THÈSE

Pour l'obtention du grade de

DOCTEUR DE L'UNIVERSITE DE POITIERS

Faculté des Sciences Fondamentales et Appliquées

Diplôme National - Arrêté du 25 mai 2016

ÉCOLE DOCTORALE SCIENCES ET INGENIERIE DES SYSTEMES, MATHEMATIQUES, INFORMATIQUE

DOMAINE DE RECHERCHE : TRAITEMENT DU SIGNAL ET DES IMAGES

Présentée par

Yu FAN

Quality assessment of stereoscopic 3D content based on binocular perception.

Directeurs de thèse :

Mohamed-Chaker LARABI

Faouzi ALAYA CHEIKH

Christine FERNANDEZ-MALOIGNE

Soutenue le 03 May 2019
Devant la Commission d'Examen

JURY

Federic Dufaux	Directeur de Recherche, CNRS L2S	Rapporteur
Mylène C. Q. Farias	Associate Professor, University of Brasília	Rapporteur
Sony George	Professor, NTNU	Examinateur
Faouzi Alaya Cheikh	Professeur, NTNU	Examinateur
Mohamed-Chaker Larabi	Maître de conférences, XLIM	Examinateur
Christine Fernandez	Professeur, XLIM	Examinateur

Acknowledgement

I would like to thank all the people who supported and helped me to conduct PhD research and complete this thesis.

First and foremost, I wish to express my deepest gratitude to my main adviser Associate Professor Mohamed-Chaker Larabi from the University of Poitiers, for not only guiding and inspiring my research work but also helping and supporting me to arrange my co-tutelle PhD program well. During the fours years of my PhD, he has invested a lot of time and efforts to teach me the fundamentals of good research methodology, experimental implementation, and lecture. Without his contributions, my PhD research would not be so productive. He always helped me to deal with the issues patiently which I encountered in my work and daily life in France, especially the most difficult period of my PhD.

I would also like to express my great appreciation to my another adviser Professor Faouzi Alaya Cheikh from the Norwegian University of Science and Technology (NTNU), who enabling collaboration between NTNU and UP to make my PhD research work more easily and life more convenient during my stay in Norway. Both advisers provided me with full freedom to explore the research topics on 3D technology and quality assessment, gave their suggestion, feedback, and foresight for my research. I would like also to thank my adviser Professor Christine Fernandez-Maloigne from UP for her guidance and encouragement. Furthermore, I also want to appreciate Professor Marius Pedersen at NTNU for his inspiration and discussion on quality assessment, and Professor Jon Yngve Hardeberg for his support and help during my short time in Norway.

In addition, I would like to sincerely appreciate evaluation committee members including, Doctor Frédéric Dufaux, who is the CNRS Research Director at Laboratoire des Signaux et Systèmes (L2S, UMR 8506), Associate Professor Mylène Farias from Department of Electrical Engineering at University of Brasilia, and Associate Professor Sony George from NTNU for their valuable feedback and comments.

Besides, I would also to thank my colleagues in Norwegian Color and Visual Computing Laboratory at NTNU, and XLIM Laboratory at UP: Congcong Wang, Xinwei Liu, Ahmed Kedir Mohammed, Mohib Ullah, Vlado Kitanovski, Sami Jaballah, Xinwen Xie, Shihang Liu and many more. Moreover, many thanks for persons who participated in my time-consuming psychophysical and subjective experiments, thanks for your support and contribution. Last but not least, I would like to appreciate my parents, Xing Fan and Guolan Wei, for their support and love. Furthermore, many warm thanks go to my friends: Jian Wu, Chennan Jiang, Hanxiang Wu, Xia Lan, and Shuiran Peng for their support

and encouragement during my difficult period.

Poitiers, France
April 4, 2019
Yu FAN

Contents

List of Figures vii

List of Tables ix

List of Abbreviations xi

I General Introduction 1

1 Introduction 3

1.1 Motivation . . . 3

1.2 Research Aims . . . 4

1.3 Research Questions . . . 4

1.3.1 Questions related to 3D-JND . . . 5

1.3.2 Questions related to SIQA . . . 5

1.4 Research Direction . . . 5

1.5 Research Methodology . . . 6

1.6 Dissertation Outline . . . 7

2 Background 9

2.1 Human visual system . . . 9

2.1.1 Depth perception . . . 10

2.2 Stereoscopic 3D imaging . . . 17

2.2.1 S3D content representation formats . . . 17

2.2.2 S3D displays . . . 18
2.2.3 Depth and binocular disparity . . . 20
2.3 Psychophysics and just noticeable difference . . . 21
2.3.1 Visual psychophysics . . . 21
2.3.2 Just noticeable difference threshold . . . 22
2.3.3 Psychophysical experiments . . . 23
2.4 Statistical tools for psychophysical experiments . . . 25
2.4.1 Subjects-related outliers . . . 26
2.4.2 Samples-related outliers . . . 26
2.4.3 Normality validation and analysis of variance . . . 27
2.5 Perceptual quality assessment of 2D and 3D images . . . 28
2.5.1 Subjective IQA methods . . . 28
2.5.2 Objective 2D-IQA methods . . . 29
2.5.3 Performance evaluation methods . . . 33

3 Summary of Results and Contributions **35**
3.1 Summary of Paper I . . . 35
3.2 Summary of Paper II . . . 37
3.3 Summary of Paper III . . . 38
3.4 Summary of Paper IV . . . 40
3.5 Summary of Paper V . . . 41
3.6 Summary of Paper VI . . . 43
3.7 Summary of Paper VII . . . 44

4 Discussion **47**
4.1 Contributions of the thesis . . . 47
4.1.1 Contributions to 3D-JND . . . 47
4.1.2 Contributions to SIQA . . . 49
4.2 Supplementary results . . . 51
4.3 Limitations and Shortcomings . . . 52

5 General Conclusions and Perspectives **59**
5.1 Conclusions . . . 59
5.2 Perspectives . . . 60

Bibliography **63**

II Included Papers **89**

6 Paper I: On the Performance of 3D Just Noticeable Difference Models **91**

6.1 Introduction . . . 92
6.2 3D-JND models . . . 92
6.3 Comparison of 3D-JND models . . . 95
6.4 Experiments . . . 97
6.5 Conclusion . . . 100

Bibliography **101**

7 Paper II: A Survey of Stereoscopic 3D Just Noticeable Difference Models **105**

7.1 Introduction . . . 106
7.2 Visual characteristics for 3D-JND models . . . 107
7.3 3D-JND models . . . 117
7.4 Comparison of 3D-JND models . . . 127
7.5 Experimental results . . . 132
7.6 Conclusion . . . 153

Bibliography **155**

8 Paper III: Just Noticeable Difference Model for Asymmetrically Distorted Stereoscopic Images **167**

8.1 Introduction . . . 168
8.2 Psychophysical experiments . . . 168
8.3 Psychophysical data analysis and modeling . . . 171
8.4 Experimental validation . . . 174
8.5 Conclusion . . . 176

Bibliography **177**

9 Paper IV: Stereoscopic Image Quality Assessment based on the Binocular Properties of the Human Visual System **181**

9.1 Introduction . . . 182
9.2 Related work . . . 183
9.3 The Proposed SIQA Method . . . 184

9.4 Experiemntal results 187
9.5 Conclusion 189

Bibliography **191**

10 Paper V: Full-Reference Stereoscopic Image Quality Assessment account for Binocular Combination and Disparity Information **195**

10.1 Introduction 196
10.2 Proposed SIQA method 198
10.3 Experimental results and analysis 201
10.4 Conclusion 204

Bibliography **205**

11 Paper VI: No-Reference Quality Assessment of Stereoscopic Images based on Binocular Combination of Local Features Statistics **209**

11.1 Introduction 210
11.2 Proposed approach 211
11.3 Experimental results 214
11.4 Conclusion 217

Bibliography **219**

12 Paper VII: Stereoscopic Image Quality Assessment based on Monocular and Binocular Visual Properties **223**

12.1 Introduction 224
12.2 Related work 226
12.3 Proposed SIQA model 229
12.4 Experimental results 235
12.5 Conclusion and future work 251

Bibliography **253**

List of Figures

1.1 Overview of thesis outline and connection with the publications. Blue blocks denote the sections in Part I, whereas red blocks represent publications described in Part II. 7

2.1 Anatomy of the human eye [26]. 10

2.2 Flowchart of a typical psychophysical HVS model presented in [34]. 10

2.3 Example of linear perspective related monocular depth cue. (@ Yu FAN) 11

2.4 Example of aerial perspective related monocular depth cue. (@ Yu FAN) 11

2.5 Example of interposition-related monocular depth cue. (@ Yu FAN) 12

2.6 Example of relative size related monocular depth cue. (@ Yu FAN) 12

2.7 Example of texture gradient related monocular depth cue. (@ Yu FAN) 13

2.8 Example of light and shadow related monocular depth cue. (@ Yu FAN) 13

2.9 Example of height in the scene related monocular depth cue. (@ Yu FAN) 14

2.10 Example of defocus blur related monocular depth cue. (@ Yu FAN) 14

2.11 Example of motion parallax monocular depth cue. (@ Yu FAN) 15

2.12 Geometry of the binocular disparity. 16

2.13 Illustration of the binocular disparity in a stereopair from the Middlebury database [54]. 16

2.14 Illustration of the horopter and the Panum's fusional area. 17

2.15 Example of JND thresholds of the pixels in an image block. 22

2.16 A general framework of a computational 3D-JND model. 23
2.17 Methodology of performance comparison for 3D-JND models. 25
2.18 Flowchart of existing NR-IQA models. 31

3.1 Block diagram of the test procedure. 36
3.2 **Paper III**: stereo pair patterns used in psychophysical experiments. 40
3.3 **Paper IV**: framework of the proposed SIQA method. 41
3.4 **Paper V**: Flowchart of the proposed SIQA method. 42
3.5 **Paper VI**: framework of the proposed NR-SIQA model 43
3.6 **Paper VII**: framework of the proposed SIQA model. 45

4.1 Scatter distribution of predicted scores versus DMOS on LIVE 3D phase I. 54
4.2 Scatter distribution of predicted scores versus DMOS on LIVE 3D phase II. 55

8.2 (a) JND thresholds for difference background luminance levels L_b and disparities d from LA experiment, (b) JND thresholds for difference L_b and noise amplitudes of the left view N_l from CM experiment. (c) Average slopes of the two curves in (b) for each L_b. 173

List of Tables

3.1 **Paper II**: comparison between the 3D-JND models. 39

4.1 Subjective test scores: quality comparison between original 3D image and noisy 3D images produced by a 3D-JND model using 12 images from Middlebury Stereo Datasets. Note that DBJND and SSJND models are respectively our model without and with considering the visual saliency effect. The higher the average of the 3D-JND model is, the better the distortion masking ability the 3D-JND is. 52

4.2 Performance of SIQA methods on LIVE 3D IQA database (Phase I). Italicized entries denote 2D-based IQA, and the results of the best-performing SIQA method are highlighted in boldface. 53

4.3 Overall performance and performances for different types of distortion of the SIQA methods on LIVE 3D-IQA Phase II database. The ranking 1st and 2nd for each criterion are highlighted with red and blue bold texts, respectively. 53

4.4 SROCC values of the NR-SIQA methods on cross-database. 54

4.5 RMSE values of the NR-SIQA methods on cross-database. 54

List of Abbreviations

2D	two dimensional 3, 4, 18, 28–31, 34, 40, 41, 44, 46–51, 59, 61
2D-JND	two dimensional just noticeable difference 5, 25, 35, 38, 44, 45, 47, 49, 50, 56, 59
3D	three dimensional 3–5, 9, 15–20, 27, 28, 35, 37, 38, 40, 41, 43–52, 55–57, 59–61
3D-JND	stereoscopic three dimensional just noticeable difference 4–8, 35, 37, 38, 40, 44, 45, 47–52, 55, 56, 59, 60
BIQA	blind image quality assessment 30–32
BJND	binocular just noticeable difference 37, 45, 47, 49, 60
DBJND	disparity-aware binocular just noticeable difference 38, 49, 51
DMOS	differential mean opinion score 28, 30, 32, 33, 51
DpM	depth map 18, 20
DsM	disparity map 40, 49, 56
FR	full-reference 29, 30, 41, 42, 44, 49, 50, 59
GM	gradient magnitude 30, 31, 43–45
GMSM	gradient magnitude similarity mean 30, 44, 56
HVS	human visual system 4, 5, 9, 16, 17, 25, 29, 30, 37, 40, 41, 43, 47, 48, 50, 59
IQA	image quality assessment 7, 27–34, 37, 40, 41, 43, 45, 46, 49–52, 56, 57, 59
ITU	international telecommunication union 25, 28
JND	just noticeable difference 4, 5, 25–27, 30, 37, 38, 40, 44–46, 48–50, 55, 56, 59, 60
JP2K	JPEG 2000 51, 52, 57
LE	local entropy 44, 45, 50
LoG	laplacian of gaussian 9, 31, 43–45, 50, 57, 59
MOS	mean opinion score 28, 30, 32, 33
NN	neural network 30, 57
NOPs	non-occluded pixels 38, 60

NR no-reference 29–32, 43, 44, 49, 50, 52, 57, 59, 61
OPs occluded pixels 38, 49
PCC Pearson linear correlation coefficient 26, 33, 36
PSNR peak-signal-to-noise-ratio 29
QA quality assessment 4, 35, 48–50, 52, 56, 61
QoE quality of experience 3, 17, 61
RF random forest 30, 57
RMSE root-mean-square error 33, 36, 52
S3D stereoscopic three dimensional 3–5, 7, 17, 18, 49, 56
SIQA stereoscopic image quality assessment 4, 5, 7, 8, 27, 29–31, 34, 36, 40–52, 56, 57, 59–61
SROCC Spearman rank order correlation coefficient 26, 33, 36, 52
SVM support vector machine 30
SVQA stereoscopic video quality assessment 52, 61
SVR support vector regression 30–32, 43, 50, 57
UQI universal quality index 29, 40, 41, 49, 56
VIF visual information fidelity 41, 49, 56
VM visual masking 5, 25, 36, 47, 48, 56, 59, 60
WN white noise 51, 52

Part I

General Introduction

Chapter 1

Introduction

In this Chapter, we briefly give a general introduction of this thesis including the motivation, research objectives, research questions, research directions, research methodology, and thesis outline.

1.1 Motivation

With the rapid advances in hardware technologies, stereoscopic three dimensional (S3D) multimedia has made a great progress over the past few years, and has been increasingly applied in the fields of entertainment (e.g., three dimensional (3D) cinema [1], 3D-Television (TV) [2–4], 3D games [5, 6]), education [7], and medical imagery [8]. S3D technology improves the quality of experience (QoE) [9] by providing more realistic and immersive viewing experience compared to two dimensional (2D) image/video, thanks to binocular depth cues. Various instances show that 3D is flourishing:

- 3D-TV has been very well received by viewers through Channel 9 in Australia and the British Broadcasting Corporation (BBC) in the United Kingdom.
- In China, the number of 3D cinemas is exploding counting over 6000 3D screens nationwide and this number is growing daily.
- 3D content on the Internet is increasing as well. YouTube has over 15000 3D videos that can be watched.

Although 3D became popular thanks to the immersive feeling, the development of 3D technologies has also brought some technical challenges and inconvenience [10–13]. One can notice a slow-down and even a decrease of the numbers of 3D computer screens and especially 3D-TV sold in the last months [13]. This is mainly due to 3D-related issues generated at each stage from capture, compression, storage, transmission, to display.

At the capture stage, there are no-common rules [2] for creating a correct 3D content besides the numerous initiatives stemming from stereographers and not always leading to unanimity. At the stages of compression and storage, 3D content format conversion and coding may probably introduce some artifacts, which usually result in depth mismatch, texture information loss, and insufficient reality [3]. Therefore, the aforementioned artifacts and defects may cause visual symptoms to viewers such as eye strain, headache, nausea, and visual fatigue [14, 15]. Furthermore, the huge volumes of S3D data produced nowadays, make the storage capacity and compression efficiency more challenging. At the display stage, if the projected 3D content is not adapted to particular factors such as display size and technology, this can lead to visual discomfort [16, 17] and a significant decrease of the QoE.

1.2 Research Aims

All above-mentioned challenges and issues motivated us to focus the PhD research on two aspects. On the one hand, to effectively and efficiently compress 3D image, it is important to account for human visual system (HVS) characteristics and properties (e.g., visual sensitivity). In particular, our research aims to investigate the spatial visibility threshold based on both monocular and binocular visual properties. This threshold is usually referred to as the just noticeable difference (JND), which determines the maximum distortion undetectable by human eyes. Moreover, since an accurate stereoscopic three dimensional just noticeable difference (3D-JND) model can be applied in performance improvement of 3D image compression and quality assessment (QA), this research also aims to propose a reliable 3D-JND model based on study and comparison between state-of-the-art 3D-JND models. On the other hand, aiming to provide the promising viewing experience for 3D content, perceptual QA for stereoscopic images is quite crucial to evaluate or optimize the performance of S3D processing algorithms/systems. Therefore, the purpose of this research is to propose accurate and efficient stereoscopic image quality assessment (SIQA) methodologies based on the investigation of binocular perception. Specifically, the most important step is to find monocular and binocular factors affecting the perceptual quality of 3D images. In addition, we need to explore and model the binocular vision properties linked to the behavior of human 3D quality judgment. Finally, the SIQA models will be proposed combining the quality-related factors and considering the binocular vision properties.

1.3 Research Questions

Based on the aforementioned aims, this PhD research focuses on two main parts: spatial visibility thresholds of the binocular perception and SIQA. Each part leads to several questions described below:

1.3.1 Questions related to 3D-JND

Q1 Which characteristics and properties of the HVS should be taken into account for 2D and 3D digital imaging ? (see **Paper II** [18])

Q2 How are the performance of the state-of-the-art 3D-JND models developed based on HVS properties and characteristics ? What are the advantages, drawbacks, and applications of these models ? How to evaluate the performance of the 3D-JND models in order to select the most appropriate model for particular applications ? (see **Paper I** [19] and **Paper II** [18])

Q3 How to develop a new reliable 3D-JND model accounting for HVS visual masking (VM) effects and depth information ? How to design the psychophysical experiment modeling VM effects and binocular disparity ? How to construct a 3D-JND model based on psychophysical data ? (see **Paper III** [20])

1.3.2 Questions related to SIQA

Q4 How does the HVS judge image quality based on binocular perception ? (see **Paper VII** [21])

Q5 What are the most influential factors for 3D image quality and to which extent are they affected ? What are the binocular perception phenomena/effects ? And how do these effects impact the perceived quality of 3D images ? (see **Paper IV** [22], **Paper V** [23], and **Paper VI** [24])

Q6 What precise and reliable methodology for SIQA that accounts for both monocular and binocular influential factors ? And how do these factors affect jointly the overall 3D quality ? (see **Paper VII** [21])

The above-mentioned *research questions* will be answered and discussed in Chapter 4, and **Paper I** to **Paper VII** mentioned above are summarized in Chapter 3.

1.4 Research Direction

This Section presents the research direction based on the questions given in Section 1.3.

It is known that human subjects may capture left and right-eye images with different qualities due to the visual asymmetry state when observing the real world. Fortunately, the HVS has the ability to correct the acceptable quality distortions, and thus create a single cyclopean view to perceive the environment. Therefore, in order to mimic the human visual perception for S3D imaging systems, we need to figure out (1) when the HVS corrects the quality difference between left and right views, (2) what is the difference threshold below which the overall 3D image quality and depth perception are guaranteed, and (3) how to reproduce the brain visual behavior regarding binocular depth cues and image quality using perceptual models ?

To deal with above-mentioned issues, we need to understand the HVS sensitivity to inter-views difference. Specifically, we propose to exploit the notion of JND to reflect the minimum changes in image's pixel that the HVS can detect. Therefore, state-of-the-art the two dimensional just noticeable difference (2D-JND) and 3D-JND models should be reviewed and studied to propose a new and reliable 3D-JND model based on visual masking effects considered in these models. This model may probably be designed based on psychophysical experiment to measure the visibility thresholds of the asymmetric noise in the stereopair. These psychophysical experiments can also help to deeply understand the behaviors/effects of binocular perception including binocular fusion, binocular rivalry, and binocular suppression. Some bio-inspired models mimicking binocular visual behavior should be explored. These perceptual models can be used in SIQA to accurately predict the overall 3D quality because it depends not only on single-view image quality or depth quality but also on the experience of binocular visual perception. As described by [25], the visual sensitivity reveals a perceptual impact of artifacts according to the spatial characteristics. Moreover, the visual sensitivity explains the tolerance of the HVS to changes/difference of pixel values in image regions, and is related to visual attention. Therefore, the visual sensitivity is proportional to visual salience and inversely proportional to JND. Integrating the models representing the visual sensitivity in the process of perceptual IQA may allow being closer to the human quality judgment as it mimics the HVS behavior.

1.5 Research Methodology

This section presents the research methodology I undertake for this thesis. In general, the research methodology used in this thesis is based on both theoretical and empirical studies. According to two research topics addressed in the thesis, the included papers are divided into **Paper I** – **Paper III** (related to 3D-JND) and **Paper IV** – **Paper VII** (related to quality assessment).

Paper I – **Paper II** firstly conduct a theoretical study (i.e., literature review) to describe and analyze the existing models, and compare them in terms of their applicability, pros, and cons. Then both papers are based on an empirical study to compare the models thanks to qualitative and quantitative experimental analysis. In addition, **Paper II** conducts the experiments using the data that we either create or obtain data from publicly available datasets. In general, the work in **Paper I** and **Paper II** employs a deductive research methodology as they provide a theoretical overview and analysis, then an experimental evaluation and comparison. In contrast, **Paper III** designs a new model using the experimental data collected from psychophysical experiments. Thus, this paper refers to the inductive study and mainly focuses on an empirical study. In addition, **Paper III** analyzes the data using some statistical methods, and then conducts a subjective evaluation to validate the proposed model.

The research in **Paper III** – **Paper VII** are deductive because they propose the quality metrics based on the binocular perception theory and hypothesis. Furthermore, quantitative experiments are used to evaluate the metrics performance. Therefore, **Paper III** – **Paper VII** are more empirically focused. In addition, in order to compare with other existing quality metrics, these papers conduct

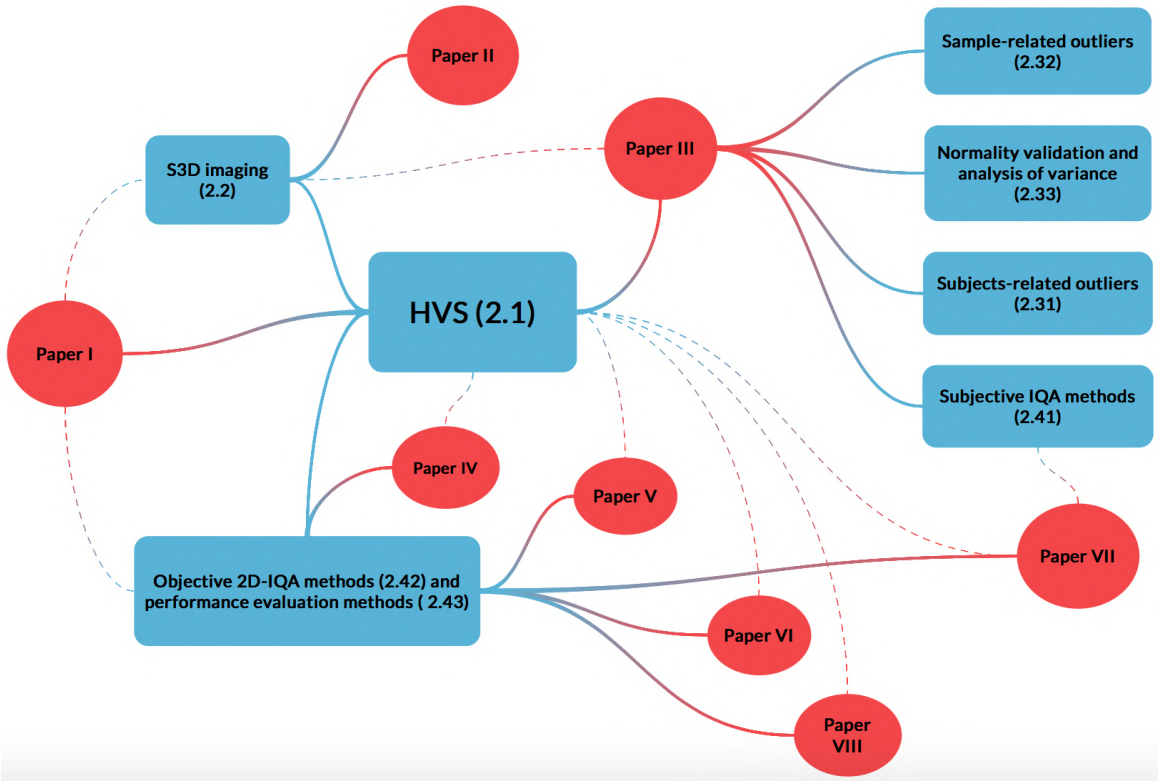


Figure 1.1 – Overview of thesis outline and connection with the publications. Blue blocks denote the sections in Part I, whereas red blocks represent publications described in Part II.

primary and secondary analysis of data. Primary data refer to the results derived from our experiments, whereas secondary data are the data collected by other researchers. The secondary data is employed because the source code of the algorithms are not publicly available and the implementation is costly.

1.6 Dissertation Outline

This dissertation contains two parts: **Part I** and **Part II**. Part I is organized as follows:

- A general introduction including PhD research motivation, objectives, directions as well as a brief organization of this dissertation.
- A research background introducing some necessary state-of-the-art knowledge with respect to human interaction with S3D technology, statistical analysis tools for psychophysical experiments and image quality assessment (IQA). This Chapter aims to provide the reader with the necessary background to understand the content of Part II.
- A summary of each included paper issue from the PhD study. In particular, we briefly present for each paper the rationale, the framework, the results, and the major contributions.

- A discussion including the thesis contributions to 3D-JND models and SIQA approaches, the personal contribution for each paper, and some remaining issues and challenges.
- A general conclusion about the conducted work and some perspectives.

Part II consists of the published and submitted papers during the PhD studies demonstrating the contributions of this work. **Paper I** – **Paper III** refer to study on 3D-JND, whereas **Paper IV** – **Paper VII** are related to SIQA. Figure 1.1 depicts the overview of this thesis organization including the background sections of Part I, papers described in Part II, and the connection between both parts.

Chapter 2

Background

This Chapter introduces the background of this doctoral thesis that is closely related to our work, including (1) human visual system, (2) stereoscopic 3D imaging, (2) visual psychophysics and just noticeable difference,(4) statistical tools for psychophysical experiments and (5) perceptual quality assessment of 2D and 3D images. Note that this Chapter aims to briefly provide some basic knowledge of relevant topics to ensure a better understanding to the readers.

2.1 Human visual system

HVS is a biological and psychological mechanism used to perform image processing tasks, and mainly includes brain, the nervous pathways and the eyes [27]. HVS models are often used to simplify the complex visual behaviors by simulating its characteristics and properties. In this section, we briefly describe the HVS biological organization and a general 3D model.

