stereo image from the Middlebury stereo database. Fig. 46 illustrates the post-processing done on these images: a global shift of the left and right images to reduce/increase their disparity values. The global shift does not change the relative disparity values within a stereo pair. Indeed, some stereo display systems allow users to change this global shift manually to choose their preferred 3D effect.

Fig. 45 A stereo image

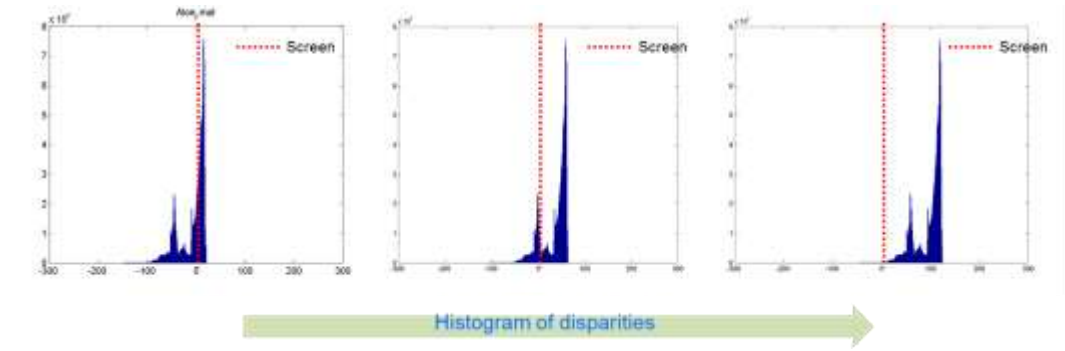

Fig. 46 Shifting the stereo image inside zone of comfortable viewing

7.2.3 Optimizing the presentation

The proposed algorithm is

1. Classify an input 3D image into either foreground or background dominant image as described in Sec. 7.2.1.
2. For the foreground dominant image, as described in Sec. 7.2.2, find the shift value that yields the maximum product of the depth resolution function and the disparity histogram.
3. For the background dominant image, find the shift value that places the closest surface on the screen.
4. Shift the left and right images so that the resulting 3D image has the desired depth according to the shift value found in Step 3 or Step 4. Then crop the undefined pixels on the boundary. For example, after an image is shifted to left by 3 columns, there are three undefined columns on the right side of the image.

The overall algorithm is described in Fig. 47.

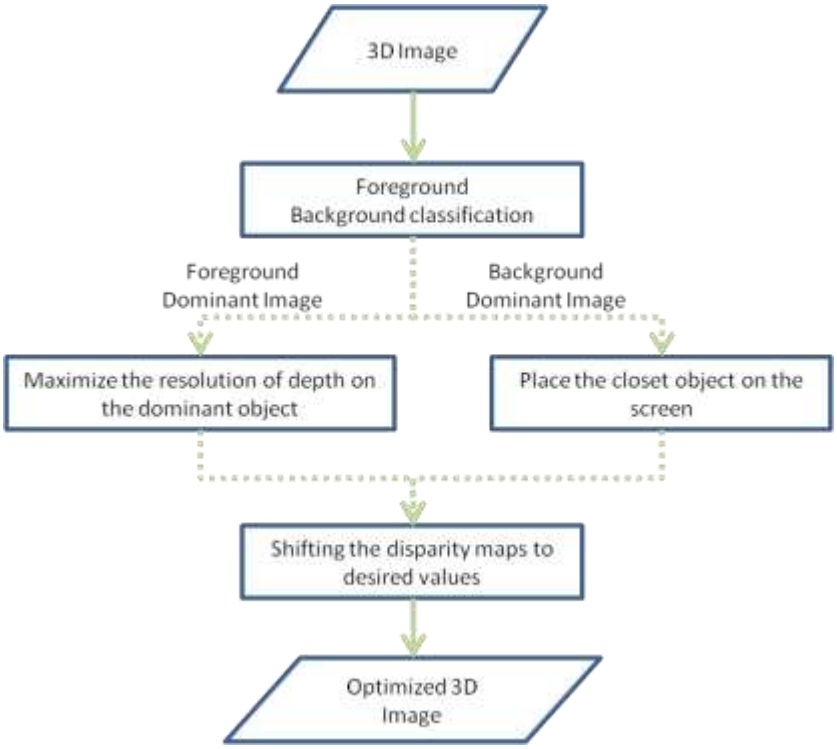

Fig. 47. Flowchart of the proposed algorithm.

A human study was conducted to assess the above algorithm. The study is described next.