The human eye, represented on Figure 2.1, is used to capture the image corresponding to real-world scenes. It consists of the cornea, aqueous humor, iris, lens, vitreous humor and retina that the light passes through respectively [28]. Once the images are captured by the eyes, the images are projected onto the retina located at their back of the eyes. The retina contains photoreceptors, bipolar cells, horizontal cells, amacrine cells and ganglion cells [29]. After preprocessing the image in the retina, the visual information is carried by the optic nerve from the ganglion cells to lateral geniculate nucleus, and finally to the visual cortex. Several previous studies show that lateral geniculate nucleus is significantly important for visual perception and processing [30–33]. For instance, binocular rivalry behavior of the HVS is correlated to lateral geniculate nucleus [32, 33]. This inspires us to use image laplacian of gaussian (LoG) response to simulate the binocular rivalry (see **Paper VI** and **Paper VII**). Finally, the visual cortex receives the visual information from both eyes and combines it to see the real-world

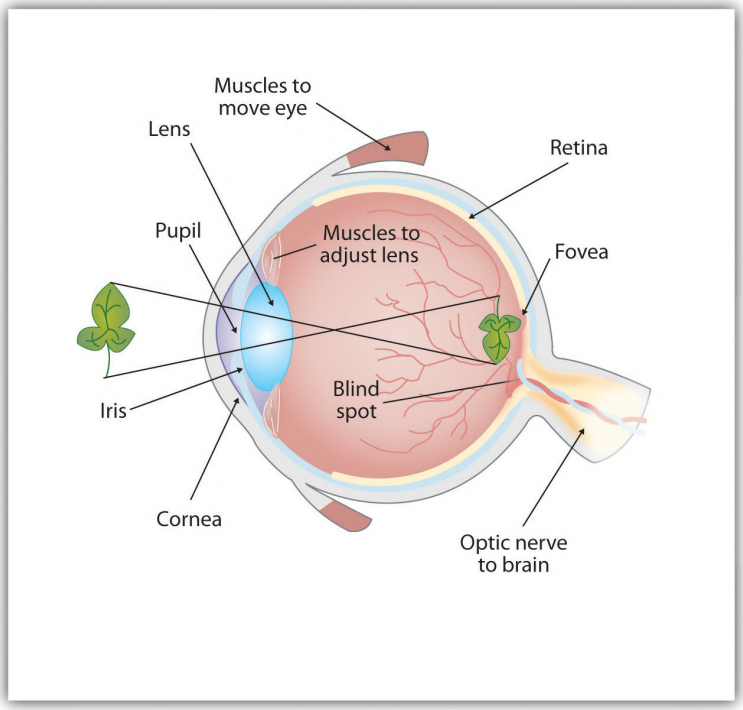


Figure 2.1 – Anatomy of the human eye [26].

in 3D. A typical psychophysical HVS model is shown in Figure 2.2. The main idea is to develop a HVS model considering its characteristics and properties including color information, contrast sensitivity, and spatial masking effects. The reader may refer to [34] for more details on this HVS model.

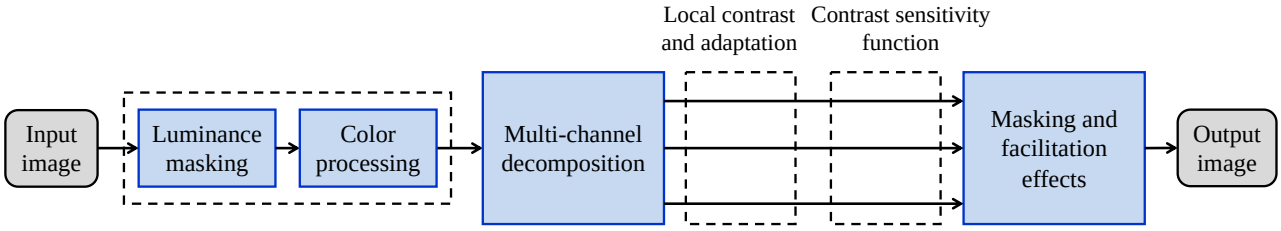


Figure 2.2 – Flowchart of a typical psychophysical HVS model presented in [34].

2.1.1 Depth perception

HVS can perceive the environment in three dimensions, and can be able to evaluate relative distances (i.e., depth) of objects thanks to depth perception [35–37]. Depth perception relies on various depth cues [38]. Since the relative distance can be perceived by the human with one eye or both eyes, depth cues are classified into two categories: monocular and binocular depth cues.

Monocular depth cues are represented in two dimensions and are seen with just one eye. According

to [39, 40], monocular depth cues mainly include:

- Linear perspective: creates an illusion of depth on a flat surface [39]. Two parallel lines farther to the viewer result in more converging to the vanishing point. An example is shown in Figure 2.3.



Figure 2.3 – Example of linear perspective related monocular depth cue. (@ Yu FAN)

- Aerial perspective: describes the depths of the objects based on the degradation of light luminance and color on the scene caused by atmospheric phenomena (e.g., fog, dust) [41]. For instance, objects that are far away from the viewer have hazy edges, lower luminance contrast and color saturation as shown in Figure 2.4.



Figure 2.4 – Example of aerial perspective related monocular depth cue. (@ Yu FAN)

- Interposition: estimates the relative distance of the objects by showing the objects occluding each other [42]. Specifically, a farther object is usually partially occluded by a nearer object. Figure 2.5 shows an example of interposition.



Figure 2.5 – Example of interposition-related monocular depth cue. (@ Yu FAN)

- Relative size: evaluates the relative depth of the physically identical objects by comparing their sizes [43]. In particular, an object with a smaller size in the scene is farther away for the viewer than the same object with a larger size. An example of wine glasses is shown in Figure 2.6.



Figure 2.6 – Example of relative size related monocular depth cue. (@ Yu FAN)

- Texture gradient: provides the depth of the objects depending on their textural strength [44]. Compared to farther away objects, nearer objects have finer and sharper details of texture. Figure 2.7 shows an example of lavenders at different distances.



Figure 2.7 – Example of texture gradient related monocular depth cue. (@ Yu FAN)

- Light and shadow: in the scene, they can jointly reflect the relative depth between objects [45]. For instance, the objects in the shadow are farther from the light source than those out of shadow (see Figure 2.8).



Figure 2.8 – Example of light and shadow related monocular depth cue. (@ Yu FAN)

- Height in the image plane: refers to vertical positions of the objects in the scene plane. In the scene, farther objects appear higher up than nearer ones. An example of this monocular depth cue is shown in Figure 2.9.



Figure 2.9 – Example of height in the scene related monocular depth cue. (@ Yu FAN)

- Defocus blur: It is related to depth of focus of the HVS [46]. Objects farther away from the viewer are generally hazier and smoother, which is exhibited in Figure 2.10.



Figure 2.10 – Example of defocus blur related monocular depth cue. (@ Yu FAN)

- Motion parallax: It is an effect where the objects closer to the viewer appear to move faster

than farther objects [47]. Thus, this depth cue exists in a dynamic scene. Figure 2.11 shows an example about motion parallax.

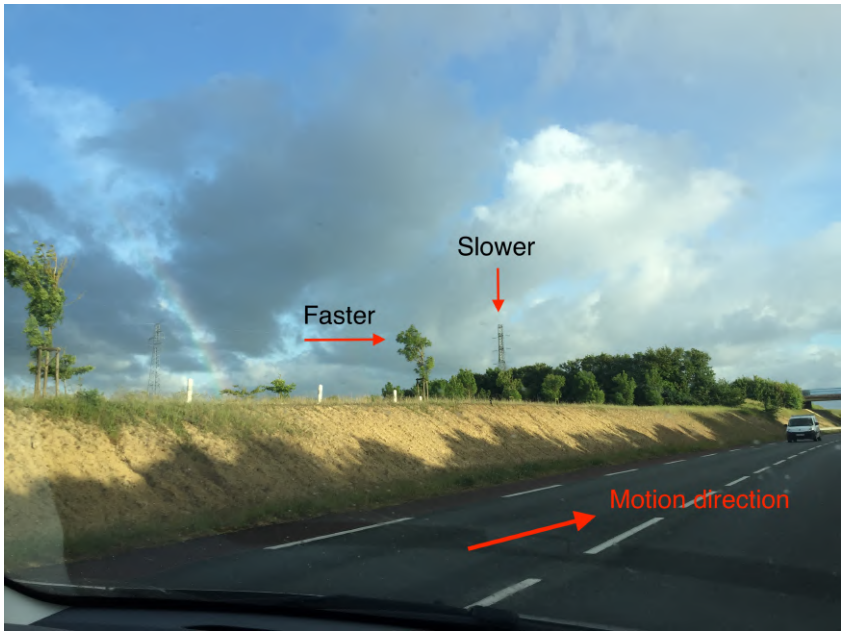


Figure 2.11 – Example of motion parallax monocular depth cue. (@ Yu FAN)

- Accommodation: It is an oculomotor cue concerning the change of lens shape of the eyes. This change is controlled by the ciliary muscles so as to adjust the focal length [48]. Accordingly, contraction/relaxing strength of ciliary muscles can reflect the distance of the objects to the fixation point.

Binocular depth cues provide depth information when observing a scene with both eyes. There are two categories of binocular cues: binocular disparity and vergence [49, 50]. Human eyes are separated by a distance of approximately 63 mm [51], which is related to parallax. Thus, for an object in the scene, both eyes can receive two similar but slightly different retinal images in terms of object's position. This difference is defined as binocular disparity or binocular parallax [52]. As shown in Figure 2.12, the binocular disparity can be represented by either the angle difference $\beta - \alpha$ or horizontal shift. Another example of disparity for an object between left and right views of a stereopair is depicted in Figure 2.13. To estimate the object's distance from viewers, human eyes extract the 3D information from texture retinal images via the binocular disparity [37, 53]. This effect is named stereopsis and is generally used in the creation of 3D images/videos with specific display and/or glasses.

Another binocular depth cue is the vergence, reflecting the simultaneous movement of both eyes in opposite directions to obtain/maintain a single binocular vision [55, 56]. Specifically, eyes rotate inward when viewing a nearer object, and outward when viewing a farther object. In fact, vergence and accommodation interact intimately and cannot be separated [57–59].

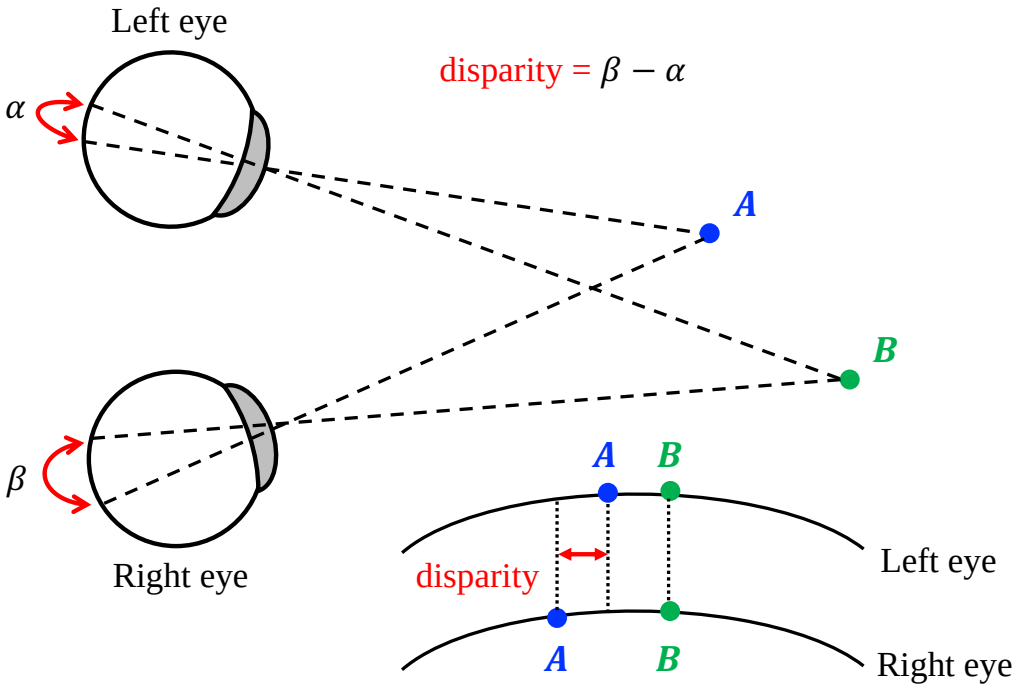


Figure 2.12 – Geometry of the binocular disparity.



Figure 2.13 – Illustration of the binocular disparity in a stereopair from the Middlebury database [54].

In addition, the objects have the same disparity if they are positioned on the corresponding locus of points in space, which is defined as the horopter [60]. In particular, horopter represents the points having the same distance from the viewer as objects of focus.

Our brain can fuse the left and right retinal images of an object into a single 3D mental image called cyclopean image [61], if this object is included in a limited area behind or in front of the horopter. This area is called Panum's fusional area [62, 63]. More details about Panum's area can be found in [64]. Figure 2.14 illustrates the horopter and Panum's area. In contrast, the image of an object outside

of this area is usually blurry, and thus results in binocular rivalry phenomenon due to the conflict between accommodation and vergence [65, 66]. Binocular fusion behavior of the HVS occurs if the left and right views are similar in terms of content or quality [67]. In contrast, if two views are different (in terms of content or quality), and this difference is relatively small, binocular suppression occurs, one of the views dominates completely the field [68]. If this difference is relatively large, binocular rivalry occurs. Two images are seen alternately, one image dominating for a moment [65, 69].

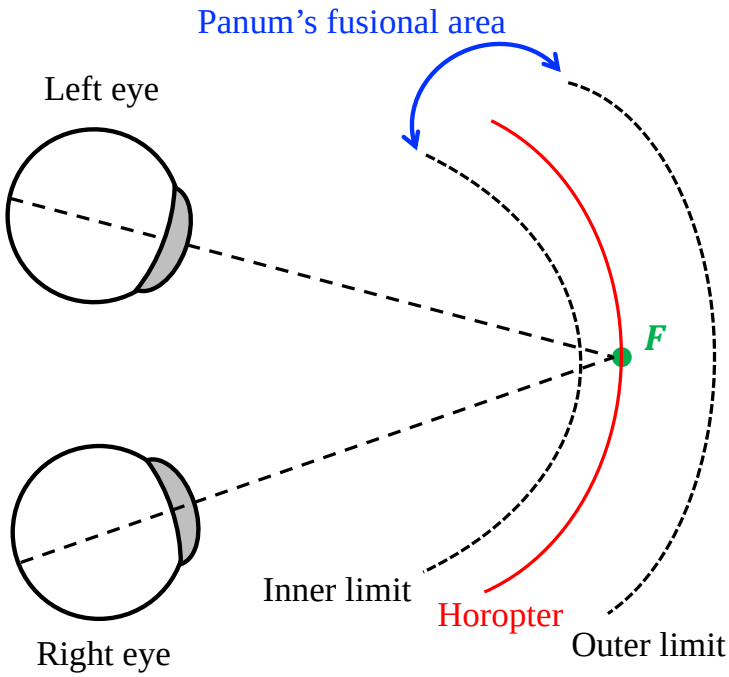


Figure 2.14 – Illustration of the horopter and the Panum's fusional area.

2.2 Stereoscopic 3D imaging

The HVS can perceive the environment in 3D thanks to binocular depth cues from binocular vision. The binocular depth perception is created based on slightly different positions of two retinal images seen by left and right eyes. This difference of positions is called binocular disparity. In other words, the presentation of slightly different images to the left and right eyes leads to the perception of depth. This inspires the development of 3D technology. In this section, we introduce different 3D content representation formats and 3D displays [70].

2.2.1 S3D content representation formats

According to [71–74], there exist several S3D representation formats including: stereo interleaving, 2D-plus-depth, layered depth video, multi-view video plus depth, and depth-enhanced formats.

Stereo interleaving format is a conventional 3D image/video format that is a multiplex of left-view and right-view images into a single image/frame. The multiplexing downsamples and places left and right images horizontally (i.e., side-by-side) or vertically (i,e., top-and-bottom). Side-by-side format was used in our psychophysical experiments (See **Paper II** and **Paper III**). 3D video with this format can match most existing 3D codecs and displays. However, stereo interleaving format can only deliver half of the image resolution and thus reduces the QoE.

2D-plus-depth format contains a single view/frame (i.e., color image) and a depth map (DpM) for 3D image/video representation [74]. This format allows people to adjust the perceived depth values easily depending on the viewing distance and display size. However, the 2D-plus-depth format may lead to occlusion issues due to mismatches between views, and provides a limited rendering of the depth range.

To overcome the constraints of 2D-plus-depth format, the layered depth format was proposed [71, 75, 76]. In addition to the texture image and DpM, this format provides a background layer with the corresponding DpM in order to improve the occlusion information and improve the quality of depth-image-based rendering. The background layer contains image content that is covered by foreground objects in the main layer. The layered depth format allows to generate the new view points for stereoscopic and auto-stereoscopic multi-view displays.

Multi-view video plus depth format uses several cameras to capture the scene from different view points [77, 78]. Therefore, This format includes multiple color images and depth maps. The depth maps can be estimated from different views. Multi-view video plus depth format allows to control the depth range based on the distance between the selected views from the multi-view arrangement. However, this format requires large storage capacity and transmission bandwidth [79].

Depth-enhanced format includes two views with high quality, DpM and occlusion layers [80]. This format provides backward compatibility and extended functionalities such as baseline adaptation and depth-based view synthesis [71].

2.2.2 S3D displays

In order to visualize 3D images/videos, a S3D-enabled display is needed. Human subjects can perceive the spatial relationship between objects thanks to various cues including monocular and binocular cues [81]. Thus, a design of 3D displays should consider the contribution of monocular depth cues in addition to binocular depth cues in order to provide a basic visual performance as a standard 2D display and 3D sensations provided by the stereoscopic cues. Several overviews of existing 3D display technologies are provided in [77, 82–85].

According to the used technology (e.g., glasses or head-mounted devices), 3D displays can be classified into three categories: stereoscopic, autostereoscopic, and head-mounted displays [82, 83, 86].

Stereoscopic displays are the visualization terminals that demand the observer to wear an optical device to direct the left and right images to the appropriate eye. Stereoscopic displays can be further

divided into:

- time-parallel displays that show left and right views of a stereopair simultaneously on the screen. In such displays, the visualization methods include: (1) color multiplexing that presents left and right images in different colors [87] ; (2) polarization multiplexing [88]. The latter uses polarized filter glasses to split left and right views ensuring that two eyes perceive different images. Linear or circular polarization can be used. The latter allows more head tilt for viewers.

  In our psychophysical experiments (described in **Paper II** and **Paper III**), the Hyundai TriDef S465D display is used to convert side-by-side 3D format to 3D image by circular polarized glasses.

- time-multiplexed displays that present left and right views alternately with a high frequency. Such technology generally uses active shutter glasses synchronized with the display system via an infrared link [89]. The idea is that the viewer can perceive neither the switching between left and right views nor related flicker if both views switch quickly enough (i.e., with a frequency greater than 58 Hz) [90]. Thus, based on the synchronization signal from the display, the active shutter glasses open alternately the image for one eye while closing the other.

Autostereoscopic displays do not require any glasses to present two-view images, but send them directly to the corresponding eyes using aligned optical elements on the surface of the displays [83]. This category of displays simplifies 3D perception for viewers and can show multiple views to them, which makes the 3D entertainment more applicable. In addition, autostereoscopic displays can present each view from a particular viewing angle along the horizontal direction and provide a comfortable viewing zone for each stereopair. Such displays can be classified into:

- binocular or multi-view based displays [91, 92]. Binocular views based displays only present a stereopair, whereas multi-view displays provide multiple views of the scene to several viewers based on light sources of different paths. The light paths can be controlled by special optical elements including for example parallax barriers, lenticular lens arrays, micropolarizers, etc.

- head-tracked displays allowing the viewers to change the viewing position by using active optics to track their head/pupil positions [82].

- volumetric displays generating the images by projection within a volume space instead of a surface in space. The image volume consists usually of voxels [93]. Each voxel on a 3D image is located physically at the supposed spatial position, and reflects omnidirectionally the light from that position to present a real image to viewers.

- holographic displays showing real and virtual images based on wave-front reconstruction. Such displays do not require any special glasses or external equipment to view the image. Such displays use holographic optical elements (e.g., lens, films, and beam splitter) to construct their projection screen [94, 95].

In addition to stereoscopic and autostereoscopic displays, head-mounted 3D displays are another and new way to present 3D images [86]. In particular, such displays require viewers to wear a particular head-tracking device containing sensors that record viewers' spatial movement information. Head-mounted displays can provide the viewers with a deep feeling of immersion. Such displays have been widely advanced and applied in entertainment, education, and medical domains with virtual and augmented reality applications [96–98].

2.2.3 Depth and binocular disparity

With the aim to ease the understanding of the used depth/disparity values described in **Paper II**, this section describes the relationship between the depth values of 3D scene z (in meters), real-world depth map (DpM) values (in pixels) and binocular disparity values (in pixels or meters).

DpM denotes the distance of the scene's objects from the viewpoint of the camera or the viewer. In other words, DpM represents a measure of the distance between objects in the image. DsM (Disparity map) refers to an image containing the distance values between two corresponding pixels in the left and right views of the stereo pair. A DpM map can be obtained using a 2D image, while a DsM map is only obtained using a stereopair. Note that the disparity value can be converted to depth value based on some specific formula and vice versa.

Typically, DpM and DsM is a 8-bit gray scale image. This map represents closer and farther objects with regards to the fixation plan by larger and smaller values, respectively. The depth image values vary between 0 and 255. Accordingly, the closest and farthest objects to the fixation point are shown as white and black pixels in depth images. Using the actual depth value z, the depth in pixel of DpM z_p is determined as follows:

$$z_p = \left\lfloor 255 \cdot \frac{\left(\frac{1}{z} - \frac{1}{z_{max}}\right)}{\left(\frac{1}{z_{min}} - \frac{1}{z_{max}}\right)} \right\rfloor, \tag{2.1}$$

where z_{min} and z_{max} respectively denote the minimal (closest objects) and maximal (farthest objects) distances in the scene to the camera/observer. Both distances are usually given by 3D content maker. In addition, $\lfloor \cdot \rfloor$ rounds the number to the lower integer. Based on Equation 2.1, the actual depth value z can be obtained from the DpM value z_p using the following equation:

$$z = \left[\frac{z_p}{255} \cdot \left(\frac{1}{z_{min}} - \frac{1}{z_{max}} \right) + \frac{1}{z_{max}} \right]^{-1}. \tag{2.2}$$

Besides, disparity values d are frequently used to synthesize the cyclopean view for stereoscopic image quality assessment. The disparity values of a scene d are obtained by converting the depth value as

follows:

$$\begin{aligned} d &= \frac{f \cdot b}{z} \\ &= f \cdot b \cdot \left[\frac{z_p}{255} \cdot \left(\frac{1}{z_{min}} - \frac{1}{z_{max}} \right) + \frac{1}{z_{max}} \right], \end{aligned} \tag{2.3}$$

where b represents the distance between the stereoscopic cameras (i.e., baseline) or interpupillary distance (approximately 63mm) for viewers. Disparity values are in pixels/meter depending on the camera focal length f in pixels/meter.

2.3 Psychophysics and just noticeable difference

This section presents some basic concepts of visual psychophysics and notion of the just noticeable difference with respect to 2D and 3D images. In addition, we briefly introduce the psychophysical experiments used in our JND measurements and modeling.

2.3.1 Visual psychophysics

Visual psychophysics quantitatively study the relationship between the physical stimuli and human sensations and perceptions [99]. Specifically, visual psychophysics refers to experimental methodologies that can be applied to study how to model the human visual process. The empirical laws of psychophysics are based on sensory threshold (i.e., visibility threshold) measurement by conducting psychophysical experiments [100]. Such a threshold corresponds to the point of intensity at which the observers can just detect the presence of a stimulus or the presence of a change between two stimuli.

According to [100], the visibility thresholds are classified into two types as described below.

- Absolute threshold: is the minimum intensity of a stimulus that subjects can detect at least 50% of the time.
- Difference threshold: is the minimum change/difference in intensity between two stimuli that subjects can detect usually 50% of the time. Thus this threshold is also called just noticeable difference (JND) threshold. To measure the JND threshold, a pair of stimuli is usually presented to subjects. One stimulus has a standard intensity and is considered thus as a reference. For the other stimulus, subjects vary its intensity until they can just barely inform that this stimulus is either more intense or less intense than the reference one.

Note that the visibility thresholds measurement presented in **Paper II** and **Paper III**) refers to JND thresholds. In fact, [100] states that absolute and difference thresholds are occasionally considered similar in principle because there is always background noise altering observers ability to notice stimuli.

2.3.2 Just noticeable difference threshold

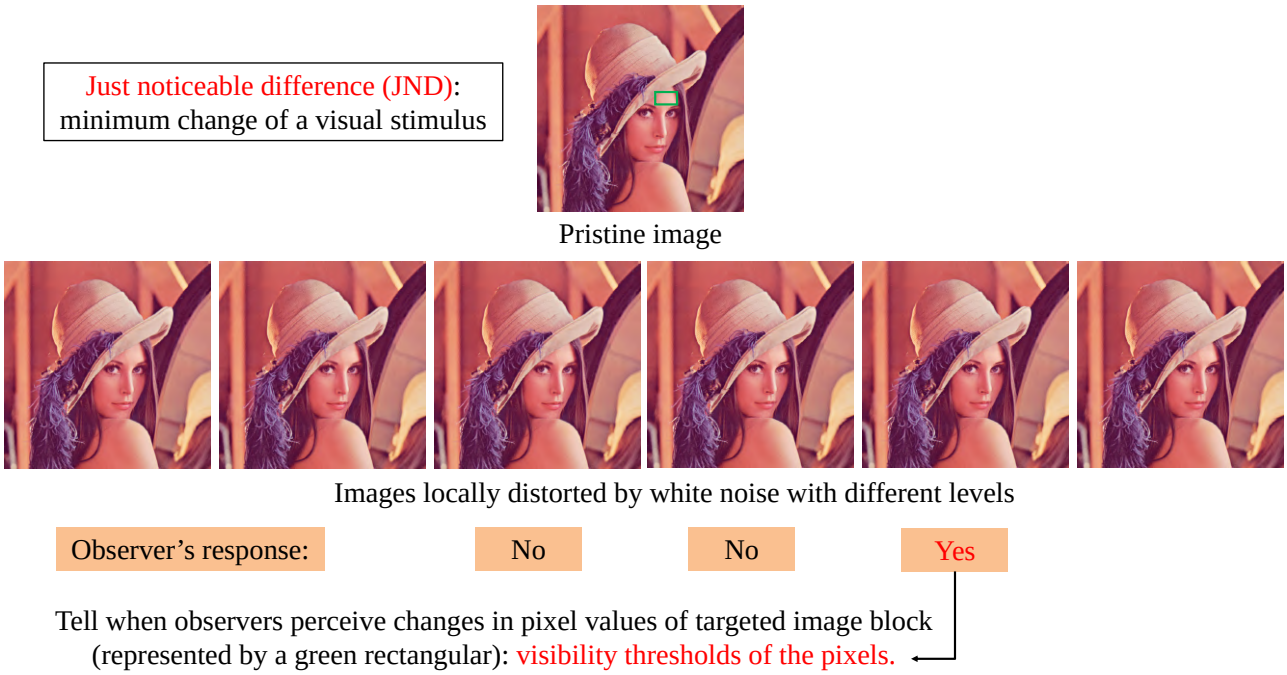


Figure 2.15 – Example of JND thresholds of the pixels in an image block.

For a 2D image, the JND of a pixel represents the visibility threshold at which human subjects are able to detect changes in pixel values. In other words, the 2D-JND reflects the maximum tolerable changes in pixel intensity to HVS. Figure 2.15 illustrates the procedure of obtaining the JND thresholds of the pixels in a 2D image block. The subjects compare the intensity difference between the pristine image and each distorted image in the targeted block (i.e., noised region). The noise level is gradually increased until the subjects report that they just detect the difference. This difference corresponds to JND thresholds of the pixels in the block.

To date, numerous 2D-JND models have been proposed by modeling contrast sensitivity, visual masking effects (e.g., luminance adaptation and contrast masking), and spatial frequency of the image local regions [101]. Any changes in the targeted image are undetectable by the HVS if they are lower than the JND threshold. Therefore, the 2D-JND models have been successfully applied to improve the algorithms of image coding, image quality assessment and enhancement. A 3D-JND model usually estimates the maximum changes in the image region that can be introduced in one view of the stereopair without causing binocularly visible differences, given the changes in the corresponding region of the other view. Therefore, as shown in Figure 2.16, in addition to monocular visual masking, the design of a 3D-JND model definitely considers the binocular vision properties (e.g., binocular masking, depth masking, and depth information). The readers can refer to **Paper III** for more details on 3D-JND models.

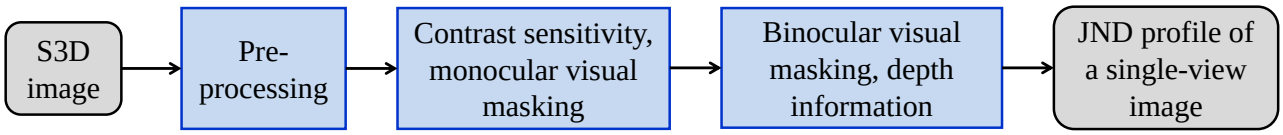


Figure 2.16 – A general framework of a computational 3D-JND model.

2.3.3 Psychophysical experiments

To measure JND thresholds, a general methodology is to conduct the psychophysical or psycho-visual experiments to determine whether the subjects can detect a stimulus, or notice the difference between two stimuli. As described in [100], psychophysical experimental methods used in JND measurements are:

- **Method of limits**: For ascending method of limits, given a stimulus (as the reference) with the property of a constant level, the property of the targeted stimulus starts out at a low level so that the difference between this stimulus and reference one could be undetectable. Then, this level is gradually increased until the subjects report that they are aware of the difference. The JND threshold is defined as the difference level of the stimulus property for which the targeted stimulus is just detected. In the experiments, the ascending and descending methods are used alternately, and the final JND threshold is obtained by averaging the thresholds of both methods. These methods are efficient because the subjects can obtain a threshold in a small number of trials. In addition, one do not need to know where the threshold is at the beginning of the experiments. Nevertheless, two disadvantages may be observed. First, the subjects may get accustomed to informing that they detect a stimulus and may continue reporting the same way even beyond the real threshold (i.e., errors of habituation). Second, subjects may expect that the stimulus will become detectable or undetectable, and thus make a premature judgment (i.e., the error of expectation).

  To overcome these potential shortcomings, a staircase (i.e.,up-and-down) method is introduced [102]. It usually starts with a stimulus having a high intensity, which is obviously detected by subjects. They give the response ('yes' or 'no'). Then, the intensity is reduced until the subjects response changes. After that, a reversal procedure starts and the intensity is increased (with the response 'no') until the response changes to 'yes'. Next, another reversal procedure is repeated until a given reversal number is reached. The JND threshold is then estimated by averaging the values of the transition points (i.e., reversal points).

- **Method of constant stimuli**: In this method, a constant comparison stimulus with each of the varied levels (ranging near the threshold) is presented repeatedly in a random order to subjects. The proportion of times causing the difference threshold is recorded. The stimulus level yielding a discrimination response in 50% of the time is considered as the JND threshold. The subjects can not predict the level of the next stimulus in the experiment. Therefore, the advantages of

this method are to reduce errors of habituation and expectation probably caused in the method of limits. However, this method is costly because it takes a lot of trials to pre-define the levels of the stimuli before conducting the experiment. In addition, the experiment using this method is time-consuming because each stimulus needs to be presented to subjects several times. It may probably reduce the accuracy of threshold measurement due to visual fatigue during the experiment.

- **Method of adjustment**: In this method, a reference stimulus with a standard level is presented to subjects. They are asked to adjust the level of the targeted stimulus, and instructed to alter it until it is the same as the level of the reference stimulus. The error level between the targeted stimulus and the reference one is recorded after each adjustment. The procedure is repeated several times. The mean of the error values for all adjustments is considered as the JND threshold. The method of adjustment is easy to implement but produces probably some unreliable results.

In **Paper II**, we used the ascending method of limits to measure the JND thresholds of the test image by comparing the reference image. Note that the stimulus corresponds to the 3D images altered by different types of distortion. The staircase method was not used because the experiment for each subject may last a long time in the case of numerous stimuli. More details about experiments could be found in **Paper II**.

For modeling JND, the staircase method is more suitable and efficient than the method of limits. Therefore, the staircase method (reversal 4) was used for detecting the just noticeable noises in the psychophysical experiments of **Paper III**. More specifically, given the noise amplitude in the left view of a stereopair (i.e., reference stimulus) at different disparities and background luminance or contrast intensities, a subject adjusts the noise amplitude in the right view with ascending and then descending orders to make the noise just detectable or undetectable. The reversal procedure is repeated twice considering the trade-off between the measurement accuracy and experiments duration. The JND thresholds of the right view are estimated by averaging the difference in noise amplitude between the reference and the test stimuli at the four reversal points. More details could be found in **Paper III**.

To compare the performance of state-of-the-art 3D-JND models, on the one hand, we evaluate the accuracy of the visibility thresholds estimation with the models (see Figure 2.17). In **Paper II**, we determine an interval of JND thresholds obtained from psychophysical experiments. Meanwhile, we estimate the JND thresholds using a 3D-JND model. Next, we verify whether the estimated JND threshold of each pixel is included in the corresponding interval. Finally, the percentage of accuracy of a 3D-JND model is obtained by dividing the number of pixels in the JND map included in the corresponding intervals by all pixels. On the other hand, we evaluate the performance of the image processing algorithms embedding 3D-JND. As shown in Figure 2.17, the 3D-JND can be used in depth or sharpness enhancement of 3D images [103, 104], to reduce the bit rate for 3D video coding [105], or to improve the prediction accuracy of the SIQA model [106]. The main idea is to evaluate the

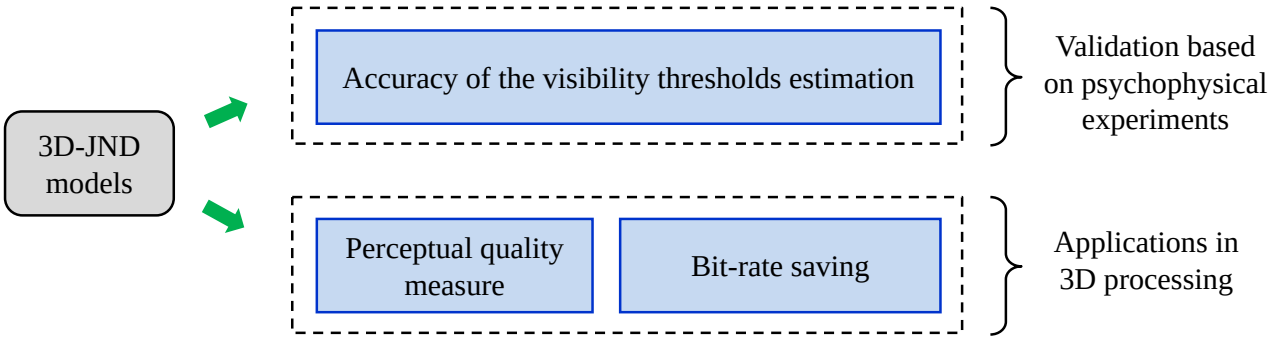


Figure 2.17 – Methodology of performance comparison for 3D-JND models.

perceptual quality and/or coding efficiency of processing algorithms embedding 3D-JND.

In **Paper III**, the proposed 3D-JND model is validated based on a subjective test and compared with other models in terms of perceptual quality at the same noise level. According to ITU-R BT.2021-1 [107], a variant of paired comparison (i.e.,stimulus-comparison) method is used in the subjective test. In particular, we use the adjectival categorical judgment method [108]. In this method, subjects assign an image to one of the categories that are typically defined in semantic terms. The categories may reflect the existence of perceptible differences (e.g., same or different), the existence and direction of perceptible differences (e.g., more, same, less). In the subjective test of **Paper III**, the subjects are shown two 3D images (injected with different noise levels) at a time, randomly arranged side by side. Then, they are asked to compare the left and the right views in terms of perceptual quality and provide a score depending on the comparison scale: 0 (the same), 1 (slightly better), 2 (better), 3 (much better), -1 (slightly worse), -2 (worse), -3 (much worse). The scores provided by all subjects for each stimulus are averaged to compute the mean opinion scores. These scores are then analyzed and processed with some methods described in the following section.

2.4 Statistical tools for psychophysical experiments

The visibility thresholds of the HVS can be determined by JND models, which are developed considering VM effects and visual sensitivity [109–119]. A comprehensive review of VM and 2D-JND models was given in **Paper II**. More details related to 2D-JND models can be found in [101].

In this section, we introduce some statistical methods for analyzing the data from psychophysical experiment or subjective test described in **Paper III**. An overview of the statistical tools used in image/video processing and computer vision can be found in [120]. To construct a reliable JND model using psychophysical data, we need to detect and remove the outliers related to the subjects and to the samples of each subject [121]. Some statistical methods of outliers detection have been introduced and reviewed in [122, 123].

2.4.1 Subjects-related outliers

To detect and remove unreliable subjects, a subject screening technique defined by the international telecommunication union (ITU) under the technical report BT 1788 [124] was used in **Paper III**. In particular, a subject is identified as an outlier if its corresponding correlation coefficient C_{min} is less than a rejection threshold.

More specifically, Let T_j^i denoting the JND threshold of the jth stimulus from i^{th} subject, where $i \in [1, N]$, and $j \in [1, M]$. M and N are the numbers of stimuli and subjects, respectively. Thereby the JND values from the ith subject T^i are defined as follows:

$$T^i = (T_1^i, T_2^i, \cdots, T_M^i), \tag{2.4}$$

and the JND values corresponding to the jth stimulus of all subjects are given by:

$$T_j = (T_j^1, T_j^2, \cdots, T_j^N). \tag{2.5}$$

Then, we define the mean vector $\overline{T}$ across all subjects as:

$$\overline{T} = (\overline{T_1}, \overline{T_2}, \cdots, \overline{T_j}, \cdots, \overline{T_M}), \quad \overline{T_j} = \frac{1}{N}\sum_{i=1}^{N} T_j^i. \tag{2.6}$$

Subsequently, we calculate the correlation coefficient between T^i and $\overline{T}$ using Pearson linear correlation coefficient (PCC) and Spearman rank order correlation coefficient (SROCC), respectively. Smaller values of $PCC(T^i, \overline{T})$ and $SROCC(T^i, \overline{T})$ are considered as C_{min}^i. The subject i is identified as an outlier and discarded if the corresponding C_{min}^i is less than a rejection threshold RT. The latter is calculated by:

$$RT = \begin{cases} C, & if\ C <= MCT \\ MCT, & otherwise, \end{cases} \tag{2.7}$$

where

$$C = \left| mean(C_{min}) - std(C_{min}) \right|, \tag{2.8}$$

where MCT denotes the maximum correlation threshold, and is set to 0.7 in the implementation of **Paper III** according to [124]. In addition, $mean(\cdot)$ and $std(\cdot)$ are the average and the standard deviation operators, respectively.

2.4.2 Samples-related outliers

After detecting and removing unreliable subjects, we further perform the rejection of outlier samples/observations for each subject. To achieve this goal, a median absolute deviation (MAD) method and the turkey fence method can be used depending on the distribution of the subject's observations [123]. These two methods are selected because they are both robust to identify the outliers when their

number represents less than 20% of all observations. MAD is used for approximately symmetric JND data distribution, whereas turkey fence method for asymmetric JND data distribution.

Based on MAD, the observation/stimulus j of a subject is labeled as an outlier if:

$$\frac{|T_j^i - \overline{T_j}|}{MAD} > 3, \tag{2.9}$$

and MAD_M can be calculated as

$$MAD = 1.4826 \cdot \underset{j=1:M}{med}(|T_j^i - \underset{j=1:M}{med}(T^i)|), \tag{2.10}$$

where $med(\cdot)$ denotes the median operator and M is the number of stimuli.

Based on turkey fence method, the observation s is identified as an outlier if:

$$T_j^i < (P_1 - 1.5 \cdot IQR) \quad or \quad T_j^i > (P_3 - 1.5 \cdot IQR), \tag{2.11}$$

where P_1 and P_3 respectively refer to the 25th and 75th percentile of all JND values for the stimulus j. IQR indicates the interquartile range of all JND values.

2.4.3 Normality validation and analysis of variance

To confirm the reliability of the JND data after outliers removal, a two-side goodness-of-fit Jarque-Bera (JB) test [125] is used to verify whether the JND data of each stimulus matches a normal distribution. JB for a stimulus is defined as follows:

$$JB = \frac{N}{6} \cdot (S^2 + \frac{(K-3)^2}{4}), \tag{2.12}$$

where N is the number of observations/samples, S is the sample skewness, and K is the sample kurtosis. The null hypothesis is that the samples come from a normal distribution with an unknown mean and variance. This null hypothesis is rejected either (1) if JB is larger than the critical value of 5% (significance level), or (2) $p-$value of the test is less than 5%.

Analysis of variance (ANOVA) is used to verify the statistical significance between different variables [126]. ANOVA measures the difference among group means in a sample. In particular, it provides a test of whether the population means of several groups are equal.

The null hypothesis for ANOVA is that there is no significant difference between variables. $p - value \leq 5\%$ rejects the null hypothesis, and indicates that there is a relationship between variables.

In fact, ANOVA is performed based on the assumptions including normality and homogeneity of variances of the data. Therefore, before carrying out ANOVA, the normality and homogeneity of variance for data distribution are checked with the Shapiro-Wilk test [127] and the Levene's test [128], respectively.

2.5 Perceptual quality assessment of 2D and 3D images

With the rapid advances in digital stereoscopic/3D image data, it becomes increasingly crucial to accurately and efficiently predict 3D image quality, which can be affected at different stages from image acquisition, compression, storage, transmission, to display. Accordingly, SIQA can be applied to evaluate/optimize the performance of 3D processing algorithms/systems (e.g., compression, enhancement, ...) [129–131]. The objective of IQA is to measure the perceived image quality, which is probably degraded in various ways. Therefore, a reliable SIQA method should measure the perceived quality highly correlating with the human judgment of quality.

2.5.1 Subjective IQA methods

The perceived quality of stereoscopic images can be assessed by either objective methods or subjective experiments. A comprehensive survey of subjective and objective IQA methods is recently presented in [132]. In subjective tests, the human subjects are asked to observe the test images and to give opinion scores. Since humans are the final receivers in most visual applications, subjective IQA methods can deliver reliable and referenced results. Many Subjective 2D- and 3D-IQA methods have been proposed by many international standardization organizations over the years. For instance, ITU recommended several standard subjective methodologies for quality assessment of 2D-TV and 3D-TV pictures including test methods, grading scales and viewing conditions [107, 108, 133]. According to [107, 108], the subjective testing methods can be usually classified into three categories: single-stimulus methods, multiple-stimulus methods, and paired comparison or stimulus comparison methods. For single-stimulus methods, only one test image is shown to subjects at any time instance and is given ratings to blindly reflect its perceived quality. For multiple-stimulus methods, several images are presented to subjects simultaneously, and the subjects rank all images based on their relative perceived quality. Finally, for paired comparison methods, a pair of images are shown either simultaneously or consecutively, and the subjects are asked to choose the one of better quality.

Subjective experiments are an important tool to construct IQA databases including reference and impaired images with different types of distortions and subjective opinions for all images, represented by either mean opinion score (MOS) or differential mean opinion score (DMOS). Over the past decades, several 2D-IQA publicly available databases were proposed to advance the work of the quality assessment community. Examples include the laboratory for image and video engineering (LIVE) database [134], tampere image database (TID2008) [135] and (TID2013) [136] databases, categorical image quality (CSIQ) database [137], LIVE multiply distorted image database (LIVEMD) [138], high dynamic range image database (HDR2014) [139], LIVE Challenge database [140], and Waterloo Exploration database [141]. More image quality databases are summarized in [142]. Subjective experiments can convincingly assess image quality and is accordingly considered as the reference results, however they are usually costly, time-consuming, and thus unsuitable for real-time application.

2.5.2 Objective 2D-IQA methods

To cope with the constraints of subjective assessment, one solution is to develop objective IQA metrics/methods to automatically assess the perceived image quality. In this section, we provide a brief review of 2D-IQA metrics closely related to our work. The readers may refer to [143–145] for recent surveys on 2D-IQA models.

Depending on how much information about the reference image is used, 2D-IQA models are typically divided into (1) full-reference (FR), (2) blind or no-reference (NR) and (3) reduced-reference (RR) metrics. FR metrics use whole reference images for quality prediction, while NR metrics assess the image quality without any cue about the reference image. A tradeoff between FR and NR metrics is represented by RR metrics, which predict image quality using only partial information (e.g., features) from the reference image. The earliest and most widely used FR metrics are the mean square error (MSE) and peak-signal-to-noise-ratio (PSNR), which simply quantify the difference between the reference I_r and the distorted I_d images, respectively. PSNR and MSE are respectively defined as follows:

$$PSNR = 10 \cdot log_{10} \cdot \frac{L_{max}^2}{MSE}, \tag{2.13}$$

and

$$MSE = \frac{1}{N} \sum_{n=1}^{N} \|I_r - I_d\|^2, \tag{2.14}$$

where N denotes the pixels number in the image I_r. L_{max} is the maximum pixel value of the image, and is usually equal to 255 for standard 8-bit images. Although PSNR is still widely used, it has a poor correlation with the human judgment of quality due to lack of consideration of the HVS properties [146]. Therefore, Wang *et al.* firstly proposed the universal quality index (UQI) [147] that is defined by:

$$\begin{aligned} UQI(I_r, I_d) &= \frac{1}{M} \sum_{m=1}^{M} UQI_{map}(I_r, I_d) \\ &= l(I_r, I_d) \cdot c(I_r, I_d) \cdot s(I_r, I_d) \\ &= \frac{2\mu_r\mu_d + C_1}{\mu_r^2 + \mu_d^2 + C_1} \cdot \frac{2\sigma_r\sigma_d + C_2}{\sigma_r^2 + \sigma_d^2 + C_2} \cdot \frac{\sigma_{rd}}{\sigma_r\sigma_d}, \end{aligned} \tag{2.15}$$

where $l(\cdot)$, $c(\cdot)$, and $s(\cdot)$ refer to luminance, contrast, and structural/correlation similarities between I_r and I_d, respectively. M is the number of local windows with size 8×8. μ_r and σ_r denote the average and variance of I_r respectively, whereas σ_{rd} is the covariance of I_r and I_d. The UQI metric is used in our proposed SIQA models (see **Paper IV** and **Paper V**).

For UQI metric, the constants C_1 and C_2 are equal to 0. To avoid $\mu_r^2 + \mu_d^2 = 0$ or $\sigma_r^2 + \sigma_d^2 = 0$, Wang *et al.* followed the idea of UQI and developed a structural similarity (SSIM) index, assuming that the HVS is sensitive to the structural information of a scene [148]. In fact, SSIM highlighted the

importance of the HVS characteristics and properties on IQA metrics design. Based on this finding, various reliable 2D FR-IQA models have been proposed over the last decade [149–153]. The existing 2D FR-IQA models from the literature can be classified into four categories described as follows:

- Scale transform-based models [149, 154–156]. The main idea is that the perceptual image quality depends on viewing distance and image resolution [157]. Thus, it is recommended to account for the appropriate [154, 156] or multiple scales [149, 155] so as to accurately predict image quality. The impact of image scale on SIQA performance is explored in **Paper VI** and **Paper VII**.
- Visual-sensitivity-based models [158–160] consider that the image pixels/regions have different visual importance in the perceived quality judgment. This visual importance can be modeled using visual saliency detection models [161–166], JND models [109–116]... In our case, saliency map based weighting on IQA is used in **Paper V** and **Paper VII**.
- Information-content-based models [150, 167, 168] measure the degradation on image content information extracted from the HVS based on scene statistics.
- Gradient-based models [151–153, 169] assume that the image low-level features (e.g., edges and textures) play an important role in perceptual IQA tasks. These low-level features correspond to image high frequency components that can be represented by image gradient magnitude (GM) and phase. Accordingly, the gradient similarity based FR 2D-IQA have been successfully proposed in the recent years. For instance, Xue *et al.* developed the gradient magnitude similarity mean (GMSM) and gradient magnitude similarity deviation models computing the quality score with average and standard deviation pooling strategies, respectively. GM-based 2D metric is used in **Paper VII** thanks to its effectiveness and efficiency. In addition, the feature similarity metric (FSIM) [151] based on gradient magnitude and phase congruency is used in **Paper VI** because of its remarkable performance.

Although previously described FR-IQA models can achieve high quality prediction accuracy, they are not applicable in real-world case because the reference image is accessible at the receiver side of the image based systems. Therefore, NR-IQA models are very important to blindly assess image quality.

As described in [170, 171], most recently proposed blind image quality assessment (BIQA) models can be divided into three main stages as shown in Figure 2.18. The first stage is to preprocess the test image in order to efficiently obtain useful image information. This stage may include color conversion, domain transform, normalization, scaling and so on. The second stage is the feature learning process containing either (1) the feature extraction and machine learning-based regression (e.g., support vector machine (SVM) and support vector regression (SVR) [172, 173], artificial neural network (NN) [174], and random forest (RF) [175] for some conventional BIQA models [176–183] or (2) deep NN for deep-learning-based BIQA models [171, 184–189], e.g., convolutional neural networks (CNNs). The last stage is to compute the quality score based on learned/regression model with/without using human opinion scores (i.e., MOS or DMOS values).

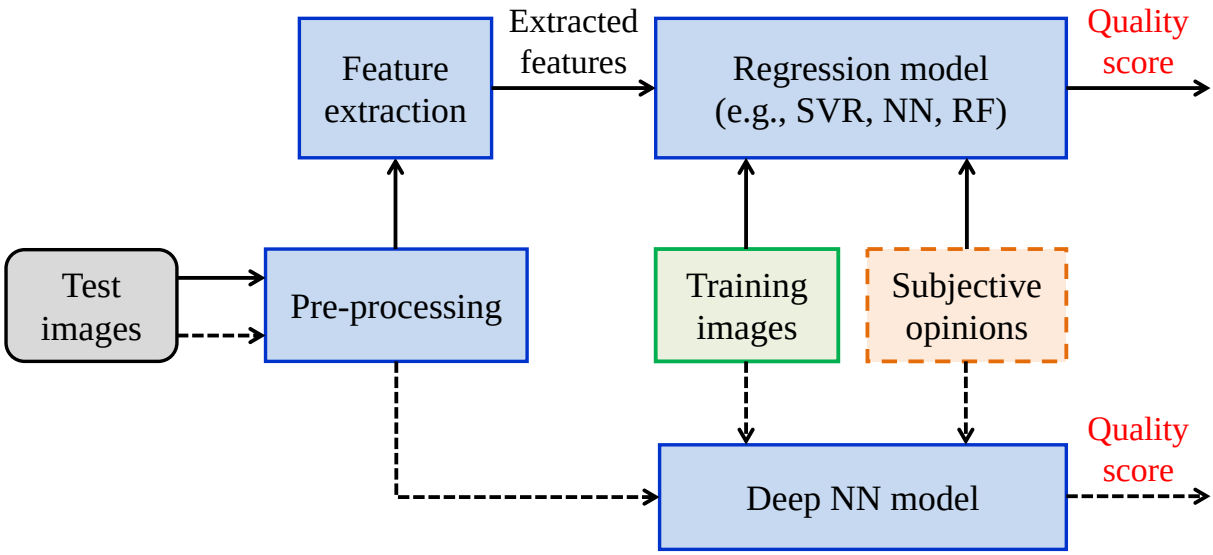


Figure 2.18 – Flowchart of existing NR-IQA models.

Based on [144, 145, 190], the existing 2D NR-IQA approaches can be mainly classified into three categories described as follows:

- Distortion-specific models target one or multiple specific distortions types and prior knowledge on the distortion properties is known. Specifically, these approaches assess the quality of images impaired by particular artifacts, e.g., "blockiness" [191, 192], "blurring" [193, 194], "ringing" [195, 196] or mixed [197, 198].
- Distortion-unaware and opinion-aware models that can evaluate the quality of distorted images without knowing the distortion types. Furthermore, such models require the subjective opinion scores for the training stage (as described in Figure 2.18). In fact, a majority of state-of-the-art NR IQA metrics belong to opinion-aware methods [176, 177, 180, 199–201]. For instance, Xue *et al.* developed a reliable and efficient BIQA model that extracts histogram-based features from the joint statistics of gradient magnitude (GM) and LoG responses [180]. Since this model delivered remarkable performance for 2D IQA, we adopt it in our SIQA metric design (see **Paper VI**). We will detail the SVR later. As described in [190], natural scene statistics can capture the natural statistical behaviors of images that are highly sensitive to different quality distortions in images. Therefore, many natural scene statistics based BIQA models [177, 199–201] were developed by measuring the destruction of "naturalness" by distortions. Meanwhile, some other BIQA models employed the quality-relevant features capturing the factors affecting image distortion [177, 202, 203].
- Distortion-generic and opinion-unaware models that do not require subjective opinion scores for training, because obtaining subjective scores can be probably expensive and time-consuming. Therefore, this category of models can be called totally blind IQA approaches [204–208]. For

instance, Zhang *et al.* first extracted a set of image local features and then fitted the feature vectors by a multivariate gaussian model [206].

In IQA tasks, the SVR regression model is often used to map the image feature vectors to subjective scores (e.g., MOS/DMOS) so as to derive the learning model used further in the testing stage. To achieve this many successful BIQA models use SVR with an ϵ-insensitive loss function (i.e., ϵ-SVR [173, 209]) to successfully predict image quality [177, 180, 181].

Given the training features $\{(v_1, s_1), (v_2, s_2), ..., (v_{n-1}, s_{n-1}), (v_n, s_n)\}$, where v_i and s_i denote the feature vector and MOS/DMOS values, respectively. n is the number of training patches/images. The prediction function is defined as follows:

$$f(\omega, x) = \sum_{i=1}^{2n} \omega_i \cdot x_i + b, \tag{2.16}$$

where b is the bias parameter. ω denotes the weight vector learned by minimizing the sum of a loss function L that is described below:

$$L(\varepsilon) = \begin{cases} 0, \ if \ \varepsilon \leq \epsilon \\ \varepsilon - \epsilon, \ otherwise, \end{cases} \tag{2.17}$$

where $\varepsilon = |s - \hat{s}|$, $\hat{s}$ denotes predicted quality score, and $\epsilon > 0$. So, the idea is to find the optimal ω and b to ensure that the ϵ-margin is maximized. In other words, we need to minimize $\|\omega\|^2$. More specifically, the optimal values (ω_*, b_*) of ϵ and b are determined by solving the optimization problem as follows:

$$(\omega_*, b_*) = argmin \ \ \frac{1}{2} \|\omega\|^2 + C \sum_{i=1}^{n} (\varepsilon_i + \varepsilon_i^*), \tag{2.18}$$

subject to:

$$\begin{cases} \sum_{i=1}^{n} \omega_n \cdot K(v_i, v) + b - s_i \leq \epsilon + \varepsilon_i, \\ s_i - \sum_{i=1}^{n} \omega_n \cdot K(v_i, v) - b \leq \epsilon + \varepsilon_i^*, \\ \varepsilon_i, \varepsilon_i^* \geq 0 \end{cases} \tag{2.19}$$

where C is a constant, K is the kernel function for features mapping. The linear radial basis function is widely used because of its high performance, and is given by:

$$K(v_i, v) = e^{-\gamma \cdot |v_i - v|^2}, \tag{2.20}$$

where γ is a constant. More detail about this approach can be found in [173, 209]. Note that the LIBSVM package was used in the implementation of **Paper VII**.

Besides, most SVR-based NR-IQA methods divide an image quality database into two non-overlapped subsets: the training and the testing subsets which respectively contain $p\%$ and $(1-p)\%$ of

all images. p varies depending on the proposed algorithm and usually is set to 80. To derive convincing results, both training and testing procedures are repeated 1000 times and the mean/median value is considered as the estimated quality score.