7.3.1 Study design

A double stimulus continuous quality scale (DSCQS) protocol [55] was adopted to obtain subjective 3D quality ratings. During a single trial, a subject compared two 3D images with different depth relative to the screen and gave both of them a subjective 3D

quality score based on his/her preference. The question given to subjects is “Give a 3D viewing quality rating”. Since it is a forced-choice procedure, the subjects could not rate both images as having equal 3D quality scores. A training session was also undertaken at the beginning of the study to familiarize them with the Graphical User Interface (GUI) of our study program. The training content was different from the images used in the study. Repeated viewing of the same 3D image was allowed before the subject gave a rating. The GUI is shown in Fig. 48.

Fig. 48 The GUI of DSCQS in our study

7.3.2 Display

An nVidia active 3D kit plus an Alienware OptX AW2310 full HD 3D monitor were used to display the 3D images. The viewing distance from subjects to screen was five times the screen height to minimize potential visual discomfort caused by the accommodation-vergence conflict.

7.3.3 Observers

Seventeen naïve observers (seven females and ten males) were recruited for the study. The subjects were pre-screened to ensure normal stereovision by asking them to distinguish the depth of three colored rectangles separated from each other by 6 arcmin in depth.

7.3.4 Stimuli

Twelve stereo images with ground truth disparity were chosen as source images. Seven of these stereo images were captured by the parallel camera configuration and were taken from the Middlebury stereo database [70], and five of them were artificial 3D stereo images (three were from MPEG 3D coding test videos, two were created by me). We used an approximately equal number of “foreground dominant” and “background dominant” image. The original resolution of the images was equal to or larger than full HD size and they were resized to full HD resolution by cropping the extra part.

To create a baseline without ground truth depth, the reference image was created by placing the closest surface inside the image at the depth of the screen. Then, four different stimuli were created by either pulling the scene in front of screen or by pushing it deeper into the screen by shifting disparities. The distance of two adjacent stimuli is 13.7 arcmin. In addition, one scene was created by maximizing the depth resolution, as described in Session 7.2.2. Finally, all stimuli have disparities that satisfy the "zone of comfort viewing" suggested by [103].

7.4 Results and analysis

Differential Mean Opinion Score (DMOS) are usually used as quality scores annotated to the content in image quality database. However, all of the stimuli in this study are pristine images, so comparing depth quality among stimuli which have different content is meaningless. On the contrary, intra-content comparisons can provide insights regarding the best perceptual 3D depth range of a stereo pair. Hence, the average ranking given by human subjects is used as a criterion to evaluate the performance of 3D images displayed with different (shifted) depth ranges.

Six different profiles were used for each source image. The ranking of each source image ranges from 1 (the best) to 6 (the worst). The performance achieved by a

3D presentation is represented by the average ranking over twelve source images. Two types of rankings were used. The first is “ranking DMOS (weighted ranking)” which is the ranking sorted by DMOS scores. The second is “ranking vote”, which only considers binary decisions (stimulus A gets one vote if one subject prefers stimulus A over stimulus B) and the ranking is sorted by the voting results. The overall ranking is the average of these two rankings.

The experimental results are shown in Fig. 49. The “closer” profile is to set the disparity value of the closest surface at -13.7 arcmin (crossed disparity) and the disparity value of the closest surface for “farther” and “farthest” profiles are 13.7 arcmin and 27.4 arcmin respectively. Four observations can be made from Fig. 49. First, the reference strategy, which places the closest surface on the screen, has a ranking of 3.25. This ranking is slightly better than the expected ranking (3.5) when the nearest surface is placed randomly inside the zone of comfortable viewing. Second, comparing the “closer” and “farther” profiles, we observe when extra computation is not allowed, the better strategy is to push the closest surface deeper into the screen rather than pull it out of the screen. Third, the strategy which maximizes the depth resolution performs better than the reference strategy, but worse than the “farther” profile. A plausible explanation is that

this strategy works for “foreground dominant” 3D images, but creates depth cue conflicts for “background dominant” 3D images. Finally, the proposed algorithm which applies both of strategies based on content gives the best performance (overall ranking is 1.83). Fig. 50 shows an example of the proposed algorithm. In Fig. 50, the left image is a foreground dominant image and we found that the best 3D presentation is to place the dominant depth on the screen. The right image is a background dominant image and the best 3D presentation is to place the closet object on the screen.

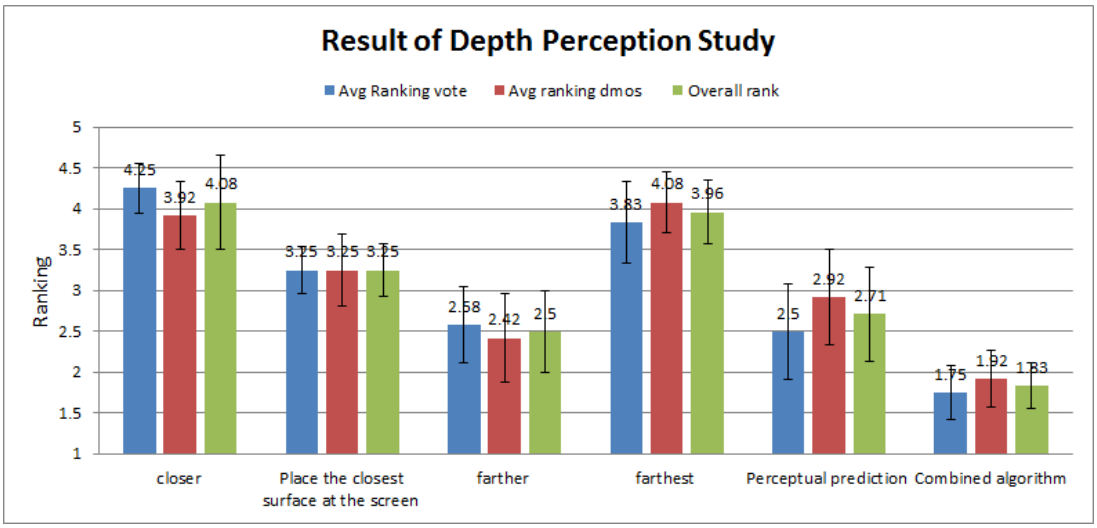

Fig. 49 Performance of different 3D image presentation strategies. Error bars represent the standard errors.

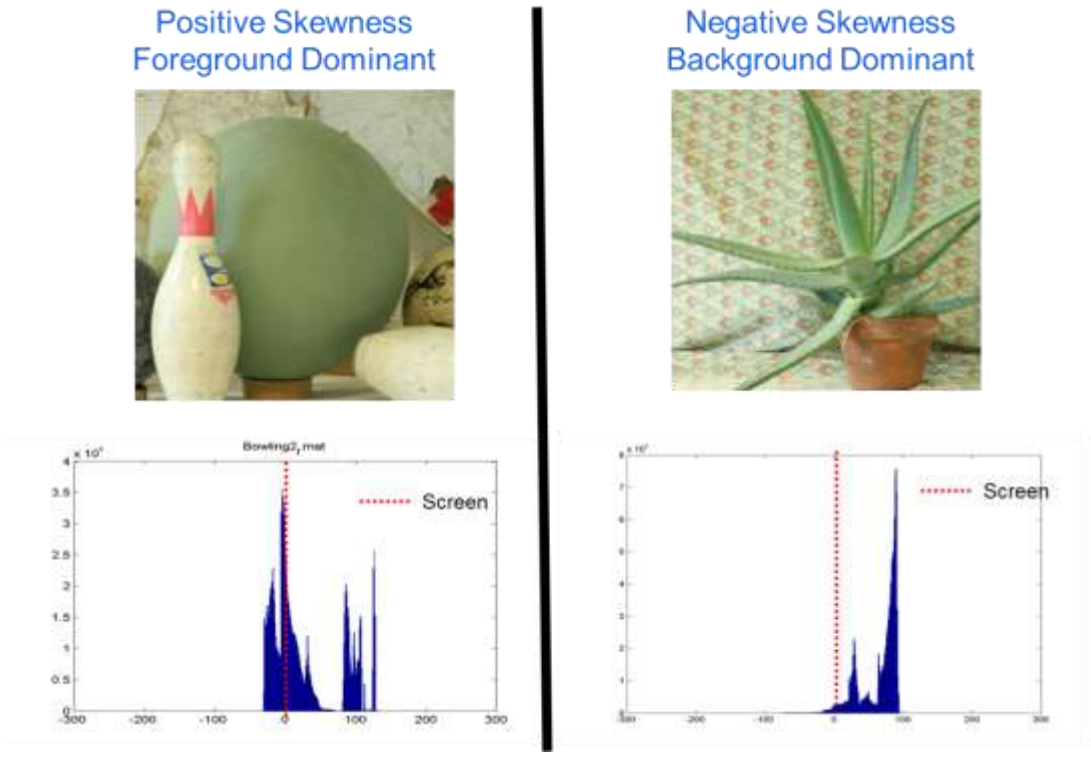

Fig. 50 Top Left: a foreground dominant image with a positively skewed disparity distribution.
Top Bottom: The histogram of disparities of the rank one stereo representation.
Top Right: a background dominant image with a negatively skewed disparity distribution.
Top Right: The histogram of disparities of the rank one stereo representation.

7.5 Conclusion

We believe that the degree of comfort in viewing a 3D image as a function of the depth range it is assigned is correlated with the stereoacuity function of the human visual system and the content of the 3D image. A human study was conducted which supports

our argument. The following points can be further considered. First, there should be a better strategy in post processing “background dominant” images. Our current strategy is to simply place the closest surface at the screen for the background dominant images. Second, other content-related factors such as object contours and the composition of a 3D image may affect the perception of a stereo 3D image.