2.5.3 Performance evaluation methods

To evaluate the performance of quality metrics, estimated quality scores should be compared to subjective scores Q_s (e.g., MOS/DMOS) based on statistical tools. In this section, we introduce the statistical tools applied in the enclosed **Paper I**, **Paper IV**, **Paper VI**, and **Paper VII**. After obtaining the predicted quality scores, we need a five-paramters logistic function (see Equation 2.21) to map the score estimates on the same scale of Q_s so as to avoid the nonlinearity brought from the IQA model [210]. This logistic function is defined as:

$$Q_p = p_1 \cdot \left[\frac{1}{2} - \frac{1}{e^{(p_2 \cdot (Q - p_3))}}\right] + p_4 \cdot Q + p_5, \tag{2.21}$$

where p_1, p_2, p_3, p_4, and p_5 are the regression parameters selected based on subjective score. The estimated quality that is represented by Q denotes the quality predicted by the metric. Q_p are the predicted scores after non-linear regression. In experimental experiments, we set $p_1 = Max(Q)$, $p_2 = Min(Q)$, $p_3 = mean(Q_s)$, $p_4 = 0.1$, and $p_3 = 40$.

Subsequently, to evaluate the metrics performance, we use three statistics-based criteria including: PCC, SROCC, and root-mean-square error (RMSE). Some other statistical criteria can be used for performance evaluation, such as Kendall rank order correlation, mean absolute error, etc.

PCC measures the level of similarity between subjective scores s and predicted scores $\hat{s}$, and is defined as follows:

$$PCC(s, \hat{s}) = \frac{\sum_{n=1}^{N}(s_n - \mu_s) \cdot (\hat{s}_n - \mu_{\hat{s}})}{\sqrt{\sum_{n=1}^{N}(s_n - \mu_s)^2} \cdot \sqrt{\sum_{n=1}^{N}(\hat{s}_n - \mu_{\hat{s}})^2}}, \tag{2.22}$$

where $\mu_s/\mu_{\hat{s}}$ denote the average value of $s/\hat{s}$. N is the number of distorted images.

SROCC evaluates the strength of correlation between the subjective score and predicted score using a monotonic function. Thus, to compute the SROCC value between s and $\hat{s}$, we first convert the raw scores to their ranks x_n and y_n, respectively. Then $SROCC(x, y)$ is calculated by:

$$SROCC(x, \hat{y}) = \frac{\sum_{n=1}^{N}(x_n - \mu_x) \cdot (y_n - \mu_y)}{\sqrt{\sum_{n=1}^{N}(x_n - \mu_x)^2} \cdot \sqrt{\sum_{n=1}^{N}(y_n - \mu_y)^2}}. \tag{2.23}$$

Both PCC and SROCC values are in $[-1, 1]$. Higher absolute PCC and SROCC values indicate higher prediction accuracy and better prediction monotonicity of the quality metrics, respectively. $PCC = 0$ or $SROCC = 0$ reflects the absence of correlation between predicted and subjective scores. RMSE is used to estimate the prediction consistency, which measures the difference between s and $\hat{s}$, and is

given as follows:

$$RMSE(s, \hat{s}) = \sqrt{\frac{1}{N}\sum_{n=1}^{N}(s - \hat{s})^2}, \tag{2.24}$$

Lower $RMSE$ indicates higher performance of the quality metrics. $RMSE$ being close to 0 means the best performance.

Based on the advance of 2D IQA, many SIQA methods were proposed. An overview of the state-of-the-art SIQA approaches is given in **Paper VI** and **Paper VII**.

Chapter 3

Summary of Results and Contributions

In this section, we give a summary for each paper included in Part II.

3.1 Summary of Paper I

Yu Fan, Mohamed-Chaker Larabi, Faouzi Alaya Cheikh, and Christine Fernandez-Maloigne. **On the performance of 3D just noticeable difference models** in *IEEE International Conference on Image Processing (ICIP)*, pages 1017–1021, September 2016.

Paper I provides a comparative study of the exiting 3D-JND models by analyzing their pros, cons, complexity, suitable 3D format and applications. Each model is briefly presented by giving its main components based on visual masking effects or psychophysical experiments. Some of the 3D-JND models are proposed as an extension of existing 2D-JND models described in Section 2.4.

In order to quantitatively evaluate the performance of each 3D-JND model, we adopt a QA framework including a 3D-JND block [211]. The framework is illustrated in Figure 3.1. First, the qualities of the left and right views of the distorted 3D images are respectively computed using the SSIM metric. To address QA for monocular and binocular regions respectively, the pixels of the distorted 3D images are categorized into occluded and non-occluded pixels based on the cross-checking method [212]. The main idea is to compare the disparity maps of the left and right views to detect the occluded regions. In this work, we use the algorithm presented in [213] to perform the classification. Given the following steps. To determine the perceptual importance of each pixel, a 2D-JND model and a 3D-JND model are used for occluded and non-occluded pixels, respectively. The red block in Figure 3.1 can be substituted by the test 3D-JND model. The JND is used to weight the SSIM score to obtain the quality score of the occluded pixels (Q^{oc}). In a similar way, the quality scores of the non-occluded pixels for

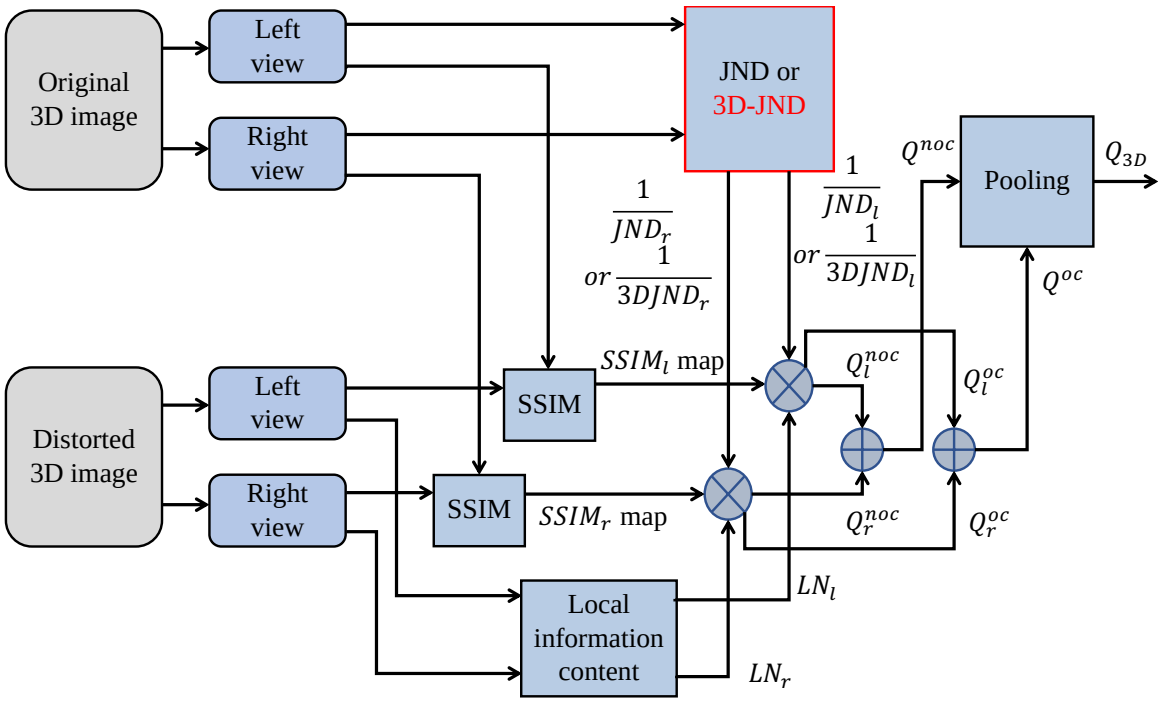


Figure 3.1 – Block diagram of the test procedure.

left and right views are computed separately. The 3D quality score of the non-occluded pixels (Q^{noc}) is determined based on a linear summation with a local entropy weighting (LN) for each view. The final 3D quality score (Q_{3D}) is calculated by combining Q^{oc} and Q^{noc}.

To compare the performance of the 3D-JND models, each model is integrated into the above-described quality metric, and the performance of the metric is estimated. In particular, their performance are compared by analyzing the image quality prediction accuracy based on LIVE 3D Phase II [214] and Waterloo IVC Phase I databases [215]. The image quality prediction accuracy is calculated by computing PCC, SROCC, and RMSE (see Section 2.5.3) between estimated quality scores and subjective scores obtained from the databases. Experimental results demonstrate that SIQA metric using stereo just noticeable difference (SJND) can deliver the best performance among all compared models thanks to various considered VM properties.

In **Paper I**, A_C for BJND description represents the visibility threshold of the right view if there is no noise in the left one, and thus denotes the maximum threshold. For SJND model, it is initially developed for 3D video quality assessment. The temporal masking effect (caused by motion) is consequently ignored to be adapted to 3D image (i.e., left and right frame). In this case, the temporal JND (TJND) of the occluded pixel is determined by the background luminance and contrast masking. More specifically, referring to [216], in our case, $f_3(bg(i), mg(i)) = P$, and $f_4(bg(i), mg(i)) = Q$ for an image, and thus $TJND(i) = JND(i)$. Then, $TJND$ and $BPJND_O$ are fused using a linear summation in order to obtain the JND for occluded pixels. It is finally combined with JND of non-

occluded pixels to derive the SJND.

The main contribution of this paper is to determine the appropriate and reliable 3D-JND model used in IQA performance improvement. Moreover, this paper allows determining the most important features considered in 3D-JND model, helping in the construction of a more accurate and efficient 3D-JND model.

3.2 Summary of Paper II

Yu Fan, Mohamed-Chaker Larabi, Faouzi Alaya Cheikh, and Christine Fernandez-Maloigne. **A Survey of Stereoscopic 3D Just Noticeable Difference Models** in *IEEE Access*, vol. 7, pp. 8621-8645, 2019.

Paper II, which is an extension of **Paper I**, provides a comprehensive and deep overview of state-of-the-art 3D-JND models. In particular, this paper firstly summarizes the contrast sensitivity of the HVS and visual masking effects that are related to monocular vision and binocular vision. Then, each model is exhaustively described by showing its framework, HVS characteristics and visual masking effects considered in the model. Next, we present a thorough comparative analysis of 3D-JND models in terms of applications, pros, and cons. Table 3.1 shows an overview of the comparison of 3D-JND models.

In addition, we carry out extensive experiments in order to compare the performance of the 3D-JND models in terms of distortion masking ability and accuracy. Specifically, we experimentally perform qualitative and quantitative evaluations using 3D images from Middlebury stereo database [54, 217, 218]. Experimental results show that the hybrid just noticeable difference (HJND) and the joint just noticeable difference (JJND) outperform the other 3D-JND models in terms of distortion tolerance ability, because they highly depend on depth information having a great effect on distortion masking. Furthermore, multi-view just noticeable difference (MJND) and binocular just noticeable difference (BJND) achieve better performance than other models in terms of edge-distortion masking ability. In addition to above-mentioned experiments, we further conduct psychophysical experiments to compare the accuracy of 3D-JND models. In particular, we first synthesize 3D images containing textures collected from the ETHZ dataset [219]. The relationship between the used disparity and depth values are determined based on Section 2.2.3. Then, we perform subjective tests to measure the visibility thresholds of the asymmetric distortion in stereopairs, which is considered as subjective JND values. Next, objective JND values are calculated using each 3D-JND model. Finally, the accuracy of each 3D-JND model is evaluated by comparing subjective and objective JND values. Experimental results demonstrate that stereo just noticeable difference (SJND) and BJND models outperform other models in terms of estimation accuracy, and achieve best and second-best performances, respectively. This finding corresponds to the conclusion in our **Paper I**.

In **Paper II**, regarding the generation of stimuli, the increment steps for each distortion type are

selected independently according to perceptual noticeable distortion in the image. Specifically, before conducting the experiments, several trials have been done to obtain an appropriate increment step for each distortion type. The criterion is to ensure that subjects are able to get the JND by increasing N times the distortion level. N is chosen randomly from the range of [5,15] to avoid the error of habituation.

The major contributions of this paper include (1) a comprehensive overview and a comparative analysis of the 3D-JND models, and (2) extensive experimental comparison between the 3D-JND models based on qualitative and quantitative performance evaluation.

3.3 Summary of Paper III

Yu Fan, Mohamed-Chaker Larabi, Faouzi Alaya Cheikh, and Christine Fernandez-Maloigne. **Just Noticeable Difference Model for Asymmetrically Distorted Stereoscopic Images** accepted in *International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Brighton,UK., 2019.

In **Paper III**, we develop a novel 3D-JND model based on psychophysical experiments, accounting for visual masking effects and binocular disparity (see Section 2.1.1). The proposed 3D-JND of one view in a stereopair estimates the maximum distortions that can be introduced in this view without binocularly evoking visible differences, given the distortions in the corresponding region of the other view.

Our proposed 3D-JND model is inspired by binocular JND (BJND) to cope with the issue of not considering the binocular disparity in the design of psychophysical experiments [220]. More specifically, we first conduct psychophysical experiments in which we measure the visibility thresholds of the asymmetric noise using stereopair patterns considering luminance adaptation, contrast masking, and binocular disparity (see Figure 3.2). The psychophysical data is processed to remove the subjects- and samples-relevant outliers using statistical tools (described in Sections 2.4.1 and 2.4.2). Then JND data after removing outliers is used to analyze the relationship between JND values and considered attributes based on analysis of variance (ANOVA) method (described in Section 2.4.3). Next, 3D image pixels are divided into non-occluded pixels (NOPs) and occluded pixels (OPs) based on [212]. The post-processed JND thresholds are used to construct a disparity-aware binocular just noticeable difference (DBJND) model, allowing to estimate the JND values for NOPs. For OPs, a reliable 2D-JND model [114] is used to obtain JND thresholds. Finally, the proposed 3D-JND model is built by weighting the JND estimates for NOPs and OPs with 3D visual saliency map [221].

To validate the effectiveness of the proposed 3D-JND model, we conduct subjective test (based on adjectival categorical judgment method, see Sections 2.3.3 and 2.5.1) using pristine stereopairs (from Middlebury stereo datasets [54, 217, 218]) and asymmetrically distorted stereopairs. Experimental results validate that the proposed 3D-JND model outperforms other models in terms of perceptual

Table 3.1 – **Paper II**: comparison between the 3D-JND models.

	JNDD	**BJND**	**JJND**	**MJND**	**SJND**	**HJND**	**DJND**
Inputs	VD, DpM	LCs NAM	LC, DsM, DpM	LCs, DpM, TI	LCs, DsM, TI	LCs, DpM	LC, DsM
VM & Features	DM	LA, CM	LA, CM, DI	LA, CM, TM, DM	LA, CM, TM, BM	LA, CM, DI, DC, GD	LA, CM, DI, DOF
3D format	2D + depth	LR images	LR images	MVD	LR frames	DIBR, MVD	LR images
Model validation	Theoretical results vs results derived from PE	Noise detection probability in S3D images	Comparison with 2D-JND in terms of SQ	MJND-based MVC vs JMVM-based MVC in terms of CE and PQ	A metric using SJND vs SVQA metrics in terms of SQ	MVC with HJND vs MVC with JJND vs JMVC in terms of CE and SQ	MVC with DJND vs MVC with 2D-JND vs JMVC in terms of CE,PQ and VC
Complexity	_____	***	*	*	**	*****	**
Pros	Extension for various 3D displays	Suitable to several 3D formats	Binocular vision properties	Multiple MEs	Multiple MEs	Considering DC and GD	Several 3D formats, VC improvement
Cons	Limit 3D format, influence of depth image quality	Disparity effect ignoring, SMA impact on JND accuracy	Accuracy decrease for low or uniform disparities, lack of comparison with 3D-JND models	Accuracy decrease for large depth range	Difficult to design the PE for model validation, many parameters in the model	Highly depending on DIBR techniques, specially designed for MVD format	Accuracy decrease for 3D image with small depth difference between FRs and BRs
Application	Depth sensation enhancement, 3D QoE enhancement, 3D video coding, S3DW	Sharpness /contrast enhancement, 3D video coding, SIQA, S3DW, S3D IR	S3DW	MVC and 3D-HEVC	SVQA	MVC	MVC

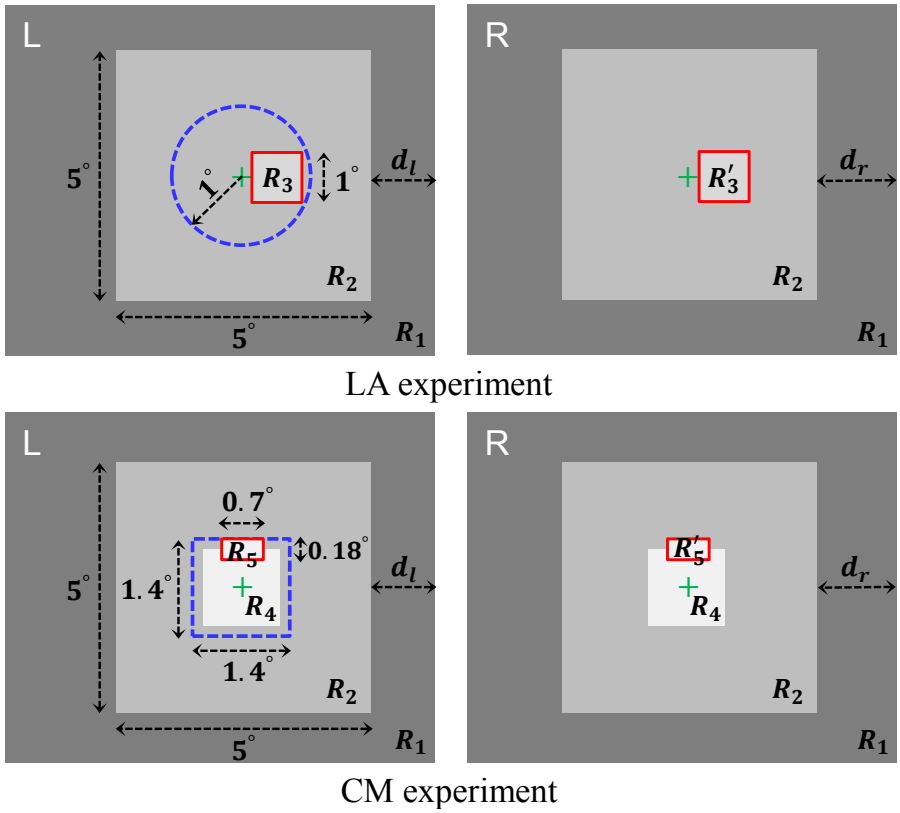


Figure 3.2 – **Paper III**: stereo pair patterns used in psychophysical experiments.

quality at the same noise level.

The main contributions of this paper include (1) new disparity-aware stereopair patterns used in psychophysical experiments for measuring luminance adaptation and contrast masking related JND threshold, (2) a novel 3D-JND model that can be applied in the improvement of 3D compression efficiency and SIQA performance.

3.4 Summary of Paper IV

Yu Fan, Mohamed-Chaker Larabi, Faouzi Alaya Cheikh, and Christine Fernandez-Maloigne. **Stereoscopic Image Quality Assessment based on the Binocular Properties of the Human Visual System** in *International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 2037–2041, March 2017.

Paper IV presents a stereoscopic IQA method based on HVS properties. Figure 3.3 exhibits the flowchart of this method. In particular, the main idea is to assess 3D quality by predicting the quality of the cyclopean image (CI) generated by fusing left and right views (see Section 2.1.1). The cyclopean image is first synthesized based on the local entropy of each view with the aim to simulate binocular

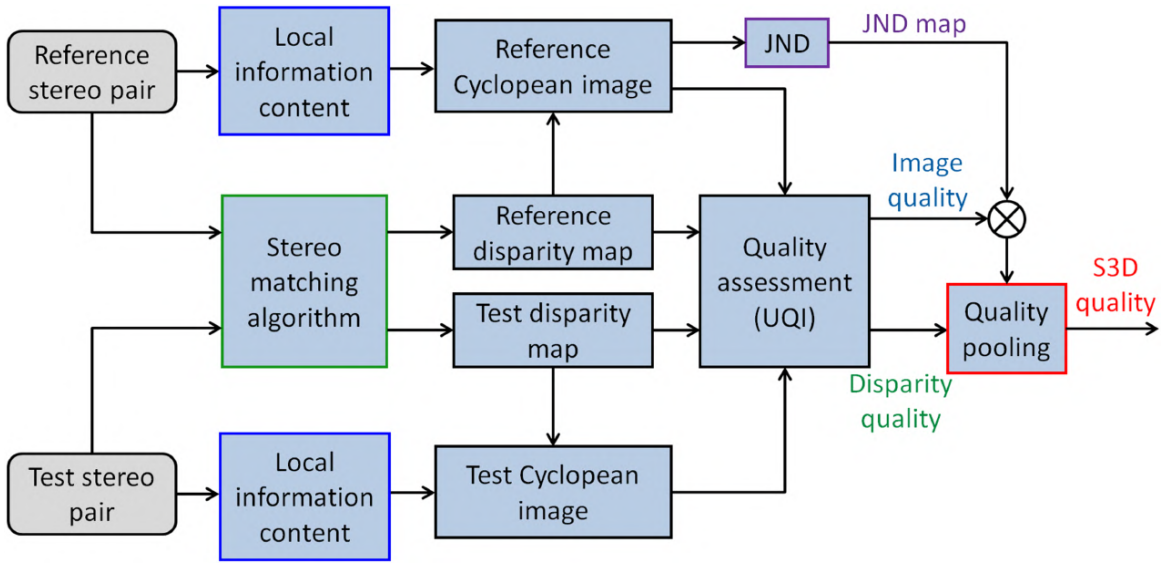


Figure 3.3 – **Paper IV**: framework of the proposed SIQA method.

rivalry or binocular suppression phenomenon. Then, the disparity information is obtained using an efficient stereo matching algorithm. Next, UQI (see 2.5.2 for details) is used to compute the quality of both cyclopean image and disparity map (DsM). In addition, the quality of the cyclopean image is modulated by visual importance of each pixel, which is determined by JND. Finally, the 3D quality is derived by combining the quality estimates of cyclopean image and DsM. Experimental results on LIVE 3D Phase II database [214] show that the proposed method outperforms some well-known 2D- and 3D-based SIQA methods in terms of prediction accuracy and computational efficiency. Moreover, the effectiveness of our method for asymmetrically distorted stereopairs has been validated. Besides, we further analyze the advantages of considering both JND and disparity quality in our proposed method.

The major contributions of this paper consist in proposing a novel SIQA metric by modeling binocular suppression or binocular rivalry phenomenon, and accounting for disparity image quality as well as the monocular spatial sensitivity of the HVS. Furthermore, it gives a comprehensive experimental evaluation and comparison between our metric and other SIQA metrics in terms of prediction accuracy and computational costs.

3.5 Summary of Paper V

Yu Fan, Mohamed-Chaker Larabi, Faouzi Alaya Cheikh, and Christine Fernandez-Maloigne. **Full-Reference Stereoscopic Image Quality Assessment account for Binocular Combination and Disparity Information** in *IEEE International Conference on Image Processing (ICIP)*, pages 760–764, September 2017.

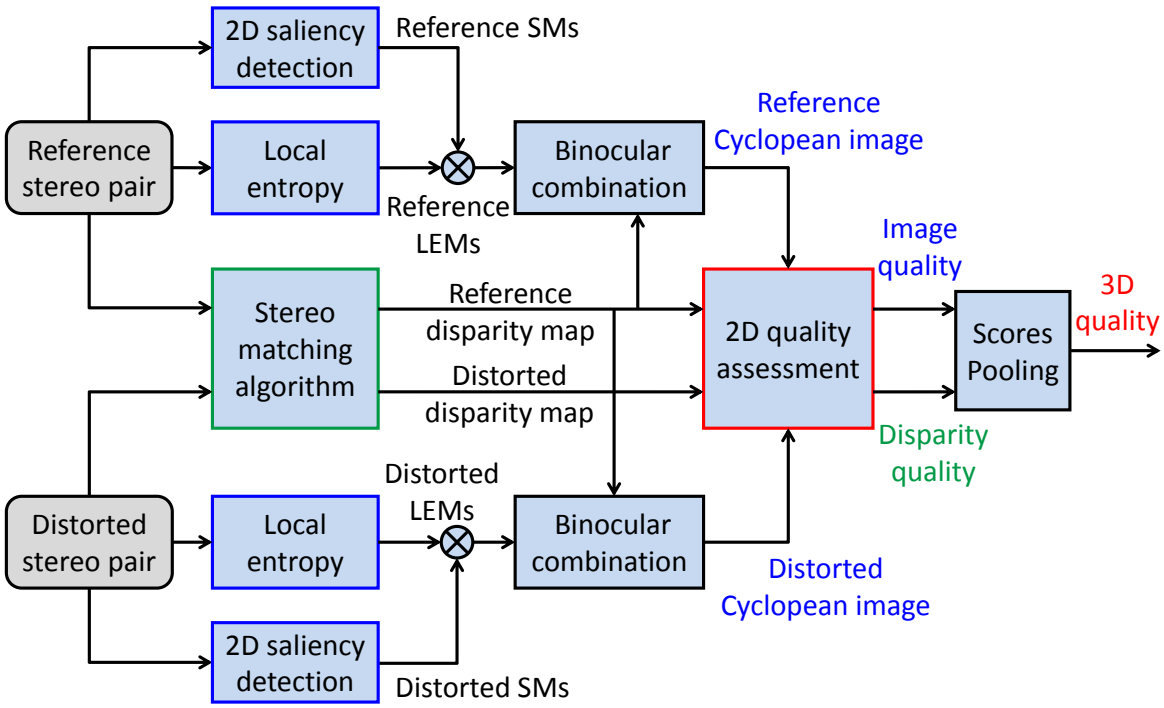


Figure 3.4 – **Paper V**: Flowchart of the proposed SIQA method.

In **Paper V**, we propose a SIQA metric accounting for binocular combination properties and disparity distortion. Similar to **Paper IV**, the qualities of the cyclopean image and disparity map are combined to yield the overall 3D quality. Figure 3.4 shows the framework of the proposed approach. To assess 2D image quality, UQI is used for LIVE 3D Phase I and II database [214, 222], whereas visual information fidelity (VIF) metric [150] (see information-content-based FR 2D-IQA models in Section 2.5.2) is used for Waterloo IVC Phase I database [215]. The proposed SIQA metric uses both local entropy and visual saliency of each view to accurately mimic the strength of the view dominance of binocular rivalry phenomenon. Experimental results on mentioned-above 3D IQA database demonstrate that our method achieves high prediction accuracy and better performance than many other SIQA methods for LIVE 3D I and II. However, the performance of the proposed method is decreased on Waterloo IVC I, because the use of mixed asymmetric distortions types in this database is making quality prediction more challenging.

The main contribution of this paper lies in development of a new FR SIQA model considering the binocular visual properties and monocular visual sensitivity. In addition, we make an extensive comparison between the proposed model and other SIQA models in terms of overall performance, performance on individual distortion and performance for symmetrically and asymmetrically distorted stereopairs.

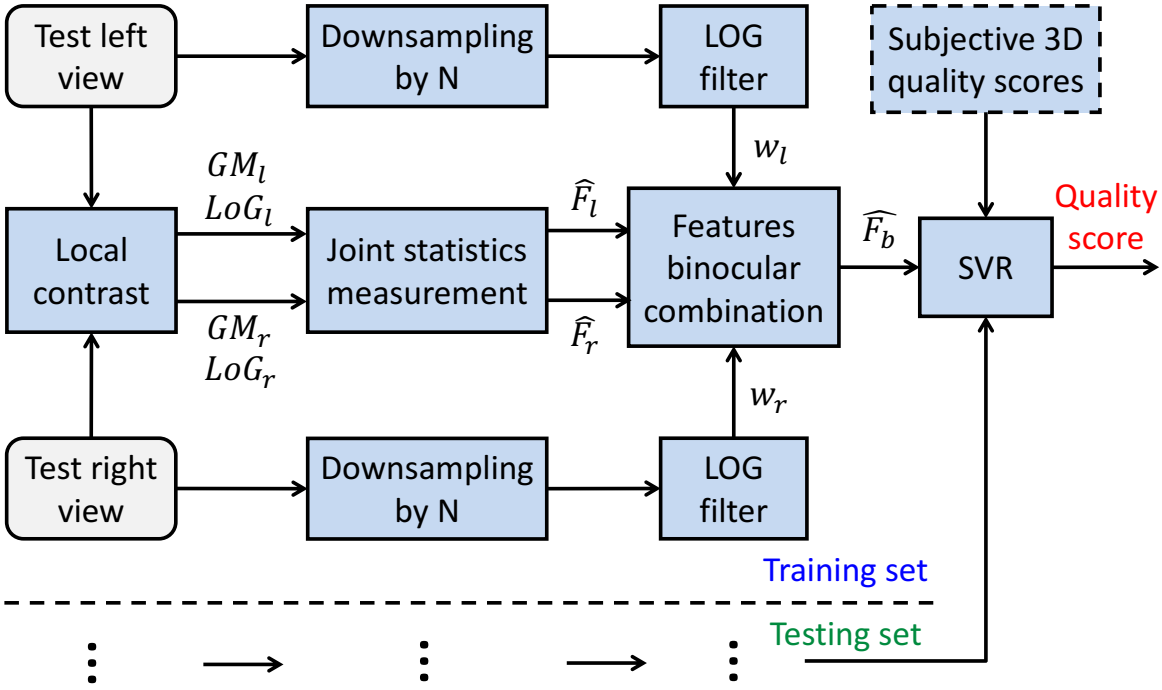


Figure 3.5 – **Paper VI**: framework of the proposed NR-SIQA model

3.6 Summary of Paper VI

Yu Fan, Mohamed-Chaker Larabi, Faouzi Alaya Cheikh, and Christine Fernandez-Maloigne. **No-Reference Quality Assessment of Stereoscopic Images based on Binocular Combination of Local Features Statistics** in *IEEE International Conference on Image Processing (ICIP)*, pages 3538–3542, October 2018.

Paper VI presents an opinion-aware NR SIQA method based on the binocular combination of monocular primitive structures, which are described by statistics of the image local contrast. The cyclopean-based SIQA methods require the depth/disparity information, which is not always available, and its estimation is probably error-prone and costly. To cope with this constraint, our proposed methods are developed without requiring disparity information unlike the methods in **Paper IV** and **Paper V**.

Specifically, following the strategy described in [180], we first extract the joint statistical features of gradient magnitude (GM) and Laplacian of Gaussian (LoG) responses for left and right views that describe the image local contrast from different perspectives. Then, the statistical features of both views (called monocular features) are combined to derive the binocular statistical features based on a linear combination model simulating the binocular rivalry phenomenon of the HVS. We estimate the LoG map for each view and use two LoG maps to calculate the weights in combination model, which simulate the strengths of the views dominance of the binocular rivalry behavior. Wang *et*

al. concluded that the image scale impacts the performance of the IQA metric, and the first- and second-ranked performances were given respectively by scale 2 and 3 [149]. Consequently, we apply the LoG filter in a single-scale image with an optimal scale size depending on image resolution. Finally, binocular statistical features and subjective scores (provided in 3D-IQA databases) are jointly employed to construct the learned regression model obtained by SVR algorithm (see Section 2.5.2 for more details).

The performance of the proposed NR SIQA method is evaluated and compared with other FR- and NR-SIQA methods on LIVE 3D Phase II [214], Waterloo IVC Phase I and II [215] databases. Experimental results show that the proposed method delivers highly competitive performance compared to other SIQA methods.

This paper includes two main contributions. First, unlike previous SIQA methods using Gabor filter magnitude or local entropy, or local variance, we employ the LoG response to mimic the strength of the view dominance of the binocular rivalry phenomenon. Second, to reduce the computational complexity, we only use binocular features combined from monocular statistical features for the training and testing processes. Third, we validate the effectiveness of our method on each of the three databases, and cross-database, and determine the importance of using an appropriate image scale for quality assessment.

Note that LoG and GM features are proposed in this quality metric to quantify the image local distortions. This is because both features represent the image structural information, which is often extracted by the HVS for quality judgment. Furthermore, the local-structure-based methods have shown their effectiveness in quality prediction [180]. Nevertheless, any additional and advanced features (e.g., biologically inspired feature [223], chrominance information [224], and energy information [225]) can be used as well. A summary of the features used in 2D/3D quality is given in [226]. Although our proposed method can successfully deal with the additive white noise and JPEG distortions, more features should be considered to improve the quality prediction for certain types of distortion (e.g., Gaussian blur and JPEG 2000). For instance, image gradient orientation information [227], which is detected by visual cortical neurons, can be used to represent the change in image anisotropy due to local distortion. In addition, we propose to use the image local directionality described in [228] to measure the information loss of the dominant structures caused by blurring-based distortion.

3.7 Summary of Paper VII

Yu Fan, Mohamed-Chaker Larabi, Faouzi Alaya Cheikh, and Christine Fernandez-Maloigne. **Stereoscopic Image Quality Assessment based on Monocular and Binocular Visual Properties** submitted in *Journal of Visual Communication and Image Representation*, 2018.

Based on **Paper IV** and **Paper V**, we propose a new FR-SIQA system in **Paper VII** considering the qualities of : (1) the left- and right-views images (with respect to monocular vision), and (2) the

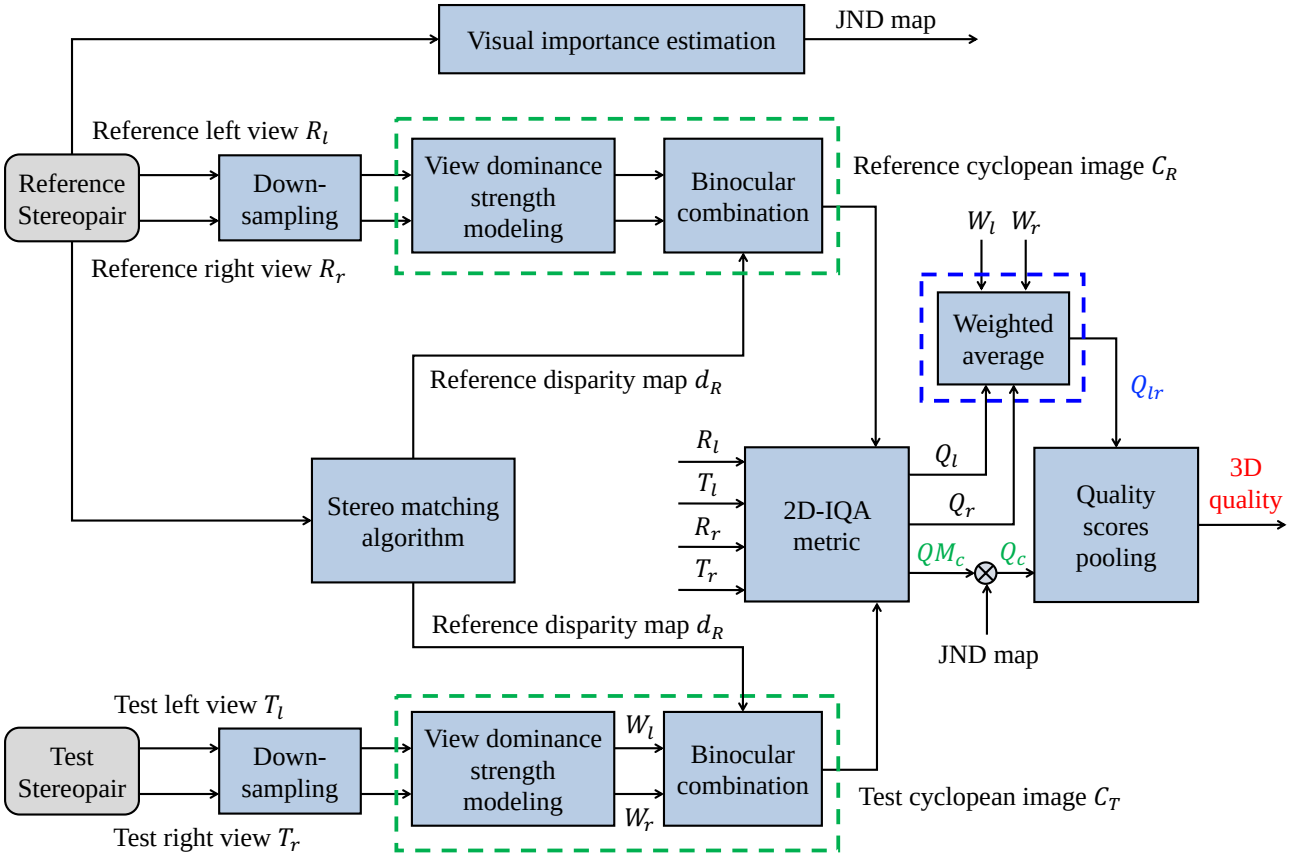


Figure 3.6 – **Paper VII**: framework of the proposed SIQA model.

cyclopean image (with respect to binocular vision). In particular, we firstly estimate the quality of left and right image separately base on GMSM (see gradient-based models in Section 2.5.2), and then linearly combine the qualities of both views into a 2D monocular image quality with the weights modeling relative stimulus strength of each view. For the proposed SIQA system, Laplacian of Gaussian (LoG), local entropy (local entropy (LE)) and gradient magnitude (GM) based weighting strategies are used respectively, to explore their effectiveness. Next, using the GMSM metric, we compute the quality score of the cyclopean image that is synthesized using a binocular combination model. Additionally, to reveal different visual sensitivities on image's region distortion, the cyclopean image quality of the test stereopair is weighted with a JND map of the reference stereopair. Two 2D-JND models and three 3D-JND models are used to compare the accuracy of the quality prediction of the proposed SIQA method integrating with JND. Finally, overall 3D quality is assessed by combining 2D monocular image quality with 3D binocular-based JND-weighted cyclopean image quality.

We achieve an extensive and comprehensive performance evaluation of our proposed model and many other state-of-the-art SIQA models on seven publicly available 3D IQA databases including: LIVE 3D Phase I and II [214, 222], Waterloo IVC Phase I and II [215], NBU 3D II [229], NBU-MDSID [230], and IEEE 3D [231]. Experimental results on these databases demonstrate the proposed

method correlates well with the human quality judgments and outperforms many recent well performing SIQA methods. Moreover, the LoG-based binocular rivalry modeling achieves better performance than LE- and GM-based binocular rivalry modeling for most databases. Besides, the proposed 3D metric using BJND [220] delivers the best performance among the metrics using other 2D-JND or 3D-JND models for databases only containing symmetrically distorted stereopairs. Furthermore, our metric without JND and with disparity-based just noticeable difference (DJND) can perform well for databases containing symmetric and asymmetric distortions.

The major contributions of this paper include:

- A new SIQA system accounting for degradation of the stereopair-based monocular scene and of the cyclopean-based binocular scene using different visual stimulus strength modeling methods.
- An overview of the existing 3D IQA databases. Comprehensive experimental evaluations of the proposed system, and exhaustive performance comparisons between our SIQA model and state-of-the-art on seven publicly available 3D IQA databases.
- Investigation of the importance of binocular-rivalry-inspired monocular 2D quality and cyclopean quality on overall 3D quality.
- Study of the impacts of different JND models and strategies of simulating the strength of view dominance on 3D quality prediction accuracy.

Chapter 4

Discussion

This Chapter first presents the technical contributions to 3D-JND and SIQA, and the answers of the *Research Questions* in Chapter 1.3, respectively, with respect to aforementioned papers in Chapter 3. Then, it provides additional results for the included papers. Finally, some limitations and shortcomings concerning 3D-JND and SIQA are discussed in this Chapter.

4.1 Contributions of the thesis

4.1.1 Contributions to 3D-JND

Paper I – **Paper III** are the main contributions to research in spatial visual sensitivity of the HVS for 3D images. The objective of the work described in these papers is to propose a reliable 3D-JND model accounting for visual sensitivity related properties of monocular vision and of binocular vision, by analyzing existing 3D-JND approaches. **Paper I** and **Paper II** contribute to overview and comparison of state-of-the-art 3D-JND models. **Paper III** contributes to the design of a new 3D-JND model based on psychophysical experiments.

4.1.1.1 Overview of the state-of-the-art 3D-JND models

Paper I summarizes the monocular and binocular VM effects and contrast sensitivity of the HVS considered in 2D-JND and 3D-JND models. This gives an answer of *Research Question* **Q1** that digital 2D/3D imaging should consider luminance adaptation, contrast masking, texture masking effects, contrast sensitivity and depth-related masking/information, which are the most important HVS characteristics and properties.

In addition, surveying existing 3D-JND models, **Paper I** briefly describes the framework of each

existing 3D-JND model in order to answer *" How are the performance of the state-of-the-art 3D-JND models developed based on HVS properties and characteristics ? "*: *Research Question* **Q2**. In particular, most of 3D-JND models were developed based on 2D-JND except that BJND [220] was constructed based on stereopair patterns used in psychophysical experiments, which is more reliable and closer to human binocular perception. However, BJND ignores the disparity effects of the visual stimuli on psychophysical experiments, making it less suitable for real-application. This inspired us to optimize BJND and thus propose a new 3D-JND model described in **Paper III**. Furthermore, we provided a thorough comparative analysis between 3D-JND models in terms of their suitable 3D formats, complexity, applications, advantages, and shortcomings (see Table 3.1). This answers the *Research Question* **Q2**: *" What are the advantages, drawbacks, and applications of these models ? "*.

Besides, to answer *"How to evaluate the performances of the 3D-JND models in order to select appropriate models for particular applications ? "*: *Research Question* **Q2**, we experimentally compared their performance with three solutions as follows:

1. Qualitatively and quantitatively evaluating the distortion tolerance ability, especially for image edges based on image from Middlebury stereo datasets [54, 217, 218].

2. Evaluating the accuracy of visibility thresholds estimation for each model by comparing the estimated JND values with JND values measured from psychophysical experiments. In addition, we created the 3D images used in psychophysical experiments based on 2D texture images from ETHZ dataset [219].

3. Estimating the prediction accuracy of a SIQA framework including the 3D-JND block.

Note that **Paper II** contributes to the abovementioned 1st and 2nd solutions and **Paper I** contributes to the 3rd solution. Both **Paper I** and **Paper II** determine the most important features that should be taken into account in 3D-JND models, and can help us to select appropriate models for QA and compression, and in the construction of more accurate and efficient 3D-JND models.

4.1.1.2 A new 3D-JND model

Based on the survey of the existing 3D-JND models in **Paper II**, we found that luminance adaptation and contrast masking are the most considered masking effects (related to monocular vision) in addition to binocular masking (related to binocular vision) for distortion in 3D images. Therefore, we designed in **Paper III** two stereopair patterns effectively simulating luminance adaptation and contrast masking, independently in the case of asymmetric noises in a stereopair. These patterns considered not only the visual field of fovea and the Percival's zone of comfort [232] with respect to binocular disparity, but also the randomness of noise location. The design of these stereopair patterns may motivate researchers to design other patterns considering more VM effects. Therefore, this can be one contribution of **Paper III**, and answers the *Research Question* **Q3**: *" How to develop a new reliable 3D-JND model accounting for HVS VM effects and depth information ? "*.

Additionally, we conducted psychophysical experiments using stereopair patterns to measure JND thresholds of one view that make the asymmetric noise binocularly visible for luminance adaptation and contrast masking experiments. This answers *" How to design the psychophysical experiment modeling VM effects and binocular disparity ? "*: *Research Question* **Q3**. The binocular JND thresholds indicated the inter-difference minimum between left and right views that human can just recognize.

Finally, we processed the psychophysical JND data to remove subjects- and samples-related outliers using statistical methods (see Sections 2.4.1 and 2.4.2) so as to obtain reliable JND data. The latter were used to construct a DBJND model inspired by BJND model [220]. As the psychophysical experiments were designed based on S3D images without occlusion effects, DBJND can estimate the visibility thresholds for NOPs. To propose a more common 3D-JND model, we used a reliable 2D-JND model [114] to compute the visibility thresholds for OPs. To further consider the visual sensitivity, a 3D-JND model was constructed by weighting the JND estimates with a 3D visual saliency map [221]. Accordingly, **Paper III** answers *" How to construct a 3D-JND model based on psychophysical data ? "*: *Research Question* **Q3**.

Unlike **Paper I** and **Paper II**, we evaluated the performance of the proposed 3D-JND model in terms of perceptual image quality with the same noise level based on subjective tests. This method can be used to either optimize 3D compression algorithms or increase the prediction accuracy of SIQA metrics.

4.1.2 Contributions to SIQA

Paper IV – Paper VII are main contributions to research in perceptual QA of 3D images. These papers aim to propose effective and efficient SIQA methods by investigating the HVS characteristics, binocular perception properties, and mechanism of the human 3D quality judgment. **Paper IV**, **Paper V**, and **Paper VII** contribute to FR-SIQA models. **Paper VI** and the work in [233] contribute to NR-SIQA models.

4.1.2.1 FR-SIQA

Paper IV predicts the overall 3D image quality combining JND-weighted cyclopean image quality with the disparity image quality. A 2D-IQA metric (i.e., UQI [147]) was used to assess the qualities of both cyclopean image and disparity image. This paper contributes to quality consideration of cyclopean image and of DsM for overall 3D image quality. In addition, based on **Paper IV**, we found that some 2D-IQA metrics (e.g., UQI [147] and VIF [150]) can perform quite well for symmetrically distorted stereopairs. This can help to propose a SIQA method that assesses 3D quality of symmetrically and asymmetrically distorted stereopairs, separately. Besides, a 2D-JND model was used in the proposed SIQA metric to increase the quality prediction accuracy. This can motivate us to apply the aforementioned 3D-JND model in this proposed SIQA framework in order to evaluate its performance. Finally, **Paper IV** also explored impacts of the disparity image quality and JND component on the

overall performance, and highlighted the importance of the disparity image quality. This answers the *Research Question* **Q5**: *" What are the most influential factors for 3D image quality and to which extent are they affected ? What are the binocular perception phenomena/effects ? And how do these effects impact the perceived quality of 3D images ? "*.

Similar to **Paper IV**, **Paper V** presents a SIQA metric accounting for both cyclopean image and disparity image qualities. In addition, this paper used both local entropy and visual saliency map of each view to accurately simulate the strength of views dominance of binocular rivalry phenomenon. This concludes that visual sensitivity of the HVS plays an important role on QA performance. This answers *" What are the most influential factors for 3D image quality and to which extent are they affected ? What are the binocular perception phenomena/effects ? And how do these effects impact the perceived quality of 3D images ? "* of *Research Question* **Q5**. Besides, **Paper IV** validated the effectiveness of the proposed SIQA metric based on three 3D-IQA databases, and compared with state-of-the-art in terms of overall performance, performance on individual distortion types and performance for symmetric and asymmetric distortions.

Based on **Paper IV** and **Paper V**, **Paper VII** proposes a general FR-SIQA framework that accounts for binocular-rivalry-inspired monocular image quality and JND-weighted cyclopean image quality. Based on this paper, we (1) understood different strategies of modeling visual stimulus strength, (2) investigated the importance of binocular-rivalry-inspired monocular quality and cycloepan quality on overall 3D quality, (3) evaluated quality prediction accuracy of the proposed SIQA model with different 2D-JND and 3D-JND, and (4) surveyed existing 3D-IQA databases, and provided comprehensive performance evaluation and exhaustive performance comparison. Thus **Paper VII** answers the *Research Question* **Q4**: *" How does the HVS judge image quality based on binocular perception ? "*, and **Q6**: *" What precise and reliable methodology for SIQA that accounts for both monocular and binocular influential factors ? And how do these factors affect jointly the overall 3D quality ? "*. In sum, **Paper VII** can further help the design of other SIQA metrics using appropriate binocular rivalry modeling approaches, and then evaluate metrics based on the described publicly available 3D quality databases.

4.1.2.2 NR-SIQA

FR-SIQA methods in **Paper IV**, **Paper V**, and **Paper VII** require the reference stereopairs, which are probably unavailable in real application. Therefore, inspired by a NR 2D-IQA approach [153], **Paper VI** proposes an opinion-aware NR-SIQA metric based on binocular-rivalry-inspired combination of monocular statistical features. SVR-based regression method was used for training and testing stages. This paper contributes to (1) combining the joint statistics of the GM and LoG into binocular statistical features based on a linear summation model with weights estimated by LoG responses of left and right images, (2) investigating the influences of appropriate image scale (for LoG-based visual stimulus strength modeling) on proposed metric performance.

To further increase the quality prediction accuracy, our work proposed in [233] improves the NR-SIQA metric from **Paper VI** by using multi-scale images, various strategies of modeling visual stimulus strength, and disparity information. Specifically, the strength of views dominance of the binocular rivalry phenomenon was estimated using LoG responses and local entropy (LE) maps of two images. LoG filter was applied in single-view image downsampled by factors 2 and 4, whereas LE map was applied in image with the original scale. As **Paper IV** showed the importance of disparity information in QA, we computed the absolute difference image between left and right views to imply disparity information [234]. This is because the ground truth disparity maps are not available and their estimation is probably error-prone. Compared to **Paper VI**, The SIQA method in [233] achieved higher correlation with human opinion scores, but low computational efficiency because of using LE-based binocular rivalry modeling. Both **Paper VI** and [233] answer *Research Question* **Q5**: *" What are the most influential factors for 3D image quality and to which extent are they affected ? What are the binocular perception phenomena/effects ? And how do these effects impact the perceived quality of 3D images ? "*.

4.2 Supplementary results

In this Section, we separately present supplementary results for **Paper III**, **Paper IV**, and **Paper VI**.

4.2.1 Supplementary results of Paper III

To further compare 3D-JND models in terms of noise masking ability, we conducted subjective tests to compare reference and distorted 3D images. The results in Table 4.1 show that the proposed DBJND outperforms other 3D-JND models. However, the proposed saliency-weighted stereoscopic JND (SSJND) performs less well, because the error introduced by the complicated image scene in subjective tests may probably decrease the saliency detection.

4.2.2 Supplementary results of Paper IV

The performance evaluation of the proposed SIQA method, and comparison to other methods on LIVE 3D-IQA Phase I database are shown in Table 4.2. The proposed method achieves the best overall performance among all the other 2D/3D IQA methods. Furthermore, we also examine the performance of the SIQA metrics for different distortions types. The proposed method outperforms the other methods for JPEG, JPEG 2000 (JP2K), and fast fading, and delivers competitive results for white noise (WN) and Gaussian blur. Overall, all 2D-based SIQA methods can achieve reasonably accurate quality prediction on LIVE 3D phase I database containing only symmetrically distorted stereopairs. In addition, Figure 4.1 and 4.2 depict scatter plots of predicted scores obtained by proposed SIQA metric versus DMOS on LIVE 3D phase I and II, respectively. It can be observed that

Table 4.1 – Subjective test scores: quality comparison between original 3D image and noisy 3D images produced by a 3D-JND model using 12 images from Middlebury Stereo Datasets. Note that DBJND and SSJND models are respectively our model without and with considering the visual saliency effect. The higher the average of the 3D-JND model is, the better the distortion masking ability the 3D-JND is.

3D image	**DBJND**		**SSJND**		**BJND**		**JJND**		**DJND**	
	Mean	p-value	Mean	p-value	Mean	p-value	Mean	p-value	Mean	p-value
Art	0.67	0.0001	0.78	0.0001	0.33	0.0001	1.11	0.0055	0.78	0.0001
Reindeer	0.11	0.0001	-0.22	0.0027	0.28	0.0001	2.28	0.0001	1.78	0.0001
Moebius	0.06	0.0001	0.50	0.0001	0.28	0.0001	1.50	0.0005	1.33	0.0018
Dolls	0	0.0001	0.11	0.0001	0.17	0.0001	1.17	0.0001	0.72	0.0001
Aloe	-0.11	0.0001	0.39	0.0001	0	0.0001	1.17	0.0004	0.61	0.0001
Baby2	0.44	0.0001	1.06	0.0001	0.22	0.0001	0.83	0.0002	1.50	0.0001
Midd2	0.33	0.0001	0.61	0.0001	0.50	0.0001	1.78	0.0001	1.44	0.0001
Plastic	0.17	0.0001	0.72	0.0001	0.61	0.0001	2.00	0.0001	1.89	0.0001
Motocycle	0.17	0.0001	0.33	0.0001	0.28	0.0001	0.94	0.0001	1.17	0.0055
Piano	0.39	0.0001	1.00	0.0001	-0.11	0.0144	2.00	0.0001	1.56	0.0001
Playroom	0.28	0.0001	0.22	0.0001	-0.11	0.0001	1.44	0.0002	1.39	0.0001
Playtable	0.50	0.0001	0.28	0.0001	0.22	0.0001	1.67	0.0001	0.89	0.0034
Average	**0.23**	**0.0001**	**0.44**	**0.0002**	**0.24**	**0.0011**	**1.38**	**0.0005**	**1.16**	**0.0008**

the scatter points are well concentrated around the fitting curves, which indicates a good correlation between the objective and subjective scores.

4.2.3 Supplementary results of Paper VI

Table 4.3 shows the performance on individual distortion types of the SIQA methods on LIVE 3D-IQA Phase II database. Experimental results demonstrate that the proposed model achieves high performance for JPEG, fast fading, and WN distortions, and yields promising results for JP2K and Gaussian blur distortions.

Tables 4.4 and 4.5 respectively show SROCC and RMSE values of the NR-SIQA methods on cross-database. Specifically, we tested the performance by training them on one database and testing on other databases. We can observe that the proposed SQSC-FW and SQSC-AW deliver competitive performance compared to other methods when using LIVE 3D II database for training. Furthermore, SQSC-AW outperforms most other methods when using Waterloo IVC II for training, especially used for testing LIVE 3D II.

4.3 Limitations and Shortcomings

This section describes some limitations and shortcomings of the work from each Paper, as well as some future trends in SIQA.

Table 4.2 – Performance of SIQA methods on LIVE 3D IQA database (Phase I). Italicized entries denote 2D-based IQA, and the results of the best-performing SIQA method are highlighted in boldface.