CHAPTER 8 CONCLUSION AND FUTURE WORKS

Assessing perceived 3D quality of stereoscopic 3D images is a much more complex task than predicting 2D quality and there exist very few 3D QA algorithms at the time of this writing. Research suggests "perceived 3D quality" is affected by "spatial image quality", "depth quality", and "visual comfort". In this dissertation, I focused on the influences of "spatial image quality" and "depth quality" on "perceived 3D quality" and minimized potential visual discomfort in all conducted studies by following the suggestion of "zone of comfortable viewing". My first study confirmed that perceived 3D quality is affected by both spatial image quality and depth quality, but it also found that different subjects tend to have different opinions on depth quality when the viewed stereo content is distorted. The result of the first study indicated that more research studies should be conducted to understand these diverse quality ratings given to depth quality before designing a QA model to predict depth quality of distorted stereo 3D content. Moreover, the spatial image quality was shown to be highly correlated to perceived 3D quality and the quality ratings given by different subjects were much more coherent. Thus, advancing the ability to predict spatial image quality should improve the performance of assessing overall perceived 3D quality.

Research suggests that spatial image quality of a stereo image is not simply the average of spatial qualities of the left and right views. Further, there exists no research which discusses depth masking effects on stereo images that are distorted by white noise, JPEG compression, JPEG2000 compression, blur, or fast fading. I conducted two more human studies to discuss possible masking effects of stereo images that are distorted by these five distortion types. The study results indicated that binocular rivalry affects the spatial image quality and there is no depth masking effect for distorted stereo images.

Based on my observations from the human studies, I designed a FR 3D QA framework and a NR 3D QA model. My proposed FR 3D QA framework can be used as a plug-in with any 2D FR QA model and improve its performance in predicting perceived 3D quality. The key idea of my proposed framework is to perform 3D quality assessment on synthesized "cyclopean" images, which models binocular rivalry using a previously proposed linear model. To verify this design, I conducted another large scale human study to create a 3D image quality database (called LIVE 3D Image Quality Database Phase II). The experimental results confirmed the intuition of the design and verified that– this framework improves the performance of 3D QA when there is binocular rivalry

while viewing a stereo image. For the case when there is no binocular rivalry, this framework performs as well as 2D QA models.

My proposed 3D NR QA model drew inspiration from research in the area of natural scene statistics on 2D images and depth maps. I showed that the statistical features from the estimated disparity map and its estimation errors are reliable features to predict perceived 3D quality of distorted stereo images. The 3D NR QA model was also verified with LIVE 3D Image Quality Database.

In the last part of my dissertation, I discussed depth quality of distortion-free stereo images, which have different 3D presentations but equal spatial resolution and quality. The objective of this part was to show that the task of predicting perceived 3D quality goes beyond modeling spatial or disparity distortions. The experimental results show that perceived depth quality is correlated with the human stereoacuity function and our high level understanding of the real world.

In summary, the contributions of this dissertation are:

1. Fundamental observations on binocular rivalry and depth masking effects on distorted stereo images. These observations demonstrate possible directions to improve the performance of 3D QA models;

2. A demonstration that modeling binocular rivalry can improve the performance of 3D QA models;
3. A demonstration that statistics of estimated disparity and the estimation errors are reliable features to predict perceived 3D Quality, and finally
4. A demonstration that 3D presentation of a distortion-free stereo image can be predicted by considering human stereoacuity function and our high-level understanding of the real world.

These contributions of this dissertation expand our knowledge of perceived 3D quality of stereo images and enable advances in the development of high-performance 3D QA models.

3D QA models to predict 3D quality of depth-image-based-rendering (DIBR) content would be a good extension of this dissertation. DIBR generated 3D images may have distortions caused by hole-filling algorithms, 3D warping algorithms, and errors from depth estimation, but a human subject may experience binocular rivalry while viewing DIBR generated 3D images. In addition, DIBR generated images have greater control on the 3D representation. Thus, the 3D presentation model proposed in this dissertation could provide a good reference to generate stereo image with depth quality.

Bibliography

[1] C. Wheatstone, "Contributions to the Physiology of Vision. Part the First. On Some Remarkable, and Hitherto Unobserved, Phenomena of Binocular Vision," *Philosophical Transactions of the Royal Society of London*, vol. 128, pp. 371-394, January 1, 1838.

[2] Fuji. *FinePix Real 3D W1 camera*. Available: http://www.lhup.edu/~dsimanek/3d/stereo/3dgallery12.htm

[3] MPAA. (2011, Theatrical Market Statistics. Available: http://www.mpaa.org/resources/5bec4ac9-a95e-443b-987b-bff6fb5455a9.pdf

[4] *List of 3D movies*. Available: http://en.wikipedia.org/wiki/List_of_3-D_films

[5] J. Cameron. (2009). *Avatar*. Available: http://en.wikipedia.org/wiki/Avatar_(2009_film)

[6] S. Chikkerur, V. Sundaram, M. Reisslein, and L. J. Karam, "Objective Video Quality Assessment Methods: A Classification, Review, and Performance Comparison," *IEEE Trans. Broadcasting*, vol. 57, pp. 165 -182, 2011.

[7] Z. Wang and A. C. Bovik, *Modern Image Quality Assessment*: Mprgan & Claypool, 2006.

[8] A. K. Moorthy, C. C. Su, A. Mittal, and A. C. Bovik, "Subjective Evaluation of Stereoscopic Visual Quality," *Signal Processing: Image Communication, Special Issue on Biologically Inspired Approaches for Visual Information Processing and Analysis*, 2012.

[9] W. J. M. Levelt, *On binocular rivalry*. The Hague; Paris: Mouton, 1968.

[10] W. J. Tam, L. B. Stelmach, and P. J. Corriveau, "Psychovisual aspects of viewing stereoscopic video sequences," *Proc. SPIE,* pp. 226-235, 1998.

[11] R. Blake and N. K. Logothetis, "Visual competition," *Nat Rev Neurosci,* vol. 3, pp. 13-21, 2002.

[12] 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. Circ. Syst. for Video Tech.,* vol. 14, pp. 381-391, 2004.

[13] R. Sekuler and R. Blake, *Perception*. New York: A.A. Knopf, 1985.

[14] N. A. Polak and R. Jones, "Dynamic interactions between accommodation and convergence," *IEEE transactions on bio-medical engineering,* vol. 37, pp. 1011-4, 1990.

[15] B. Julesz, *Foundations of cyclopean perception*. Chicago: University of Chicago Press, 1971.

[16] I. P. Howard and B. J. Rogers, *Binocular vision and stereopsis*. New York: Oxford University Press, 1995.

[17] C. Schor, I. Wood, and J. Ogawa, "Binocular sensory fusion is limited by spatial resolution," *Vision research,* vol. 24, pp. 661-5, 1984.

[18] C. M. Schor and C. W. Tyler, "Spatio-temporal properties of Panum's fusional area," *Vision research,* vol. 21, pp. 683-92, 1981.

[19] J. Burge, M. A. Peterson, and S. E. Palmer, "Ordinal configural cues combine with metric disparity in depth perception," *Journal of Vision,* vol. 5, June 22 2005.

[20] D. M. Hoffman, A. R. Girshick, K. Akeley, and M. S. Banks, "Vergence-accommodation conflicts hinder visual performance and cause visual fatigue," *Journal of vision,* vol. 8, pp. 1-30, 2008.

[21] W. J. Tam, F. Speranza, S. Yano, K. Shimono, and H. Ono, "Stereoscopic 3D-TV: Visual Comfort," *IEEE Trans. Broadcasting,* vol. 57, pp. 335-346, 2011.

[22] M. T. M. Lambooij, W. A. Ijsselsteijn, and I. Heynderickx, "Visual discomfort in stereoscopic displays: a review " *Proc. SPIE,* vol. 6490, p. 64900I, 2007.

[23] W. Lili, K. Teunissen, T. Yan, C. Li, Z. Panpan, Z. Tingting, and I. Heynderickx, "Crosstalk Evaluation in Stereoscopic Displays," *Journal of Display Technology,* vol. 7, pp. 208-214, 2011.

[24] N. S. Holliman, N. A. Dodgson, G. E. Favalora, and L. Pockett, "Three-Dimensional Displays: A Review and Applications Analysis," *IEEE Trans. Broadcasting,* vol. 57, pp. 362-371, 2011.

[25] F. Speranza, W. J. Tam, R. Renaud, and N. Hur, "Effect of disparity and motion on visual comfort of stereoscopic images " *Proc. SPIE,* vol. 6055, 2006.

[26] S. Winkler, "On the properties of subjective ratings in video quality experiments," in *QoMEx* 2009, pp. 139-144.

[27] 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. Appl. Percept.*, vol. 3, pp. 95-109, 2006.

[28] 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," *J. Exp. P.:Appl.*, vol. 7, pp. 143-53, 2001.

[29] 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*, vol. 13, pp. 600-612, 2004.

[30] Z. Wang, E. Simoncelli, and A. Bovik, "Multi-scale Structural Similarity for Image Quality Assessment," *Asilomar Conf. on Signals, Syst. and Computers*, vol. 2, pp. 1398-1402, 2003.

[31] Z. Wang and E. P. Simoncelli, "Reduced-reference image quality assessment using a wavelet-domain natural image statistic model [5666-20]," *Proc. SPIE*, vol. 5666, pp. 149-159, 2005.