Distortion type	Criteria	*SSIM [148]*	*MS-SSIM [149]*	*FSIM [151]*	*VIF [150]*	*UQI [147]*	Wang [235]	Fezza [211]	Fezza [236]	Chen [214]	**Proposed**
WN	LCC	*0.944*	*0.952*	*0.931*	*0.930*	*0.927*	0.949	0.941	0.947	**0.955**	0.931
	SROCC	*0.939*	*0.942*	*0.929*	*0.931*	*0.926*	0.947	0.935	0.944	**0.948**	0.926
	RMSE	*5.500*	*5.070*	*6.094*	*6.103*	*6.240*	5.254	5.620	5.351	**4.963**	6.068
JPEG	LCC	*0.475*	*0.633*	*0.623*	*0.603*	*0.769*	0.473	0.274	0.706	0.527	**0.791**
	SROCC	*0.435*	*0.613*	*0.582*	*0.580*	*0.737*	0.450	0.246	0.657	0.521	**0.754**
	RMSE	*5.755*	*5.063*	*5.116*	*5.216*	*4.178*	5.762	6.289	4.632	5.557	**4.001**
JP2K	LCC	*0.858*	*0.930*	*0.908*	*0.888*	*0.944*	0.875	0.783	0.937	0.920	**0.953**
	SROCC	*0.857*	*0.892*	*0.905*	*0.902*	*0.910*	0.856	0.774	0.896	0.887	**0.911**
	RMSE	*6.663*	*4.752*	*5.424*	*5.959*	*4.270*	6.272	8.822	4.532	5.070	**3.927**
GB	LCC	*0.907*	*0.944*	*0.933*	***0.962***	*0.952*	0.893	0.908	0.934	0.943	0.957
	SROCC	*0.879*	*0.925*	*0.922*	***0.934***	*0.925*	0.871	0.867	0.909	0.924	0.926
	RMSE	*8.774*	*4.790*	*5.205*	***3.955***	4.451	6.512	6.058	5.173	4.813	4.182
FF	LCC	*0.670*	*0.803*	*0.815*	*0.862*	*0.879*	0.644	0.641	0.783	0.776	**0.885**
	SROCC	*0.584*	*0.722*	*0.729*	*0.804*	***0.833***	0.525	0.515	0.693	0.700	0.828
	RMSE	*9.277*	*7.405*	*7.199*	*6.306*	*5.925*	9.508	9.541	7.730	7.832	**5.780**
ALL	LCC	*0.877*	*0.856*	*0.915*	*0.925*	***0.943***	0.868	0.833	0.821	0.922	**0.943**
	SROCC	*0.877*	*0.824*	*0.928*	*0.920*	*0.937*	0.868	0.823	0.922	0.914	**0.939**
	RMSE	*7.889*	*8.472*	*6.614*	*6.230*	*5.478*	8.131	9.063	9.358	6.351	**5.468**

Table 4.3 – Overall performance and performances for different types of distortion of the SIQA methods on LIVE 3D-IQA Phase II database. The ranking 1st and 2nd for each criterion are highlighted with red and blue bold texts, respectively.

SIQA	**JPEG**		**JP2K**		**WN**		**GB**		**FF**	
method	SROCC	RMSE	SROCC	RMSE	SROCC	RMSE	SROCC	RMSE	SROCC	RMSE
Chen-FR [214]	0.843	3.865	0.814	5.562	0.940	3.368	**0.908**	3.747	0.884	4.966
Chen-NR [237]	**0.867**	3.342	0.867	4.298	0.950	3.531	0.900	4.725	**0.933**	4.180
SINQ [238]	0.839	3.476	**0.909**	**3.463**	**0.957**	**2.519**	**0.909**	**2.481**	0.924	**3.803**
SSQA [225]	0.858	**3.068**	**0.908**	**4.022**	0.940	3.536	0.901	**2.570**	0.924	3.879
SQSC-AW	**0.901**	**2.929**	0.864	4.152	**0.944**	**2.907**	0.882	3.107	**0.934**	**3.784**

Paper I gives an overview of existing 3D-JND models, and compares them adopting the QA framework described in [211]. Specifically, the performance (i.e., prediction accuracy) of the SIQA

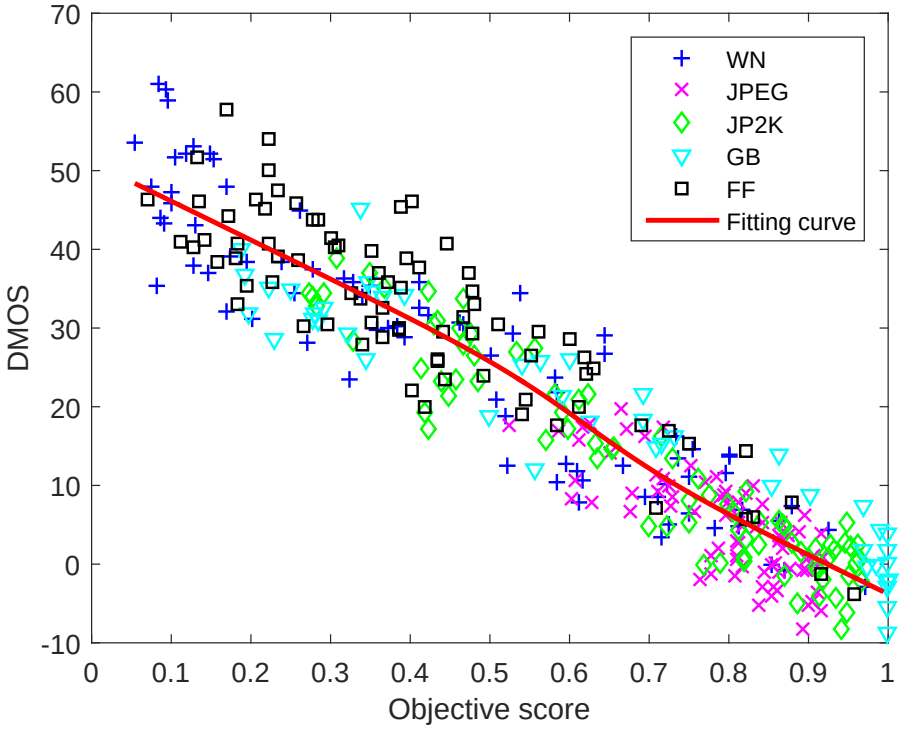


Figure 4.1 – Scatter distribution of predicted scores versus DMOS on LIVE 3D phase I.

Table 4.4 – SROCC values of the NR-SIQA methods on cross-database.

Training database	**Testing database**	Chen-NR [237]	SINQ [238]	SSQA [225]	**SQSC (FW)**	**SQSC (AW)**
LIVE 3D II	Waterloo IVC I	0.414	0.557	**0.653**	**0.561**	0.544
	Waterloo IVC II	0.491	0.439	**0.686**	**0.624**	0.493
Waterloo IVC II	LIVE 3D II	0.441	0.535	**0.669**	0.550	**0.729**
	Waterloo IVC I	0.823	**0.908**	**0.911**	0.899	0.904

Table 4.5 – RMSE values of the NR-SIQA methods on cross-database.

Training database	**Testing database**	Chen-NR [237]	SINQ [238]	SSQA [225]	**SQSC (FW)**	**SQSC (AW)**
LIVE 3D II	Waterloo IVC I	13.957	13.426	11.832	**10.048**	**10.765**
	Waterloo IVC II	16.404	17.092	14.118	**12.375**	**14.111**
Waterloo IVC II	LIVE 3D II	9.875	9.052	**8.159**	9.101	**7.347**
	Waterloo IVC I	8.750	6.331	**6.021**	6.252	**6.154**

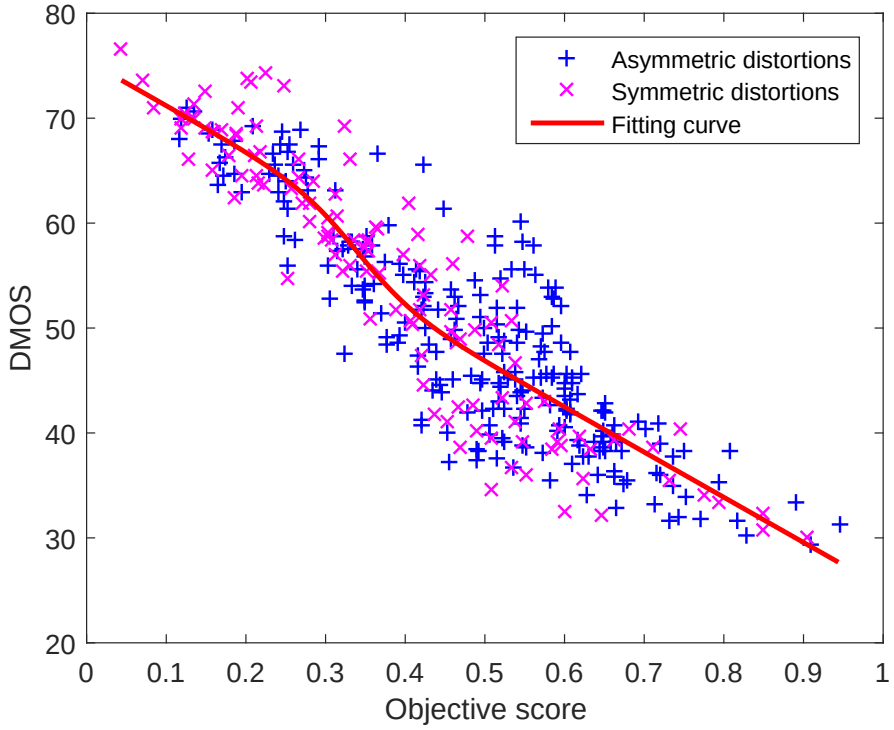


Figure 4.2 – Scatter distribution of predicted scores versus DMOS on LIVE 3D phase II.

metric integrating a block of 3D-JND showed the performance of the 3D-JND model. The SIQA metric performance was evaluated based on two 3D IQA databases including symmetrically and asymmetrically distorted stereopairs. However, this performance was evaluated using database containing only symmetric distortions, such as LIVE 3D Phase I database [222], and NBU 3D II database [229]. In addition, using only one QA framework including 3D-JND cannot generally compare the performance between 3D-JND models. Therefore, to provide a fair comparison between 3D-JND models, some other SIQA or stereoscopic video quality assessment (SVQA) approaches including 3D-JND should be used [106, 239–242].

Compared to **Paper I**, **Paper II** provides a more comprehensive comparative study on state-of-the-art 3D-JND models in terms of theoretical analysis and experiments. Although the distortion masking ability of each 3D-JND models was evaluated, perceptual quality of JND-noise-contaminated 3D images was not subjectively assessed [243]. In particular, a subjective test should be conducted to compare the noisy images by 3D-JND with the pristine images. The noisy images with higher quality score show that the corresponding 3D-JND model has better distortion masking ability. In addition, only seven texture images were used to create the 3D images used in psychophysical experiments in order to evaluate the estimation accuracy of each 3D-JND model. More various types of images should be employed to synthesize the 3D images for psychophysical experiments, because the real-

world scenes are complicated and diversified. For instance, we may use the images containing persons, objects, and animals with different strength of textures and edges. Furthermore, more appropriate disparity values should be considered in psychophysical experiments so as to measure more reliable JND thresholds. Besides, even if we controlled the duration of the psychophysical experiments, the visual fatigue caused by accommodation-vergence conflict may probably decrease the JND estimation accuracy.

Paper III designs the stereopair patterns used in psychophysical experiments in order to construct a new 3D-JND model. However, other VM effects (such as temporal masking) were not considered in these patterns. One challenge is how to create a stereopair pattern simulating multiple VM effects aiming to develop a more accurate 3D-JND model. Moreover, a few noise levels, luminance contrast and disparity values were used in psychophysical experiments due to limit of experiments duration. To construct a more reliable 3D-JND model, more appropriate noise levels, luminance contrast, and disparity values should be taken into account. In addition, we need to consider a trade-off between the number of visual stimulus attributes and experiments duration. Several trials of selecting appropriate noise increment/decrement steps are time-consuming and error-prone. Inappropriate noise variation steps may result in misjudgment of JND thresholds. It is very challenging to synthesize S3D patterns used in psychophysical experiments consider various VM effects and binocular disparity effect jointly. Besides, the performance of the proposed 3D-JND model was not investigated using SIQA framework as described in **Paper I**, or compression framework.

Paper IV and **Paper V** compute the 3D image quality combining cyclopean image quality and DsM quality. Although two metrics proposed in **Paper IV** and **Paper V** achieved the competitive performance compared with other SIQA metrics, the weights determining the relative importance of cyclopean image quality and DsM quality were not investigated. In addition, performance of the proposed SIQA method in **Paper IV** was not evaluated on other 3D-IQA databases containing only symmetrically distorted stereopairs (e.g., LIVE 3D Phase I database [222]). In **Paper V**, UQI and VIF were used to assess the monocular image quality for LIVE 3D Phase I and II databases ([214, 222]) and Waterloo IVC Phase I database [215], respectively. This limited the generality of SIQA metric because the metric using UQI/VIF may preform less well on other databases. Furthermore, the proposed metric showed low prediction accuracy for 3D images distorted with asymmetric distortion types in Waterloo IVC Phase I database.

Compared to **Paper IV** and **Paper V**, **Paper VII** used gradient magnitude similarity mean (GMSM) metric to compute monocular image quality. However, the proposed SIQA model in **Paper VII** need to select appropriate downsampling factor of GMSM for different 3D-IQA databases in order to achieve the promising performance. Moreover, impact of the visual saliency component, which is used in view dominance strength modeling, on QA performance was not discussed in **Paper VII**. In

addition, the performance of the SIQA framework including the 3D-JND model proposed in **Paper III** was not evaluated and compared with other 2D-JND and 3D-JND models.

Paper VI [24] and the work in [233] present NR-SIQA approaches based on binocular-rivalry-inspired combination of monocular statistical features, and/or disparity information. Performance of these proposed SIQA models were not evaluated on symmetric-distortion-related 3D-IQA databases (e.g., LIVE 3D Phase I [222], and NBU 3D II [229]). Furthermore, other machine learning-based regression methods (such as artificial NN [174] and RF [175]) were not used to compare with SVR. In addition, the computational complexity of the proposed metrics should be investigated and compared with other efficient NR-SIQA metrics. Besides, both SIQA metrics perform less well for blurring artifacts (e.g., Gaussian blur and JP2K) in 3D images because the smoothing effects in LoG response may probably decrease the accuracy of mincing the stimulus strength. In [233], the proposed model accounting for disparity features had not compared to this without disparity features in terms of prediction accuracy. Moreover, as the proposed SIQA metric was developed based on multiple scales, the performance of the metric with a single scale should be further investigated.

In addition to the above-described limitations, we discuss the challenges and difficulties one may face in 3D quality assessment, and try to give some possible solutions and future work.

Firstly, despite several publicly available 3D IQA databases have been proposed, they are less comprehensive compared with 2D IQA database, and are created based on different protocols. For example, existing 3D IQA databases (e.g., LIVE 3D [214], Waterloo-IVC [215]) cover a few distortion types in contrast with 2D ones (refer to [136]). Furthermore, some issues are probably involved in creating 3D IQA databases such as acquisition protocols, depth and visual discomfort control, image formats, and asymmetric distortion control. Besides, current 3D IQA databases are established under controlled conditions by introducing the graded simulated distortions onto images. However, real-world 3D images have large content variation, and may be altered by complex mixtures of diverse distortions, which are not typically modeled by the synthetic distortions considered in existing databases. Only NBU-MDSID Phase I and II databases [230, 244] have considered the effects of multiple distortions simultaneously presented on images. Similar to some extensive 2D IQA databases [140, 141], future new 3D IQA databases, which contain a large amount of images with a diverse range of 3D content and multiple distortions types, are thus needed to achieve a fair performance evaluation of quality metrics. Besides, collecting means opinion scores based on subject test is costly when the samples number is large. Therefore, the robust and efficient performance evaluation criteria could be used for future studies [141].

Secondly, most state-of-the-art SIQA models focus on the judgment of the quality discrepancy between the reference and distorted 3D images. However, when 3D content is presented to viewers, they mainly care about the quality of experience (QoE) [245, 246] influenced by diverse perceptual factors such as image quality, image aesthetics [247], depth quality [248], visual comfort [12, 249], naturalness [250], etc. In particular, the geometry distortion in the depth map may probably influence

QoE [251]. Image aesthetics, referring to the experience of beauty for viewers when perceiving an image, may be an interesting factor affecting 3D QoE [252]. Accordingly, effectively predicting the QoE of real-world 3D image/video would be a more appropriate way to evaluate the overall quality [253]. A major challenge is on how to model the relationships between the above QoE-related factors and to combine them all together to achieve final QoE prediction. In addition, some work may explore how to develop approaches to improve QoE by estimating the optimal capture and display parameters of 3D scene content.

Finally, even though several machine learning based NR-SIQA models have achieved outstanding performance on specific SIQA databases [230, 254], there still exists some deficiencies and issues for real applications. These models need large training database (containing diverse contents) to sufficiently represent the real-world natural images. Therefore, the performance of these SIQA models is database- or distortion types-dependent [255, 256]. Moreover, several SIQA models may suffer from the over-fitting problem due to the training dataset, and thus one can not fairly evaluate their generalization ability. Thus, it is difficult to choose the right size of the training and the testing samples on one specific dataset. Even though learning from cross-databases may avoid this shortcoming [257], accurately predicting 3D quality on cross-databases is still challenging [225, 258]. Specifically, a small number of samples in the training stage may result in under-fitting. Besides, most NR-SIQA approaches often involve supervised learning and/or manual parameters adjustments to deliver promising performance for specific databases. These approaches could not generally deal with real-world images with richer content variations and various quality degradations. Moreover, a regression model is often trained with parameters to map the extracted image features to subjective scores. Optimizing these parameters for the best quality prediction is challenging, such as the parameters of SVR [209]. A more effective regression model with a few parameters could be used to improve the generalization ability of the quality metrics [259]. Meanwhile, efficient and robust pooling strategies of monocular and binocular features/qualities for overall 3D quality formulation can be used to increase the quality prediction accuracy [244, 260]. Besides, most existing NR-SIQA methods are opinion-aware, which require human subjective scores to train the regression model. However, let remainder that obtaining subjective scores based on subjective experiments are often expensive and time-consuming. Furthermore, the opinion-aware methods highly depend on image content, distortion types, and relative distortion level on the left and right views. Consequently, inspired by existing blind 2D quality metrics [205, 206], defining how to design reliable opinion-unaware NR-SIQA metrics is a worthwhile and challenging future work.

Chapter 5

General Conclusions and Perspectives

5.1 Conclusions

This thesis focuses on the investigation of the HVS characteristics and properties of the binocular perception, which are taken into account in order to establish the methodologies of 3D image visibility threshold estimation and SIQA. The work described here is divided into two parts.

In the first part, we explored the human visual sensitivity to detect the changes in an image. If these changes are lower than JND threshold the HVS cannot detect them. While 2D-JND models have been extensively investigated and advanced in the recent years [101], research on 3D-JND is still at an early stage. Therefore, we first provided a comprehensive overview of state-of-the-art 3D-JND models, and a deep comparison between them from various aspects including advantages, shortcomings, and applications. Moreover, the performance of the 3D-JND models were experimentally compared in terms of distortion masking ability, estimation accuracy and prediction accuracy of SIQA framework including 3D-JND block. This overview and comparative study of 3D-JND models are related to **Paper I** and **Paper II**. The major contributions of these papers are to propose design methodologies to compare existing 3D-JND models, and analyze these models. This can help other researchers design more accurate and particular 3D-JND models for image/video processing tasks. Besides, based on the survey of existing 3D-JND models, we proposed a new 3D-JND model accounting for monocular VM effects, binocular disparity, and visual attention. To achieve this, we designed and conducted psychophysical experiments as described in **Paper III**. This new model, which can be further applied in 3D compression and SIQA, is one of the core contributions for this thesis.

In the second part of the dissertation, we explored the SIQA methodology considering binocular perception behaviors (i.e., binocular fusion and binocular rivalry) and visual sensitivity of the HVS, which can affect human quality judgment for 3D images [65, 261]. We developed 3D quality metrics

from two different aspects. First, we proposed two FR-SIQA approaches. One was developed based on the binocular-based cyclopean image quality and disparity image quality (see **Paper IV** and **Paper V**), the other accounted for binocular-based cyclopean image and binocular-rivalry-inspired monocular image quality (see **Paper VII**). **Paper VII** is one core contribution of this dissertation, because it (1) presented a new SIQA framework considering different binocular rivalry modeling strategies, (2) investigated the importance of cyclopean image quality and combined monocular 2D quality, and (3) provided an overview of existing 3D-IQA databases, and compared the proposed metrics with various competitive SIQA metrics in terms of prediction accuracy based on these databases. Second, we proposed two NR-SIQA approaches based on the combination of monocular statistical features of the image local contrast with (see **Paper VI**) or without disparity information [233]. **Paper VI** and [233] contribute to validate the effectiveness of using image LoG responses to model the visual stimulus strength of the binocular rivalry phenomenon, and highlight the impacts of image scale on 3D-IQA.

Measuring the depth impairment is crucial for developing reliable 3D-JND and SIQA models. In particular, degradation on disparity/depth maps may cause visual discomfort/fatigue that definitely influences the overall quality of the 3D images [248]. On the one hand, modeling 3D scene statistics with regards to depth perception has not been deeply explored to date. Moreover, the relationship between the 2D quality and depth information/quality in SIQA task remains poorly understood. On the other hand, the effect of disparity masking on 3D visibility threshold has been validated in our proposed 3D-JND model, even though this effect is relatively less than luminance adaptation and contrast masking. In sum, depth impairment has effects on 3D quality assessment and 3D-JND modeling depending on the relative degradation strength between depth information/map and 2D images.

5.2 Perspectives

The research work of this thesis can be extended from various perspectives, some of which are described below.

Existing 3D-JND models were compared using a SIQA model including 3D-JND in **Paper I**. There exist other SIQA approaches employing BJND to determine the visibility thresholds for NOPs in a stereopair [106, 239–242]. Therefore, the idea is to replace BJND by other 3D-JND model in these SIQA models to provide a more extensive comparison between 3D-JND models. In addition, aiming to optimize perceptual quality and bitrate saving in 3D compression, applying the proposed 3D-JND model (see **Paper III**) in 3D high efficiency video coding will be worthwhile and very interesting. Besides, we will propose a new framework that uses the 3D-JND model to enhance the perceptual quality of 3D image/video.

Paper III can be extended by considering other types of distortions (e.g., "blurring" and "blockiness") in addition to noise artifacts in psychophysical experiments. Furthermore, in order to propose a

more general and accurate 3D-JND model, we will design the psychophysical experiments accounting for both symmetric and asymmetric distortions in stereopairs. This new model can probably improve the prediction accuracy for symmetrically distorted stereopairs using SIQA metrics. Finally, the design of a stereopairs pattern used in psychophysical experiments jointly considering different important VM effects and binocular disparity will be challenging and is still an open issue. Moreover, to evaluate the performance of future proposed 3D-JND models, we need to create a 3D-JND-related database containing original and distorted 3D images associating with subjective JND values (considered as ground truth), as well as objective JND values estimated based on existing 3D-JND models.

Selecting the appropriate image scale of the proposed NR-SIQA metrics in **Paper VI**, **Paper VII**, and [233] are time-consuming and error-prone, the future work can focus on development of a learning-based method to automatically select the best single/multiple scales. Although other regression methods (e.g., artificial neural network [174] and random forest [175]) can be used in our proposed NR-SIQA models [24, 233], obtaining the subjective scores are obviously costly. Consequently, opinion-unaware and distortion-unaware NR-SIQA approaches considering the binocular visual properties based on deep neural networks are mandatory for real industrial applications. Besides, in order to guarantee good QoE [245, 262], a reliable methodology of 3D image QoE assessment can be further investigated, because QoE depends on various factors including 2D image quality, depth quality, visual comfort, naturalness and immersive sensation [251, 263]. Finally, SVQA advanced slowly over the last decade due to the complexity of video spatial and temporal features, and their interaction in binocular perception. Our proposed SIQA frameworks in this thesis could be extended in SVQA for real-time QA and distortion optimization.

In addition to the above-mentioned future work, several long-term perspectives are listed as follows:

- **Binocular vision modeling**: Although there exits several binocular combination models [264, 265], information on how to apply the appropriate models in SIQA task remains an open question [266]. In addition, much more efforts should be made to deeply understand and effectively model the binocular vision behaviors (i.e., binocular fusion, binocular rivalry, and binocular suppression) and their interaction in with human 3D quality judgment [267, 268]. Finally, more work for SIQA should be done to jointly model the scene statistics of depth information and natural images [269].

- **Learning-based 3D-JND models**: The current 3D-JND models are mainly based on 2D-JND and psychophysical experiments. To accurately determine the visibility thresholds due to multiple distortions in the real-world 3D image, learning-based JND modeling could be a worthwhile research direction [270]. For instance, an interesting work in [271] studied how to estimate the just noticeable distortion thresholds from image quality scores. Besides, the creation of a reliable stereoscopic images databases with 3D-JND ground truth is important and imperative to evaluate the estimation accuracy of the existing 3D-JND models. Finally, it may be worthwhile to investigate how to directly compare state-of-the-art 3D-JND models without

conducting subjective tests [272].

- **Perception-driven omnidirectional (360-degree) content processing**: Virtual reality (VR) contents are in the form of 360-degree image/video. VR multimedia technology is becoming very popular because of providing a more interactive and immersive viewing experience to viewers. Accordingly, future research directions with respect to VR content may include: (1) JND models of 360-degree image/video, (2) quality or QoE assessment approaches for 360-degree image/video [273, 274], (3) visual saliency models of 360-degree image/video [275, 276] and (4), 360-degree image/video databases with ground-truth quality or QoE [277, 278].