[32] M.-J. Chen and A. C. Bovik, "No-reference image blur assessment using multiscale gradient," in *QoMEx*, 2009, pp. 70-74.

[33] H. R. Sheikh, A. C. Bovik, and L. Cormack, "No-reference quality assessment using natural scene statistics: JPEG2000," *IEEE Trans. Image Processing*, vol. 14, pp. 1918-27, 2005.

[34] A. K. Moorthy and A. C. Bovik, "A Two-Step Framework for Constructing Blind Image Quality Indices," *IEEE Signal Processing Letters,* vol. 17, pp. 513-516, 2010.

[35] Y. Peng and D. Doermann, "No-Reference Image Quality Assessment Using Visual Codebooks," *IEEE Trans. Image Processing,* vol. 21, pp. 3129 -3138, 2012.

[36] A. Mittal, A. K. Moorthy, and A. C. Bovik, "Blind/referenceless image spatial quality evaluator," *IEEE Trans. Image Processing,* vol. To appear, 2012.

[37] M. A. Saad and A. C. Bovik, "Blind Image Quality Assessment: A Natural Scene Statistics Approach in the DCT Domain," *IEEE Trans. Image Processing,* vol. 21, pp. 3339-3352, 2012.

[38] A. K. Moorthy and A. C. Bovik, "Blind image quality assessment: from natural scene statistics to perceptual quality," *IEEE Trans. Image Processing,* vol. 20, pp. 3350-3364, 2011.

[39] S. L. P. Yasakethu, C. T. E. R. Hewage, W. A. C. Fernando, and A. M. Kondoz, "Quality Analysis for 3D Video Using 2D Video Quality Models," *IEEE Trans. Consumer Electronics,* vol. 54, pp. 1969-1976, 2008.

[40] 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,* vol. 44, pp. 963-965, 2008.

[41] P. Gorley and N. Holliman, "Stereoscopic image quality metrics and compression " *Proc. SPIE,* vol. 6803, p. 05, 2008.

[42] D. G. Lowe, "Object recognition from local scale-invariant features," *ICCV*, 1999.

[43] M. A. Fischler and R. C. Bolles, "Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography," *ACM Commun.*, vol. 24, pp. 381--395, 1981.

[44] A. Benoit, P. Le Callet, P. Campisi, and R. Cousseau, "Quality Assessment of Stereoscopic Images," *EURASIP Journal on Image and Video Processing*, vol. 2008, pp. 1-14, 2008.

[45] C.-C. Su, A. C. Bovik, and L. K. Cormack, "Natural scene statistics of color and range," *Vision*, vol. 11, p. 1190, 2011.

[46] J. You, L. Xing, A. Perkis, and X. Wang, "Perceptual quality assessment for stereoscopic images based on 2D image quality metrics and disparity analysis," presented at the Int. Workshop Video Processing and Quality Metrics, 2010.

[47] Z. Zhu and Y. Wang, "Perceptual distortion metric for stereo video quality evaluation," *WSEAS Trans. Signal Process.*, vol. 5, pp. 241-250, 2009.

[48] J. Yang, C. Hou, Y. Zhou, Z. Zhang, and J. Guo, "Objective quality assessment method of stereo images," in *3DTV Conference: The True Vision - Capture, Transmission and Display of 3D Video*, 2009, pp. 1-4.

[49] A. Maalouf and M. C. Larabi, "CYCLOP: A stereo color image quality assessment metric," in *IEEE Int. Conf. Acoustics, Speech and Signal Processing*, 2011, pp. 1161-1164.

[50] M.-J. Chen, A. C. Bovik, and L. K. Cormack, "Study on distortion conspicuity in stereoscopically viewed 3D images," in *IVMSP Workshop, 2011 IEEE 10th*, 2011, pp. 24-29.

[51] R. Bensalma and M.-C. Larabi, "A perceptual metric for stereoscopic image quality assessment based on the binocular energy," *Multidimensional Systems and Signal Processing*, pp. 1-36, 2012.

[52] C. T. E. R. Hewage and M. G. Martini, "Reduced-reference quality metric for 3D depth map transmission," in *3DTV Conference: The True Vision - Capture, Transmission and Display of 3D Video*, 2010, pp. 1-4.

[53] R. Akhter, Z. M. P. Sazzad, Y. Horita, and J. Baltes, "No-reference stereoscopic image quality assessment," *Proc. SPIE*, vol. 7524, p. 7524 0T, 2010.

[54] L. Goldmann, F. De Simone, T. Ebrahimi, P. Three-Dimensional Image, and Applications, "A comprehensive database and subjective evaluation methodology for quality of experience in stereoscopic video," *Proc. SPIE*, vol. 7526, 2010.