Bibliography

[1] Bernard Mendiburu. *3D movie making: stereoscopic digital cinema from script to screen*. Focal press, 2012. 3

[2] Liang Zhang, Carlos Vazquez, and Sebastian Knorr. 3D-TV content creation: automatic 2D-to-3D video conversion. *IEEE Transactions on Broadcasting*, 57(2):372–383, 2011. 3, 4

[3] Anthony Vetro, Alexis M Tourapis, Karsten Muller, and Tao Chen. 3d-tv content storage and transmission. *IEEE Transactions on Broadcasting*, 57(2):384–394, 2011. 3, 4

[4] Wa James Tam, Filippo Speranza, Sumio Yano, Koichi Shimono, and Hiroshi Ono. Stereoscopic 3D-TV: visual comfort. *IEEE Transactions on Broadcasting*, 57(2):335–346, 2011. 3

[5] Petronel Bigioi, George Susanu, Igor Barcovschi, Piotr Stec, Larry Murray, Alexandru Drimbarean, and Peter Corcoran. Stereoscopic (3D) panorama creation on handheld device, June 16 2011. US Patent App. 12/879,003. 3

[6] Jonas Schild, Joseph LaViola, and Maic Masuch. Understanding User Experience in Stereoscopic 3D Games. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '12, pages 89–98, New York, NY, USA, 2012. ACM. ISBN 978-1-4503-1015-4. doi: 10.1145/2207676.2207690. URL http://doi.acm.org/10.1145/2207676.2207690. 3

[7] Hannes Kaufmann and Dieter Schmalstieg. Mathematics and geometry education with collaborative augmented reality. In *ACM SIGGRAPH 2002 Conference Abstracts and Applications*, pages 37–41, New York, NY, USA, 2002. ACM. 3

[8] Manisha Mistry, Victoria A Roach, and Timothy D Wilson. Application of stereoscopic visualization on surgical skill acquisition in novices. *Journal of surgical education*, 70(5):563–570, 2013. 3

[9] Chaminda TER Hewage and Maria G Martini. Quality of experience for 3D video streaming. *IEEE Communications Magazine*, 51(5):101–107, 2013. 3

[10] Peter A Howarth. Potential hazards of viewing 3-D stereoscopic television, cinema and computer games: A review. *Ophthalmic and Physiological Optics*, 31(2):111–122, 2011. 3

[11] Monika Pölönen, Toni Järvenpää, and Beatrice Bilcu. Stereoscopic 3D entertainment and its effect on viewing comfort: Comparison of children and adults. *Applied Ergonomics*, 44(1):151–160, 2013. 3

[12] Matthieu Urvoy, Marcus Barkowsky, and Patrick Le Callet. How visual fatigue and discomfort impact 3D-TV quality of experience: a comprehensive review of technological, psychophysical, and psychological factors. *annals of telecommunications-annales des télécommunications*, 68 (11-12):641–655, Dec. 2013. 3, 57

[13] P. Rotter. Why did the 3D revolution fail?: the present and future of stereoscopy [Commentary]. *IEEE Technology and Society Magazine*, 36(1):81–85, March 2017. ISSN 0278-0097. doi: 10.1109/MTS.2017.2654294. 3

[14] Shunnan Yang, Tawny Schlieski, Brent Selmins, Scott C Cooper, Rina A Doherty, Philip J Corriveau, and James E Sheedy. Stereoscopic viewing and reported perceived immersion and symptoms. *Optometry and vision science*, 89(7):1068–1080, 2012. 4

[15] Syed Ali Arsalan Naqvi, Nasreen Badruddin, Aamir Saeed Malik, Wan Hazabbah, and Baharudin Abdullah. Does 3D produce more symptoms of visually induced motion sickness? In *35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 6405–6408, July 2013. 4

[16] Marc Lambooij, Marten Fortuin, Ingrid Heynderickx, and Wijnand IJsselsteijn. Visual discomfort and visual fatigue of stereoscopic displays: A review. *Journal of Imaging Science and Technology*, 53(3):30201–1–30201–14, 2009. 4

[17] Jing Li, Marcus Barkowsky, and Patrick Le Callet. Visual discomfort of stereoscopic 3D videos: Influence of 3D motion. *Displays*, 35(1):49–57, 2014. 4

[18] Y. Fan, M. Larabi, F. Alaya Cheikh, and C. Fernandez-Maloigne. A Survey of Stereoscopic 3D Just Noticeable Difference Models. *IEEE Access*, 7:8621–8645, 2019. ISSN 2169-3536. doi: 10.1109/ACCESS.2018.2887276. 5

[19] Yu Fan, Mohamed-Chaker Larabi, Faouzi Alaya Cheikh, and Christine Fernandez-Maloigne. On the Performance of 3D Just Noticeable Difference Models. In *IEEE International Conference on Image Processing (ICIP)*, pages 1017–1021, September 2016. doi: 10.1109/ICIP.2016.7532511. 5

[20] Yu Fan, Mohamed-Chaker Larabi, Faouzi Alaya Cheikh, and Christine Fernandez-Maloigne. Just Noticeable Difference Model for Asymmetrically Distorted Stereoscopic Images. In *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2019. Accepted. 5

[21] Yu Fan, Mohamed-Chaker Larabi, Faouzi Alaya Cheikh, and Christine Fernandez-Maloigne. Stereoscopic Image Quality Assessment based on Monocular and Binocular Visual Properties. *Journal of Visual Communication and Image Representation*, 2018. Submitted. 5

[22] Yu Fan, Mohamed-Chaker Larabi, Faouzi Alaya Cheikh, and Christine Fernandez-Maloigne. Stereoscopic Image Quality Assessment based on the Binocular Properties of the Human Visual System. In *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 2037–2041, March 2017. doi: 10.1109/ICASSP.2017.7952514. 5

[23] Yu Fan, Mohamed-Chaker Larabi, Faouzi Alaya Cheikh, and Christine Fernandez-Maloigne. Full-Reference Stereoscopic Image Quality Assessment account for Binocular Combination and Disparity Information. In *IEEE International Conference on Image Processing (ICIP)*, pages 760–764, September 2017. doi: 10.1109/ICIP.2017.8296383. 5

[24] Yu Fan, Mohamed-Chaker Larabi, Faouzi Alaya Cheikh, and Christine Fernandez-Maloigne. No-Reference Quality Assessment of Stereoscopic Images Based on Binocular Combination of Local Features Statistics. In *IEEE International Conference on Image Processing (ICIP)*, pages 3538–3542, October 2018. doi: 10.1109/ICIP.2018.8451490. 5, 57, 61

[25] J. Kim and S. Lee. Deep learning of human visual sensitivity in image quality assessment framework. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1969–1977, July 2017. doi: 10.1109/CVPR.2017.213. 6

[26] Key structures of the eye and the role they paly in vision. `http://cerritos.instructure.com/courses/82/pages/seeing`. vii, 10

[27] William F Schreiber. *Fundamentals of electronic imaging systems: some aspects of image processing*, volume 15. Springer Science & Business Media, 2012. 9

[28] HR Wu, KR Rao, Makoto Miyahara, and Ryoichi Kawada. Philosophy of picture quality scale. In *Digital Video Image Quality and Perceptual Coding*, pages 181–224. CRC/Taylor & Francis, Boca Raton, Fla., 1 edition, 2005. 9

[29] Clyde W Oyster and Nancy Haver. *The human eye: structure and function*, volume 1. Sinauer Associates, Sunderland, Mass., 1999. 9

[30] AM Derrington and P Lennie. Spatial and temporal contrast sensitivities of neurones in lateral geniculate nucleus of macaque. *The Journal of physiology*, 357(1):219–240, 1984. 9

[31] Andrew M Derrington, John Krauskopf, and Peter Lennie. Chromatic mechanisms in lateral geniculate nucleus of macaque. *The Journal of physiology*, 357(1):241–265, 1984. 9

[32] Klaus Wunderlich, Keith A Schneider, and Sabine Kastner. Neural correlates of binocular rivalry in the human lateral geniculate nucleus. *Nature neuroscience*, 8(11):1595, 2005. 9

[33] John-Dylan Haynes, Ralf Deichmann, and Geraint Rees. Eye-specific effects of binocular rivalry in the human lateral geniculate nucleus. *Nature*, 438(7067):496, 2005. 9

[34] Atanas Boev, Maija Poikela, Atanas Gotchev, and Anil Aksay. Modelling of the stereoscopic HVS. *Report on Mobile 3DTV*, 2009. Available: `http://sp.cs.tut.fi/mobile3dtv/results/`. vii, 10

[35] Ian P Howard and Brian J Rogers. *Perceiving in depth, Vol. 1: basic mechanisms*. Oxford University Press, 2012. 10

[36] Ian P Howard and Brian J Rogers. *Perceiving in depth, Vol. 3: other mechanisms of depth perception*. Oxford University Press, 2012. 10

[37] Ian P Howard and Brian J Rogers. *Perceiving in depth, Vol. 2: Stereoscopic vision*. Oxford University Press, 2012. 10, 15

[38] Stanley Coren, Lawrence M Ward, and James T Enns. *Sensation and perception*. NJ:Wiley, Hoboken, 6 edition, 2004. 10

[39] Stephan Reichelt, Ralf Häussler, Gerald Fütterer, and Norbert Leister. Depth cues in human visual perception and their realization in 3D displays. In *Proc.SPIE*, volume 7690, pages 7690–7690–12, 2010. 11

[40] Nicholas Wade and Mike Swanston. *Visual perception: An introduction*. Psychology Press, London, England, 3rd ed. edition, 2013. 11

[41] Robert P O'Shea, Shane G Blackburn, and Hiroshi Ono. Contrast as a depth cue. *Vision research*, 34(12):1595–1604, 1994. 11

[42] Philip J Kellman and Thomas F Shipley. A theory of visual interpolation in object perception. *Cognitive psychology*, 23(2):141–221, 1991. 12

[43] Walter C Gogel. Size cue to visually perceived distance. *Psychological Bulletin*, 62(4):217, 1964. 12

[44] Ken-Ichiro Tsutsui, Hideo Sakata, Tomoka Naganuma, and Masato Taira. Neural correlates for perception of 3D surface orientation from texture gradient. *Science*, 298(5592):409–412, 2002. 13

[45] Pascal Mamassian, David C Knill, and Daniel Kersten. The perception of cast shadows. *Trends in cognitive sciences*, 2(8):288–295, 1998. 13

[46] George Mather. The use of image blur as a depth cue. *Perception*, 26(9):1147–1158, 1997. 14

[47] Brian Rogers and Maureen Graham. Motion parallax as an independent cue for depth perception. *Perception*, 8(2):125–134, 1979. 15

[48] Tsunehiro Takeda, Keizo Hashimoto, Nobuyuki Hiruma, and Yukio Fukui. Characteristics of accommodation toward apparent depth. *Vision Research*, 39(12):2087–2097, 1999. 15

[49] JEW Mayhew and HC Longuet-Higgins. A computational model of binocular depth perception. *Nature*, 297(5865):376, 1982. 15

[50] Andrew J Parker. Binocular depth perception and the cerebral cortex. *Nature Reviews Neuroscience*, 8(5):379, 2007. 15

[51] Neil A. Dodgson. Variation and extrema of human interpupillary distance. In *Proc.SPIE*, volume 5291, pages 5291 – 5291 – 11, 2004. 15

[52] Ning Qian. Binocular disparity and the perception of depth. *Neuron*, 18(3):359–368, 1997. 15

[53] Ian P Howard and Brian J Rogers. *Binocular vision and stereopsis*. Oxford University Press, 1996. 15

[54] Heiko Hirschmuller and Daniel Scharstein. Evaluation of cost functions for stereo matching. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1–8, June 2007. vii, 16, 37, 38, 48

[55] Lawrence Stark, Robert V Kenyon, VV Krishnan, and Kenneth J Ciuffreda. Disparity vergence: a proposed name for a dominant component of binocular vergence eye movements. *American journal of optometry and physiological optics*, 57(9):606–609, 1980. 15

[56] Mark S. Hughes. Dictionary of eye terminology. *Archives of Ophthalmology*, 109(9):1208–1208, 09 1991. 15

[57] Clifton M Schor and John C Kotulak. Dynamic interactions between accommodation and convergence are velocity sensitive. *Vision Research*, 26(6):927–942, 1986. 15

[58] Clifton Schor. The influence of interactions between accommodation and convergence on the lag of accommodation. *Ophthalmic and Physiological Optics*, 19(2):134–150, 1999. 15

[59] David M Hoffman, Ahna R Girshick, Kurt Akeley, and Martin S Banks. Vergence–accommodation conflicts hinder visual performance and cause visual fatigue. *Journal of vision*, 8(3):33–33, 2008. 15

[60] Peter A Howarth. The geometric horopter. *Vision research*, 51(4):397–399, 2011. 16

[61] Julesz Bela. *Foundations of cyclopean perception.* U. Chicago Press, Oxford, England, 1971. 16

[62] FH Verhoeff. Panum's areas and some other prevailing misconceptions concerning binocular vision. *Transactions of the American Ophthalmological Society*, 57:37, 1959. 16

[63] Jacek Turski. On binocular vision: the geometric horopter and cyclopean eye. *Vision research*, 119:73–81, 2016. 16

[64] Kenneth N Ogle. *Researches in binocular vision.* WB Saunders, Oxford, England, 1950. 16

[65] Willem JM Levelt. *On binocular rivalry.* PhD thesis, 1965. Available: https://pure.mpg.de/rest/items/item_77195/component/file_2424565/content. 17, 59

[66] Willem JM Levelt. The alternation process in binocular rivalry. *British Journal of Psychology*, 57(3-4):225–238, 1966. 17

[67] Jeremy M Wolfe. Stereopsis and binocular rivalry. *Psychological review*, 93(3):269, 1986. 17

[68] Jan Brascamp, Hansem Sohn, Sang-Hun Lee, and Randolph Blake. A monocular contribution to stimulus rivalry. *Proceedings of the National Academy of Sciences of the USA*, 110(21):8337–8344, 2013. 17

[69] Randolph Blake and Nikos K Logothetis. Visual competition. *Nature Reviews Neuroscience*, 3 (1):13–21, 2002. 17

[70] Haldun M Ozaktas and Levent Onural. *Three-Dimensional Television: Capture, Transmission, Display.* Springer Science & Business Media, Berlin, 2007. 17

[71] A. Smolic, K. Mueller, P. Merkle, P. Kauff, and T. Wiegand. An overview of available and emerging 3D video formats and depth enhanced stereo as efficient generic solution. In *Picture Coding Symposium*, pages 1–4, 2009. 17, 18

[72] Anthony Vetro. Representation and coding formats for stereo and multiview video. In *Intelligent Multimedia Communication: Techniques and Applications*, pages 51–73. Springer, 2010. 17

[73] Anthony Vetro, Thomas Wiegand, and Gary J Sullivan. Overview of the stereo and multiview video coding extensions of the H.264/MPEG-4 AVC standard. *Proceedings of the IEEE*, 99(4): 626–642, 2011. 17

[74] K. Müller, P. Merkle, and T. Wiegand. 3-D video representation using depth maps. *Proceedings of the IEEE*, 99(4):643–656, April 2011. 17, 18

[75] Liwei He Jonathan Shade, Steven Gortler and Richard Szeliski. Layered depth images. In *Proceedings of the 25th annual Conference on Computer Graphics and Interactive Techniques*, pages 231–242. ACM Press, July 1998. 18

[76] C Lawrence Zitnick, Sing Bing Kang, Matthew Uyttendaele, Simon Winder, and Richard Szeliski. High-quality video view interpolation using a layered representation. In *ACM Transactions on Graphics*, volume 23, pages 600–608, Aug. 2004. 18

[77] Philip Benzie, John Watson, Phil Surman, Ismo Rakkolainen, Klaus Hopf, Hakan Urey, Ventseslav Sainov, and Christoph Von Kopylow. A survey of 3DTV displays: techniques and technologies. *IEEE Transactions on Circuits and Systems for Video Technology*, 17(11):1647–1658, 2007. 18

[78] Philipp Merkle, Aljoscha Smolic, Karsten Muller, and Thomas Wiegand. Multi-view video plus depth representation and coding. In *IEEE International Conference on Image Processing (ICIP)*, volume 1, pages I–201–I–204, Sep. 2007. 18

[79] Karsten Müller, Aljoscha Smolic, Kristina Dix, Philipp Merkle, and Thomas Wiegand. Coding and intermediate view synthesis of multiview video plus depth. In *IEEE International Conference on Image Processing (ICIP)*, pages 741–744, 2009. 18

[80] WHA Bruls, Chris Varekamp, R Klein Gunnewiek, Bart Barenbrug, and Amaud Bourge. Enabling introduction of stereoscopic (3D) video: Formats and compression standards. In *IEEE International Conference on Image Processing(ICIP)*, volume 1, pages 89–92, Sep. 2007. 18

[81] George Mather. *Foundations of sensation and perception*. Psychology Press, London, 2016. 18

[82] Nicolas S Holliman, Neil A Dodgson, Gregg E Favalora, Lachlan Pockett, et al. Three-dimensional displays: a review and applications analysis. *IEEE transactions on Broadcasting*, 57(2):362, 2011. 18, 19

[83] Hakan Urey, Kishore V Chellappan, Erdem Erden, and Phil Surman. State of the art in stereoscopic and autostereoscopic displays. *Proceedings of the IEEE*, 99(4):540–555, 2011. 18, 19

[84] Jisoo Hong, Youngmin Kim, Hee-Jin Choi, Joonku Hahn, Jae-Hyeung Park, Hwi Kim, Sung-Wook Min, Ni Chen, and Byoungho Lee. Three-dimensional display technologies of recent interest: principles, status, and issues. *Appl. Opt*, 50(34):H87–H115, 2011. 18

[85] John P McIntire, Paul R Havig, and Eric E Geiselman. Stereoscopic 3D displays and human performance: A comprehensive review. *Displays*, 35(1):18–26, 2014. 18

[86] Jason Geng. Three-dimensional display technologies. *Advances in optics and photonics*, 5(4): 456–535, 2013. 18, 20

[87] Luc Froehly, S Nieto Martin, T Lasser, C Depeursinge, and Florian Lang. Multiplexed 3D imaging using wavelength encoded spectral interferometry: a proof of principle. *Optics Communications*, 222(1-6):127–136, 2003. 19

[88] Jin Hee Jung. Polarized stereoscopic display apparatus and manufacturing method thereof, August 19 2008. US Patent 7,414,782. 19

[89] Michael G. Robinson Gary D. Sharp. Enabling stereoscopic 3D technology. In *Proc.SPIE*, volume 6490, page 64900X, 2007. 19

[90] Selig Hecht and Emil L Smith. Intermittent stimulation by light: VI. Area and the relation between critical frequency and intensity. *The Journal of general physiology*, 19(6):979–989, 1936. 19

[91] J Harrold, DJ Wilkes, and GJ Woodgate. Switchable 2D/3D display–solid phase liquid crystal microlens array. In *Proc. IDW*, volume 11, pages 1495–1496, 2004. 19

[92] Alexander Schmidt and Armin Grasnick. Multiviewpoint autostereoscopic dispays from 4D-Vision GmbH. In *Proc.SPIE*, volume 4660, pages 4660–4660–10, 2002. 19

[93] Gregg E Favalora. Volumetric 3D displays and application infrastructure. *Computer*, 38(8): 37–44, 2005. 19

[94] R Häussler, A Schwerdtner, and N Leister. Large holographic displays as an alternative to stereoscopic displays. In *Proc.SPIE*, volume 6803, pages 6803–6803–9, 2008. 19

[95] Fahri Yaraş, Hoonjong Kang, and Levent Onural. State of the art in holographic displays: a survey. *Journal of display technology*, 6(10):443–454, 2010. 19

[96] DWF Van Krevelen and Ronald Poelman. A survey of augmented reality technologies, applications and limitations. *International journal of virtual reality*, 9(2):1, 2010. 20

[97] Mark Billinghurst, Adrian Clark, Gun Lee, et al. A survey of augmented reality. *Foundations and Trends® in Human–Computer Interaction*, 8(2-3):73–272, 2015. 20

[98] Julia Diemer, Georg W Alpers, Henrik M Peperkorn, Youssef Shiban, and Andreas Mühlberger. The impact of perception and presence on emotional reactions: a review of research in virtual reality. *Frontiers in psychology*, 6:26, 2015. 20

[99] Z.L. Lu and B. Dosher. *Visual Psychophysics: From Laboratory to Theory*. MIT Press, 2013. ISBN 9780262019453. URL https://books.google.no/books?id=nYr6AQAAQBAJ. 21

[100] G.A. Gescheider. *Psychophysics: The Fundamentals*. Mahwah, NJ: Lawrence Erlbaum Associates, 1997. ISBN 9781134801299. URL https://books.google.no/books?id=fLYWFcuamPwC. 21, 23

[101] Ee-Leng Tan and Woon-Seng Gan. Computational models for just-noticeable differences. In *Perceptual Image Coding with Discrete Cosine Transform*, pages 3–19. Springer, Singapore, 2015. doi: 10.1007/978-981-287-543-3_2. 22, 25, 59

[102] Tom N. Cornsweet. The staircase-method in psychophysics. *The American Journal of Psychology*, 75(3):485–491, 1962. ISSN 00029556. URL http://www.jstor.org/stable/1419876. 23

[103] Seung-Won Jung and Sung-Jea Ko. Depth sensation enhancement using the just noticeable depth difference. *IEEE Transactions on Image Processing*, 21(8):3624–3637, 2012. 24

[104] Seung-Won Jung, Jae-Yun Jeong, and Sung-Jea Ko. Sharpness enhancement of stereo images using binocular just-noticeable difference. *Image Processing, IEEE Transactions on*, 21(3):1191–1199, 2012. 24

[105] D Varuna SX De Silva, Erhan Ekmekcioglu, Warnakulasuriya Anil Chandana Fernando, and Stewart T Worrall. Display dependent preprocessing of depth maps based on just noticeable depth difference modeling. *IEEE Journal of Selected Topics in Signal Processing*, 5(2):335–351, 2011. 24

[106] Yu Cao, Wenhao Hong, and Lu Yu. Full-reference perceptual quality assessment for stereoscopic images based on primary visual processing mechanism. In *IEEE International Conference on Multimedia and Expo (ICME)*, pages 1–6, 2016. 24, 55, 60

[107] ITU-R BT.2021-1. Subjective methods for the assessment of stereoscopic 3DTV systems. *International Telecommunication Union*, 2015. Available: https://www.itu.int/rec/R-REC-BT.2021-1-201502-I/en. 25, 28

[108] ITU-R BT.500-8. Methodology for the subjective assessment of the quality of television pictures. *International Telecommunication Union*, 1998. Available: https://www.itu.int/rec/R-REC-BT.500-8-199802-S/en. 25, 28

[109] Chun-Hsien Chou and Yun-Chin Li. A perceptually tuned subband image coder based on the measure of just-noticeable-distortion profile. *IEEE Transactions on Circuits and Systems for Video Technology*, 5(6):467–476, 1995. 25, 30

[110] Ingo Hontsch and Lina J Karam. Adaptive image coding with perceptual distortion control. *IEEE Transactions on Image Processing*, 11(3):213–222, 2002. 25, 30

[111] XK Yang, WS Ling, ZK Lu, Ee Ping Ong, and SS Yao. Just noticeable distortion model and its applications in video coding. *Signal Processing: Image Communication*, 20(7):662–680, 2005. 25, 30

[112] XH Zhang, WS Lin, and Ping Xue. Improved estimation for just-noticeable visual distortion. *Signal Processing*, 85(4):795–808, 2005. 25, 30

[113] Xiaohui Zhang, Weisi Lin, and Ping Xue. Just-noticeable difference estimation with pixels in images. *Journal of Visual Communication and Image Representation*, 19(1):30–41, 2008. 25, 30

[114] Anmin Liu, Weisi Lin, Manoranjan Paul, Chenwei Deng, and Fan Zhang. Just noticeable difference for images with decomposition model for separating edge and textured regions. *IEEE Transactions on Circuits and Systems for Video Technology*, 20(11):1648–1652, 2010. 25, 30, 38, 49

[115] Jinjian Wu, Guangming Shi, Weisi Lin, Anmin Liu, and Fei Qi. Just noticeable difference estimation for images with free-energy principle. *IEEE Transactions on Multimedia*, 15(7): 1705–1710, 2013. 25, 30

[116] Jinjian Wu, Leida Li, Weisheng Dong, Guangming Shi, Weisi Lin, and C-C Jay Kuo. Enhanced just noticeable difference model for images with pattern complexity. *IEEE Transactions on Image Processing*, 26(6):2682–2693, 2017. 25, 30

[117] Hadi Hadizadeh, Atiyeh Rajati, and Ivan V Bajić. Saliency-guided just noticeable distortion estimation using the normalized laplacian pyramid. *IEEE Signal Processing Letters*, 24(8): 1218–1222, 2017. 25

[118] Xinfeng Zhang, Shiqi Wang, Ke Gu, Weisi Lin, Siwei Ma, and Wen Gao. Just-noticeable difference-based perceptual optimization for JPEG compression. *IEEE Signal Processing Letters*, 24(1):96–100, 2017. 25

[119] Sung-Ho Bae and Munchurl Kim. A DCT-based total JND profile for spatio-temporal and foveated masking effects. *IEEE Transactions on Circuits and Systems for Video Technology*, 27 (6):1196–1207, 2017. 25

[120] Gareth James, Daniela Witten, Trevor Hastie, and Robert Tibshirani. *An introduction to statistical learning*, volume 112. Springer, New York, USA, 2013. 25

[121] Haiqiang Wang, Ioannis Katsavounidis, Jiantong Zhou, Jeonghoon Park, Shawmin Lei, Xin Zhou, Man-On Pun, Xin Jin, Ronggang Wang, Xu Wang, Yun Zhang, Jiwu Huang, Sam Kwong, and C.-C. Jay Kuo. VideoSet: A large-scale compressed video quality dataset based on JND measurement. *Journal of Visual Communication and Image Representation*, 46:292–302, 2017. 25

[122] Denis Cousineau and Sylvain Chartier. Outliers detection and treatment: a review. *International Journal of Psychological Research*, 3(1):58–67, 2010. 25

[123] Pete R Jones. A note on detecting statistical outliers in psychophysical data. *bioRxiv*, page 074591, 2016. Available: https://www.biorxiv.org/content/biorxiv/early/2016/09/12/074591.full.pdf. 25, 26

[124] ITU-R BT.1788. Methodology for the subjective assessment of video quality in multimedia applications. *International Telecommunication Union*, 2007. Available: https://www.itu.int/rec/R-REC-BT.1788-0-200701-I/en. 26

[125] Carlos M Jarque and Anil K Bera. A test for normality of observations and regression residuals. *International Statistical Review/Revue Internationale de Statistique*, 55(2):163–172, 1987. 27

[126] Gerald Van Belle. *Statistical rules of thumb*. Wiley, New York, 2011. 27

[127] Samuel Sanford Shapiro and Martin B Wilk. An analysis of variance test for normality (complete samples). *Biometrika*, 52(3/4):591–611, 1965. 27

[128] Howard Levene et al. Robust tests for equality of variances. *Contributions to probability and statistics*, 1:278–292, 1960. 27

[129] Zhou Wang. Applications of objective image quality assessment methods. *IEEE Signal Processing Magazine*, 28(6):137–142, 2011. 28

[130] J. Wang, S. Wang, and Z. Wang. Asymmetrically compressed stereoscopic 3D videos: quality assessment and rate-distortion performance evaluation. *IEEE Transactions on Image Processing*, 26(3):1330–1343, March 2017. 28

[131] F. Battisti, M. Carli, P. Le Callet, and P. Paudyal. Toward the assessment of quality of experience for asymmetric encoding in immersive media. *IEEE Transactions on Broadcasting*, 64(2):392–406, June 2018. 28

[132] Pedram Mohammadi, Abbas Ebrahimi-Moghadam, and Shahram Shirani. Subjective and Objective Quality Assessment of Image: A Survey. *CoRR*, abs/1406.7799, 2014. Available: http://arxiv.org/abs/1406.7799. 28

[133] ITU-R BT.2022. General viewing conditions for subjective assessment of quality of SDTV and HDTV television pictures on flat panel displays. *International Telecommunication Union*, 2015. Available: https://www.itu.int/rec/R-REC-BT.2021-1-201502-I/en. 28

[134] A. C. Bovik H. R. Sheikh, Z. Wang and L. K. Cormack. Image and video quality assessment research at live [online]. Available: http://live.ece.utexas.edu/research/quality/. 28

[135] Nikolay Ponomarenko, Vladimir Lukin, Alexander Zelensky, Karen Egiazarian, Jaakko Astola, Marco Carli, and Federica Battisti. TID2008 - A Database for evaluation of full-Reference visual

quality assessment metrics. *Advances of Modern Radioelectronics*, 26:1330–1343, 2009. Available: http://www.ponomarenko.info/tid2008. 28

[136] Nikolay Ponomarenko, Lina Jin, Oleg Ieremeiev, Vladimir Lukin, Karen Egiazarian, Jaakko Astola, Benoit Vozel, Kacem Chehdi, Marco Carli, Federica Battisti, and C.-C. Jay Kuo. Image database TID2013: Peculiarities, results and perspectives. *Signal Processing: Image Communication*, 30:57 – 77, 2015. 28, 57

[137] E. C. Larson and D. M. Chandler. Most apparent distortion: full- reference image quality assessment and the role of strategy. *Journal of Electronic Imaging*, 19:19–19–21, 2010. 28

[138] E. C. Larson and D. M. Chandler. LIVE multiply distorted image database. In *IEEE Asilomar Conference on Signals, Systems and Computers*, page 1693–1697, 2012. 28

[139] M. Liu, G. Zhai, S. Tan, Z. Zhang, K. Gu, and X. Yang. HDR2014 - A high dynamic range image quality database. In *IEEE International Conference on Multimedia and Expo Workshops (ICMEW)*, pages 1–6, July 2014. 28

[140] D. Ghadiyaram and A. C. Bovik. Massive online crowdsourced study of subjective and objective picture quality. *IEEE Transactions on Image Processing*, 25(1):372–387, 2016. 28, 57

[141] Kede Ma, Zhengfang Duanmu, Qingbo Wu, Zhou Wang, Hongwei Yong, Hongliang Li, and Lei Zhang. Waterloo Exploration Database: New challenges for image quality assessment models. *IEEE Transactions on Image Processing*, 26(2):1004–1016, Feb. 2017. Available: https://ece.uwaterloo.ca/~k29ma/exploration/. 28, 57

[142] Stefan Winkler. Annotated image quality databases. Available: https://stefan.winkler.site/resources.html. 28

[143] Weisi Lin and C.-C. Jay Kuo. Perceptual visual quality metrics: A survey. *Journal of Visual Communication and Image Representation*, 22(4):297 – 312, 2011. ISSN 1047-3203. doi: https://doi.org/10.1016/j.jvcir.2011.01.005. URL http://www.sciencedirect.com/science/article/pii/S1047320311000204. 29

[144] Redzuan Abdul Manap and Ling Shao. Non-distortion-specific no-reference image quality assessment: A survey. *Information Sciences*, 301:141–160, 2015. ISSN 0020-0255. doi: https://doi.org/10.1016/j.ins.2014.12.055. URL http://www.sciencedirect.com/science/article/pii/S0020025515000079. 29, 31

[145] Salvador Gabarda, Gabriel CristÃbal, and Navdeep Goel. Anisotropic blind image quality assessment: Survey and analysis with current methods. *Journal of Visual Communication and Image Representation*, 52:101 – 105, 2018. ISSN 1047-3203. doi: https://doi.org/10.1016/j.jvcir.2018.