[55] I. T. U. R. Assembly and U. International Telecommunication, *Methodology for the subjective assessment of the quality of television pictures*. Geneva, Switzerland: International Telecommunication Union, 2003.

[56] W. Richards, "Stereopsis and stereoblindness," *Experimental brain research. Experimentelle Hirnforschung. Expérimentation cérébrale*, vol. 10, pp. 380-8, 1970.

[57] 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, vol. 44, pp. 891-900, February 1 2003.

[58] P. Seuntiëns. (2006). *Visual experience of 3D TV*. Available: http://library.tue.nl/csp/dare/LinkToRepository.csp?recordnumber=609714

[59] R. Blake, D. H. Westendorf, and R. Overton, "What is suppressed during binocular rivalry?," *Perception,* vol. 9, pp. 223-31, 1980.

[60] P. Whittle, "Binocular rivalry and the contrast at contours," *The Quart. J. of Expt. Psych.*, vol. 17, pp. 217-226, 1965.

[61] I. T. Kaplan and W. Metlay, "Light intensity and binocular rivalry," *Journal of Experimental Psychology*, vol. 67, pp. 22-26, 1964.

[62] M. Fahle, "Binocular rivalry: suppression depends on orientation and spatial frequency," *Vision research,* vol. 22, pp. 787-800, 1982.

[63] N. K. Logothetis and J. D. Schall, "Neuronal correlates of subjective visual perception," *Science*, vol. 245, pp. 761-763, 1989.

[64] D. A. Leopold and N. K. Logothetis, "Activity changes in early visual cortex reflect monkeys' percepts during binocular rivalry," *Nature*, vol. 379, pp. 549-553, 1996.

[65] D. Alais and R. Blake, "Grouping visual features during binocular rivalry," *Vision research,* vol. 39, pp. 4341-53, 1999.

[66] M. K. Kapadia, M. Ito, C. D. Gilbert, and G. Westheimer, "Improvement in visual sensitivity by changes in local context: parallel studies in human observers and in V1 of alert monkeys," *Neuron,* vol. 15, pp. 843-56, 1995.

[67] D. J. Field, A. Hayes, and R. F. Hess, "Contour integration by the human visual system: evidence for a local "association field"," *Vision research,* vol. 33, pp. 173-93, 1993.

[68] R. P. O'Shea, A. J. Sims, and D. G. Govan, "The effect of spatial frequency and field size on the spread of exclusive visibility in binocular rivalry," *Vision research,* vol. 37, pp. 175-83, 1997.

[69] R. Blake, R. P. O'Shea, and T. J. Mueller, "Spatial zones of binocular rivalry in central and peripheral vision," *Visual neuroscience,* vol. 8, pp. 469-78, 1992.

[70] D. Scharstein and R. Szeliski, "A Taxonomy and Evaluation of Dense Two-Frame Stereo Correspondence Algorithms," *International Journal of Computer Vision,* vol. 47, pp. 7-42, 2002.

[71] A. Klaus, M. Sormann, and K. Karner, "Segment-Based Stereo Matching Using Belief Propagation and a Self-Adapting Dissimilarity Measure," in *ICPR 2006.* , 2006, pp. 15-18.

[72] D. J. Field, "Relations between the Statistics of Natural Images and the Response Properties of Cortical-Cells," *Journal of the Optical Society of America a-Optics Image Science and Vision,* vol. 4, pp. 2379-2394, Dec 1987.

[73] C.-C. Su, A. Bovik, and L. Cormack, "Natural scene statistics of color and range," *Journal of Vision,* vol. 11, p. 1190, September 23, 2011 2011.

[74] C. W. Tyler, "Stereoscopic depth movement: two eyes less sensitive than one," *Science,* vol. 174, pp. 958-61, 1971.

[75] 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*, vol. 15, pp. 3440-3451, 2006.

[76] M. G. Perkins, "Data compression of stereopairs," *IEEE Trans. Commun.*, vol. 40, pp. 684-696, 1992.

[77] F. Allenmark and J. Read, "Spatial Stereoresolution for Depth Corrugations May Be Set in Primary Visual Cortex," *Plos Computational Biology*, vol. 7, p. e1002142, 2011.

[78] B. N. Vlaskamp, G. Yoon, and M. S. Banks, "Neural and optical constraints on stereoacuity," *Perception 37 ECVP Abstract Supplement*, p. 2, 2008.

[79] F. W. Campbell and D. G. Green, "Optical and retinal factors affecting visual resolution," *J. Physiology*, vol. 181, pp. 576-593, 1965.

[80] A. Vetro, A. M. Tourapis, K. Muller, and C. Tao, "3D-TV Content Storage and Transmission," *Broadcasting, IEEE Transactions on*, vol. 57, pp. 384-394, 2011.

[81] M. Lambooij, W. Ijsselsteijn, D. G. Bouwhuis, and I. Heynderickx, "Evaluation of Stereoscopic Images: Beyond 2D Quality," *IEEE TRANSACTIONS ON BROADCASTING*, vol. 57, pp. 432-444, 2011.

[82] M.-J. Chen, D.-K. Kwon, and A. C. Bovik, "Study of Subject Agreement on Stereoscopic Video Quality," presented at the Proc. IEEE Southwest Symp. Image Anal., Interp, Santa Fe, New Mexico, USA 2012.

[83] P. Ye and D. Doermann, "No-Reference Image Quality Assessment Using Visual Codebooks," *IEEE Trans. Image Processing*, vol. 21, pp. 3129 -3138, 2012.

[84] R. Granit, *Receptors and sensory perception*: Yale University Press, 1955.

[85] D. J. Heeger, "Normalization of cell responses in cat striate cortex," *Visual neuroscience,* vol. 9, pp. 181-197, 1992.

[86] A. Srivastava, A. B. Lee, E. P. Simoncelli, and S.-C. Zhu, "On advances in statistical modeling of natural images," *Journal of mathematical imaging and vision,* vol. 18, pp. 17-33, 2003.

[87] D. L. Ruderman, "The statistics of natural images," *Network: Computation in Neural Systems,* vol. 5, pp. 517-548, 1994.

[88] K. Sharifi and A. Leon-Garcia, "Estimation of shape parameter for generalized Gaussian distributions in sbuband decomposition of video," *IEEE Trans. Circ. Syst. for Video Tech.,* vol. 5, pp. 52-56, 1995.

[89] J. Huang, A. B. Lee, and D. Mumford, "Statistics of Range Images," in *CVPR*, 2000, pp. 324--331.

[90] Z. Yang and D. Purves, "Image/source statistics of surfaces in natural scenes," *Network (Bristol, England),* vol. 14, pp. 371-390, 2003.

[91] P. Hibbard, "A statistical model of binocular disparity," *Visual Cognition,* vol. 15, pp. 149-165, 2007.

[92] Y. Liu, A. C. Bovik, and L. K. Cormack, "Disparity statistics in natural scenes," *Journal of Vision,* vol. 8, 2008.

[93] C. C. Chang and L. C. J., "LIBSVM: A library for suport vector machines," 2001.

[94] M.-J. Chen, C.-C. Su, D.-K. Kwon, L. K. Cormack, and A. C. Bovik, "Full-Reference Quality Assessment of Stereopairs Accounting for Rivalry," *Asilomar Conf. on Signals, Syst. and Computers*, 2012.

[95] HTC. (2011). *HTC EVO 3D*. Available: http://www.htc.com/us/products/evo3d-sprint

[96] E. F. Fincham and J. Walton, "The reciprocal actions of accommodation and convergence," *The Journal of physiology*, vol. 137, pp. 488-508, 1957.

[97] S. J. Daly, R. T. Held, and D. M. Hoffman, "Perceptual Issues in Stereoscopic Signal Processing," *IEEE Trans. Broadcasting*, vol. 57, pp. 347-361, 2011.

[98] Y. Y. Yeh and L. D. Silverstein, "Limits of fusion and depth judgment in stereoscopic color displays," *Human factors*, vol. 32, pp. 45-60, 1990.

[99] N. A. Valyus, *Stereoscopy*. New York: Focal Press, 1966.

[100] T. Shibata, J. Kim, D. M. Hoffman, and M. S. Banks, "The zone of comfort: Predicting visual discomfort with stereo displays," *Journal of vision*, vol. 11, 2011.

[101] S. Yano, S. Ide, T. Mitsuhashi, and H. Thwaites, "A study of visual fatigue and visual comfort for 3D HDTV/HDTV images," *Displays.*, vol. 23, pp. 191-201, 2002.

[102] Y. Nojiri, H. Yamanoue, A. Hanazato, M. Emoto, and F. Okano, "Visual comfort/discomfort and visual fatigue caused by stereoscopic HDTV viewing," in *Prof. SPIE*, 2004, pp. 303-313.

[103] M. Wopking, "Viewing comfort with stereoscopic pictures: An experimental study on the subjective effects of disparity magnitude and depth of focus," *Journal of the Society for Information Display*, vol. 3, pp. 101-103, 1995.

[104] Y. Le Grand, *Light, colour and vision*. London: Chapman & Hall, 1968.

[105] D. J. Field, D. W. Jones, and G. Kneen, "1,5-Shift of Unsaturated Groups," *Journal of the Chemical Society-Chemical Communications*, pp. 873-874, 1976.

[106] D. M. Hoffman, A. R. Girshick, K. Akeley, and M. S. Banks, "Vergence–accommodation conflicts hinder visual performance and cause visual fatigue," *Journal of Vision*, vol. 8, March 28 2008.