02.008. URL http://www.sciencedirect.com/science/article/pii/S1047320318300403. 29, 31

[146] Zhou Wang and Alan C. Bovik. Mean squared error: Love it or leave it? a new look at signal fidelity measures. *IEEE Signal Processing Magazine*, 26(1):98–117, 2009. 29

[147] Zhou Wang and Alan C Bovik. A universal image quality index. *IEEE signal processing letters*, 9(3):81–84, 2002. 29, 49, 53

[148] Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4): 600–612, 2004. 29, 53

[149] Z. Wang, E. P. Simoncelli, and A. C. Bovik. Multiscale structural similarity for image quality assessment. In *The Thrity-Seventh Asilomar Conference on Signals, Systems Computers*, volume 2, pages 1398–1402, Nov. 2003. 30, 44, 53

[150] H. R. Sheikh and A. C. Bovik. Image information and visual quality. In *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, volume 3, pages iii–709, May 2004. 30, 42, 49, 53

[151] L. Zhang, L. Zhang, X. Mou, and D. Zhang. FSIM: a feature similarity index for image quality assessment. *IEEE transactions on Image Processing*, 20(8):2378–2386, 2011. 30, 53

[152] Anmin Liu, Weisi Lin, and Manish Narwaria. Image quality assessment based on gradient similarity. *IEEE Transactions on Image Processing*, 21(4):1500–1512, 2012. 30

[153] Wufeng Xue, Lei Zhang, Xuanqin Mou, and Alan C Bovik. Gradient magnitude similarity deviation: A highly efficient perceptual image quality index. *IEEE Transactions on Image Processing*, 23(2):684–695, 2014. 30, 50

[154] Ke Gu, Guangtao Zhai, Xiaokang Yang, and Wenjun Zhang. Self-adaptive scale transform for iqa metric. In *IEEE International Symposium on Circuits and Systems (ISCAS)*, pages 2365–2368, 2013. 30

[155] Min Liu, Guangtao Zhai, Ke Gu, Qi Xu, Xiaokang Yang, Xianghui Sun, Wanhong Chen, and Ying Zuo. A new image quality metric based on mix-scale transform. In *IEEE Workshop on Signal Processing Systems (SiPS)*, pages 266–271, 2013. 30

[156] Ke Gu, Leida Li, Hong Lu, Xiongkuo Min, and Weisi Lin. A fast reliable image quality predictor by fusing micro-and macro-structures. *IEEE Transactions on Industrial Electronics*, 64(5):3903–3912, 2017. 30

[157] Ke Gu, Min Liu, Guangtao Zhai, Xiaokang Yang, and Wenjun Zhang. Quality assessment considering viewing distance and image resolution. *IEEE Transactions on Broadcasting*, 61(3): 520–531, 2015. 30

[158] Nikolay Ponomarenko, Flavia Silvestri, Karen Egiazarian, Marco Carli, Jaakko Astola, and Vladimir Lukin. On between-coefficient contrast masking of DCT basis functions. In *Proceedings of the third international workshop on video processing and quality metrics*, volume 4, 2007. 30

[159] Ke Gu, Guangtao Zhai, Xiaokang Yang, Li Chen, and Wenjun Zhang. Nonlinear additive model based saliency map weighting strategy for image quality assessment. In *2012 IEEE 14th International Workshop on Multimedia Signal Processing (MMSP)*, pages 313–318, Sep. 2012. 30

[160] Lin Zhang, Ying Shen, and Hongyu Li. VSI: A visual saliency-induced index for perceptual image quality assessment. *IEEE Transactions on Image Processing*, 23(10):4270–4281, 2014. 30

[161] Jonathan Harel, Christof Koch, and Pietro Perona. Graph-based visual saliency. In *Advances in neural information processing systems*, pages 545–552, 2007. 30

[162] Hamed Rezazadegan Tavakoli, Esa Rahtu, and Janne Heikkilä. Fast and efficient saliency detection using sparse sampling and kernel density estimation. In *Scandinavian Conference on Image Analysis*, pages 666–675. Springer, 2011. 30

[163] Lijuan Duan, Chunpeng Wu, Jun Miao, Laiyun Qing, and Yu Fu. Visual saliency detection by spatially weighted dissimilarity. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 473–480, June 2011. 30

[164] Xiaodi Hou, Jonathan Harel, and Christof Koch. Image signature: Highlighting sparse salient regions. *IEEE transactions on Pattern Analysis and Machine Intelligence*, 34(1):194–201, 2012. 30

[165] Huawei Tian, Yuming Fang, Yao Zhao, Weisi Lin, Rongrong Ni, and Zhenfeng Zhu. Salient region detection by fusing bottom-up and top-down features extracted from a single image. *IEEE Transactions on Image Processing*, 23(10):4389–4398, 2014. 30

[166] Yuming Fang, Weisi Lin, Zhijun Fang, Zhenzhong Chen, Chia-Wen Lin, and Chenwei Deng. Visual acuity inspired saliency detection by using sparse features. *Information Sciences*, 309: 1–10, 2015. 30

[167] Eero P Simoncelli and Bruno A Olshausen. Natural image statistics and neural representation. *Annual review of neuroscience*, 24(1):1193–1216, 2001. 30

[168] Wilson S Geisler and Randy L Diehl. A bayesian approach to the evolution of perceptual and cognitive systems. *Cognitive Science*, 27(3):379–402, 2003. 30

[169] Hossein Ziaei Nafchi, Atena Shahkolaei, Rachid Hedjam, and Mohamed Cheriet. Mean deviation similarity index: Efficient and reliable full-reference image quality evaluator. *IEEE Access*, 4: 5579–5590, 2016. 30

[170] Ma, Kede. *Blind Image Quality Assessment: Exploiting New Evaluation and Design Methodologies*. PhD thesis, 2017. URL http://hdl.handle.net/10012/12538. 30

[171] Jongyoo Kim, Hui Zeng, Deepti Ghadiyaram, Sanghoon Lee, Lei Zhang, and Alan C Bovik. Deep convolutional neural models for picture-quality prediction: Challenges and solutions to data-driven image quality assessment. *IEEE Signal Processing Magazine*, 34(6):130–141, 2017. 30

[172] Robert C. Williamson Bernhard Scholkopf, Alex J. Smola and Peter L. Bartlett. A no-reference metric for perceived ringing artifacts in images. *New support vector algorithms*, 12(5):1207–1245, May 2000. 30

[173] Debasish Basak, Srimanta Pal, and Dipak Chandra Patranabis. Support vector regression. *Neural Information Processing-Letters and Reviews*, 11(10):203–224, 2007. 30, 32

[174] David E Rumelhart, Geoffrey E Hinton, and Ronald J Williams. Learning representations by back-propagating errors. *nature*, 323(6088):533, 1986. 30, 57, 61

[175] Leo Breiman. Random forests. *Machine Learning*, 45(1):5–32, Oct 2001. ISSN 1573-0565. doi: 10.1023/A:1010933404324. URL https://doi.org/10.1023/A:1010933404324. 30, 57, 61

[176] Anush Krishna Moorthy and Alan Conrad Bovik. A two-step framework for constructing blind image quality indices. *IEEE Signal processing letters*, 17(5):513–516, 2010. 30, 31

[177] Huixuan Tang, Neel Joshi, and Ashish Kapoor. Learning a blind measure of perceptual image quality. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 305–312, June 2011. 30, 31, 32

[178] M. Narwaria and W. Lin. SVD-based quality metric for image and video using machine learning. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, 42(2):347–364, April 2012. 30

[179] Peng Ye, Jayant Kumar, Le Kang, and David Doermann. Real-time no-reference image quality assessment based on filter learning. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 987–994, June 2013. 30

[180] Wufeng Xue, Xuanqin Mou, Lei Zhang, Alan C Bovik, and Xiangchu Feng. Blind image quality assessment using joint statistics of gradient magnitude and laplacian features. *IEEE Transactions on Image Processing*, 23(11):4850–4862, 2014. 30, 31, 32, 43, 44

[181] Jingtao Xu, Peng Ye, Qiaohong Li, Haiqing Du, Yong Liu, and David Doermann. Blind image quality assessment based on high order statistics aggregation. *IEEE Transactions on Image Processing*, 25(9):4444–4457, 2016. 30, 32

[182] L. Zhang, Z. Gu, X. Liu, H. Li, and J. Lu. Training quality-aware filters for no-reference image quality assessment. *IEEE MultiMedia*, 21(4):67–75, 2014. 30

[183] Huixuan Tang, Neel Joshi, and Ashish Kapoor. Blind image quality assessment using semi-supervised rectifier networks. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2877–2884, June 2014. 30

[184] Le Kang, Peng Ye, Yi Li, and David Doermann. Convolutional neural networks for no-reference image quality assessment. In *IEEE conference on computer vision and pattern recognition (CVPR)*, pages 1733–1740, June 2014. 30

[185] Simone Bianco, Luigi Celona, Paolo Napoletano, and Raimondo Schettini. On the use of deep learning for blind image quality assessment. *Signal, Image and Video Processing*, 12(2):355–362, 2018. 30

[186] Kede Ma, Wentao Liu, Kai Zhang, Zhengfang Duanmu, Zhou Wang, and Wangmeng Zuo. End-to-end blind image quality assessment using deep neural networks. *IEEE Transactions on Image Processing*, 27(3):1202–1213, 2018. 30

[187] Sebastian Bosse, Dominique Maniry, Klaus-Robert Müller, Thomas Wiegand, and Wojciech Samek. Deep neural networks for no-reference and full-reference image quality assessment. *IEEE Transactions on Image Processing*, 27(1):206–219, 2018. 30

[188] Jie Gu, Gaofeng Meng, Judith A Redi, Shiming Xiang, and Chunhong Pan. Blind image quality assessment via vector regression and object oriented pooling. *IEEE Transactions on Multimedia*, 20(5):1140–1153, 2018. 30

[189] Hossein Talebi and Peyman Milanfar. Nima: Neural image assessment. *IEEE Transactions on Image Processing*, 27(8):3998–4011, 2018. 30

[190] Zhou Wang and Alan C Bovik. Reduced-and no-reference image quality assessment. *IEEE Signal Processing Magazine*, 28(6):29–40, 2011. 31

[191] Shizhong Liu and Alan C Bovik. Efficient dct-domain blind measurement and reduction of blocking artifacts. *IEEE Transactions on Circuits and Systems for Video Technology*, 12(12): 1139–1149, 2002. 31

[192] Zhou Wang, Hamid R Sheikh, and Alan C Bovik. No-reference perceptual quality assessment of JPEG compressed images. In *IEEE International Conference on Image Processing (ICIP)*, volume 1, pages 474–477, Sep. 2002. 31

[193] Hanghang Tong, Mingjing Li, Hongjiang Zhang, and Changshui Zhang. Blur detection for digital images using wavelet transform. In *IEEE International Conference on Multimedia and Expo (ICME)*, pages 17–20, July 2004. 31

[194] Rony Ferzli and Lina J Karam. A no-reference objective image sharpness metric based on the notion of just noticeable blur (JNB). *IEEE transactions on Image Processing*, 18(4):717–728, 2009. 31

[195] Hamid R Sheikh, Alan C Bovik, and Lawrence Cormack. No-reference quality assessment using natural scene statistics: JPEG2000. *IEEE Transactions on Image Processing*, 14(11):1918–1927, 2005. 31

[196] Hantao Liu, Nick Klomp, and Ingrid Heynderickx. A no-reference metric for perceived ringing artifacts in images. *IEEE Transactions on Circuits and Systems for Video Technology*, 20(4): 529–539, 2010. 31

[197] X. Zhu and P. Milanfar. A no-reference sharpness metric sensitive to blur and noise. In *International Workshop on Quality of Multimedia Experience*, pages 64–69, July 2009. 31

[198] Pina Marziliano, Frederic Dufaux, Stefan Winkler, and Touradj Ebrahimi. Perceptual blur and ringing metrics: application to JPEG2000. *Signal processing: Image communication*, 19(2): 163–172, 2004. 31

[199] Anush Krishna Moorthy and Alan Conrad Bovik. Blind image quality assessment: From natural scene statistics to perceptual quality. *IEEE transactions on Image Processing*, 20(12):3350–3364, 2011. 31

[200] Michele A Saad, Alan C Bovik, and Christophe Charrier. Blind image quality assessment: A natural scene statistics approach in the dct domain. *IEEE transactions on Image Processing*, 21 (8):3339–3352, 2012. 31

[201] Anish Mittal, Anush Krishna Moorthy, and Alan Conrad Bovik. No-reference image quality assessment in the spatial domain. *IEEE Transactions on Image Processing*, 21(12):4695–4708, 2012. 31

[202] Peng Ye and David Doermann. No-reference image quality assessment using visual codebooks. *IEEE Transactions on Image Processing*, 21(7):3129–3138, 2012. 31

[203] Peng Ye, Jayant Kumar, Le Kang, and David Doermann. Unsupervised feature learning framework for no-reference image quality assessment. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1098–1105, June 2012. 31

[204] Anish Mittal, Gautam S Muralidhar, Joydeep Ghosh, and Alan C Bovik. Blind image quality assessment without human training using latent quality factors. *IEEE Signal Processing Letters*, 19(2):75–78, 2012. 31

[205] Wufeng Xue, Lei Zhang, and Xuanqin Mou. Learning without human scores for blind image quality assessment. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 995–1002, June 2013. 31, 58

[206] Lin Zhang, Lei Zhang, and Alan C Bovik. A feature-enriched completely blind image quality evaluator. *IEEE Transactions on Image Processing*, 24(8):2579–2591, 2015. 31, 32, 58

[207] Weilong Hou, Xinbo Gao, Dacheng Tao, and Xuelong Li. Blind image quality assessment via deep learning. *IEEE transactions on Neural Networks and Learning Systems*, 26(6):1275–1286, 2015. 31

[208] Kede Ma, Wentao Liu, Tongliang Liu, Zhou Wang, and Dacheng Tao. dipIQ: Blind image quality assessment by learning-to-rank discriminable image pairs. *IEEE Transactions on Image Processing*, 26(8):3951–3964, 2017. 31

[209] Alex J Smola and Bernhard Schölkopf. A tutorial on support vector regression. *Statistics and computing*, 14(3):199–222, 2004. 32, 58

[210] Hamid R Sheikh, Muhammad F Sabir, and Alan C Bovik. A statistical evaluation of recent full reference image quality assessment algorithms. *IEEE Transactions on Image Processing*, 15(11): 3440–3451, 2006. 33

[211] Sid Ahmed Fezza, Mohamed-Chaker Larabi, and Kamel Mohamed Faraoun. Stereoscopic image quality metric based on local entropy and binocular just noticeable difference. In *IEEE International Conference on Image Processing*, pages 2002–2006, Oct. 2014. 35, 53

[212] Daniel Scharstein and Richard Szeliski. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. *International Journal of Computer Vision*, 47(1):7–42, Apr. 2002. 35, 38

[213] Geoffrey Egnal and Richard P Wildes. Detecting binocular half-occlusions: Empirical comparisons of five approaches. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 24 (8):1127–1133, 2002. 35

[214] Ming-Jun Chen, Che-Chun Su, Do-Kyoung Kwon, Lawrence K Cormack, and Alan C Bovik. Full-reference quality assessment of stereopairs accounting for rivalry. *Signal Processing: Image Communication*, 28(9):1143–1155, 2013. 36, 41, 42, 44, 45, 53, 56, 57

[215] Jiheng Wang, Abdul Rehman, Kai Zeng, Shiqi Wang, and Zhou Wang. Quality prediction of asymmetrically distorted stereoscopic 3D images. *IEEE Transactions on Image Processing*, 24 (11):3400–3414, 2015. 36, 42, 44, 45, 56, 57

[216] Feng Qi, Debin Zhao, Xiaopeng Fan, and Tingting Jiang. Stereoscopic video quality assessment based on visual attention and just-noticeable difference models. *Signal, Image and Video Processing*, pages 1–8, 2015. 36

[217] Daniel Scharstein and Chris Pal. Learning conditional random fields for stereo. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1–8, 2007. 37, 38, 48

[218] Daniel Scharstein, Heiko Hirschmüller, York Kitajima, Greg Krathwohl, Nera Nešić, Xi Wang, and Porter Westling. High-resolution stereo datasets with subpixel-accurate ground truth. In *German Conference on Pattern Recognition*, pages 31–42. Springer, 2014. 37, 38, 48

[219] Dengxin Dai, Hayko Riemenschneider, and Luc Van Gool. The synthesizability of texture examples. In *IEEE Conference on Computer Vision and Pattern Recognition(CVPR)*, pages 3027–3034, June 2014. 37, 48

[220] Yin Zhao, Zhenzhong Chen, Ce Zhu, Yap-Peng Tan, and Lu Yu. Binocular just-noticeable-difference model for stereoscopic images. *IEEE Signal Processing Letter*, 18(1):19–22, 2011. 38, 46, 48, 49

[221] Yuming Fang, Junle Wang, Manish Narwaria, Patrick Le Callet, and Weisi Lin. Saliency detection for stereoscopic images. *IEEE Transactions on Image Processing*, 23(6):2625–2636, 2014. 38, 49

[222] Anush Krishna Moorthy, Che-Chun Su, Anish Mittal, and Alan Conrad Bovik. Subjective evaluation of stereoscopic image quality. *Signal Processing: Image Communication*, 28(8):870–883, 2013. 42, 45, 55, 56, 57

[223] Fei Gao and Jun Yu. Biologically inspired image quality assessment. *Signal Processing*, 124:210 – 219, 2016. ISSN 0165-1684. doi: https://doi.org/10.1016/j.sigpro.2015.08.012. URL http://www.sciencedirect.com/science/article/pii/S0165168415002856. Big Data Meets Multimedia Analytics. 44

[224] P. G. Freitas, W. Y. L. Akamine, and M. C. Q. Farias. No-reference image quality assessment using orthogonal color planes patterns. *IEEE Transactions on Multimedia*, 20(12):3353–3360, Dec 2018. ISSN 1520-9210. 44

[225] Lixiong Liu, Bing Yang, and Hua Huang. No-reference stereopair quality assessment based on singular value decomposition. *Neurocomputing*, 275:1823–1835, 2018. 44, 53, 54, 58

[226] Fucui Li, Feng Shao, Qiuping Jiang, Randi Fu, Gangyi Jiang, and Mei Yu. Local and global sparse representation for no-reference quality assessment of stereoscopic images. *Information Sciences*, 422:110 – 121, 2018. ISSN 0020-0255. doi: https://doi.org/10.1016/j.ins.2017.09.011. URL http://www.sciencedirect.com/science/article/pii/S0020025516312749. 44

[227] Lixiong Liu, Yi Hua, Qingjie Zhao, Hua Huang, and Alan Conrad Bovik. Blind image quality assessment by relative gradient statistics and adaboosting neural network. *Signal Processing: Image Communication*, 40:1 – 15, 2016. ISSN 0923-5965. doi: https://doi.org/10.1016/j.image.2015.10.005. URL http://www.sciencedirect.com/science/article/pii/S0923596515001708. 44

[228] L. Ding, H. Huang, and Y. Zang. Image quality assessment using directional anisotropy structure measurement. *IEEE Transactions on Image Processing*, 26(4):1799–1809, April 2017. ISSN 1057-7149. doi: 10.1109/TIP.2017.2665972. 44

[229] Junming Zhou, Gangyi Jiang, Xiangying Mao, Mei Yu, Feng Shao, Zongju Peng, and Yun Zhang. Subjective quality analyses of stereoscopic images in 3DTV system. In *Visual Communications and Image Processing (VCIP)*, pages 1–4, Nov 2011. 45, 55, 57

[230] Feng Shao, Weijun Tian, Weisi Lin, Gangyi Jiang, and Qionghai Dai. Learning sparse representation for no-reference quality assessment of multiply distorted stereoscopic images. *IEEE Transactions on Multimedia*, 19(8):1821–1836, 2017. 45, 57, 58

[231] IEEE Standards Association Stereo Image Database. http://grouper.ieee.org/groups/3dhf/. 45

[232] Frédéric Devernay and Paul Beardsley. Stereoscopic cinema. In Rémi Ronfard and Gabriel Taubin, editors, *Image and Geometry Processing for 3-D Cinematography*, pages 11–51. Springer Berlin Heidelberg, Berlin, Heidelberg, 2010. ISBN 978-3-642-12392-4. doi: 10.1007/978-3-642-12392-4_2. URL https://doi.org/10.1007/978-3-642-12392-4_2. 48

[233] Yu Fan, Mohamed-Chaker Larabi, Faouzi Alaya Cheikh, and Christine Fernandez-Maloigne. Blind Stereopair Quality Evaluation based on Statistics of Binocular Contrast Features. In *COmpression et REprésentation des Signaux Audiovisuels (CORESA)*, November 2018. 49, 51, 57, 60, 61

[234] Wei Zhang, Chenfei Qu, Lin Ma, Jingwei Guan, and Rui Huang. Learning structure of stereoscopic image for no-reference quality assessment with convolutional neural network. *Pattern Recognition*, 59:176–187, 2016. 51

[235] X. Wang, S. Kwong, Y. Zhang, and Y. Zhang. Considering binocular spatial sensitivity in stereoscopic image quality assessment. In *Visual Communications and Image Processing (VCIP)*, pages 1–4, Taiwan, Nov 2011. doi: 10.1109/VCIP.2011.6116015. 53

[236] S. A. Fezza and M. C. Larabi. Stereoscopic 3D image quality assessment based on cyclopean view and depth map. In *IEEE Fourth International Conference on Consumer Electronics Berlin (ICCE-Berlin)*, pages 335–339, Sept. 2014. doi: 10.1109/ICCE-Berlin.2014.7034289. 53

[237] Ming-Jun Chen, Lawrence K Cormack, and Alan C Bovik. No-reference quality assessment of natural stereopairs. *IEEE Transactions on Image Processing*, 22(9):3379–3391, 2013. 53, 54

[238] Lixiong Liu, Bao Liu, Che-Chun Su, Hua Huang, and Alan Conrad Bovik. Binocular spatial activity and reverse saliency driven no-reference stereopair quality assessment. *Signal Processing: Image Communication*, 58:287–299, 2017. 53, 54

[239] Feng Shao, Weisi Lin, Shanbo Gu, Gangyi Jiang, and Thambipillai Srikanthan. Perceptual full-reference quality assessment of stereoscopic images by considering binocular visual characteristics. *IEEE Transactions on Image Processing*, 22(5):1940–1953, 2013. 55, 60

[240] Feng Shao, Gang yi Jiang, Mei Yu, Fucui Li, Zongju Peng, and Randi Fu. Binocular energy response based quality assessment of stereoscopic images. *Digital Signal Processing*, 29:45–53, 2014. ISSN 1051-2004. doi: https://doi.org/10.1016/j.dsp.2014.03.003. URL http://www.sciencedirect.com/science/article/pii/S1051200414000967. 55, 60

[241] Wujie Zhou, Gangyi Jiang, Mei Yu, Feng Shao, and Zongju Peng. PMFS: a perceptual modulated feature similarity metric for stereoscopic image quality assessment. *IEEE Signal Processing Letters*, 21(8):1003–1006, 2014. 55, 60

[242] Yancong Lin, Jiachen Yang, Wen Lu, Qinggang Meng, Zhihan Lv, and Houbing Song. Quality index for stereoscopic images by jointly evaluating cyclopean amplitude and cyclopean phase. *IEEE Journal of Selected Topics in Signal Processing*, 11(1):89–101, 2017. 55, 60

[243] Shiqi Wang, Lin Ma, Yuming Fang, Weisi Lin, Siwei Ma, and Wen Gao. Just noticeable difference estimation for screen content images. *IEEE Transactions on Image Processing*, 25(8):3838–3851, 2016. 55

[244] F. Shao, Y. Gao, Q. Jiang, G. Jiang, and Y. Ho. Multistage pooling for blind quality prediction of asymmetric multiply-distorted stereoscopic images. *IEEE Transactions on Multimedia*, 20 (10):2605–2619, Oct 2018. ISSN 1520-9210. doi: 10.1109/TMM.2018.2817072. 57, 58

[245] Alan C. Bovik Anush K. Moorthy. A survey on 3D quality of experience and 3D quality assessment. volume 8651, pages 8651 – 8651 – 11, 2013. 57, 61

[246] Che-Chun Su, Anush Krishna Moorthy, and Alan Conrad Bovik. Visual quality assessment of stereoscopic image and video: challenges, advances, and future trends. In *Visual Signal Quality Assessment*, pages 185–212. Springer, 2015. 57

[247] D. Joshi, R. Datta, E. Fedorovskaya, Q. Luong, J. Z. Wang, J. Li, and J. Luo. Aesthetics and emotions in images. *IEEE Signal Processing Magazine*, 28(5):94–115, Sep. 2011. ISSN 1053-5888. 57

[248] J. Wang, S. Wang, K. Ma, and Z. Wang. Perceptual depth quality in distorted stereoscopic images. *IEEE Transactions on Image Processing*, 26(3):1202–1215, March 2017. ISSN 1057-7149. doi: 10.1109/TIP.2016.2642791. 57, 60

[249] H. G. Kim, H. Jeong, H. Lim, and Y. M. Ro. Binocular Fusion Net: Deep Learning Visual Comfort Assessment for Stereoscopic 3D. *IEEE Transactions on Circuits and Systems for Video Technology*, pages 1–1, 2018. ISSN 1051-8215. 57

[250] Pieter Seuntiens. Visual experience of 3D TV. *Doctoral thesis, Eindhoven University of Technology*, 2006. 57

[251] Wei Chen, Fournier Jérôme, Marcus Barkowsky, and Patrick Le Callet. Exploration of quality of experience of stereoscopic images: Binocular depth. In *Sixth International Workshop on Video Processing and Quality Metrics for Consumer Electronics-VPQM 2012*, pages 1–6, Jan. 2012. 58, 61

[252] E. Mavridaki and V. Mezaris. A comprehensive aesthetic quality assessment method for natural images using basic rules of photography. In *2015 IEEE International Conference on Image Processing (ICIP)*, pages 887–891, Sep. 2015. doi: 10.1109/ICIP.2015.7350927. 58

[253] A. Mittal, A. K. Moorthy, J. Ghosh, and A. C. Bovik. Algorithmic assessment of 3d quality of experience for images and videos. In *2011 Digital Signal Processing and Signal Processing Education Meeting (DSP/SPE)*, pages 338–343, Jan 2011. doi: 10.1109/DSP-SPE.2011.5739236. 58

[254] W. Zhou, Z. Chen, and W. Li. Dual-stream interactive networks for no-reference stereoscopic image quality assessment. *IEEE Transactions on Image Processing*, pages 1–1, 2019. ISSN 1057-7149. doi: 10.1109/TIP.2019.2902831. 58

[255] Q. Jiang, F. Shao, W. Gao, Z. Chen, G. Jiang, and Y. Ho. Unified no-reference quality assessment of singly and multiply distorted stereoscopic images. *IEEE Transactions on Image Processing*, 28(4):1866–1881, April 2019. ISSN 1057-7149. doi: 10.1109/TIP.2018.2881828. 58

[256] Sid Ahmed Fezza, Aladine Chetouani, and Mohamed-Chaker Larabi. Using distortion and asymmetry determination for blind stereoscopic image quality assessment strategy. *Journal of Visual Communication and Image Representation*, 49:115–128, 2017. 58

[257] H. Oh, S. Ahn, J. Kim, and S. Lee. Blind deep s3d image quality evaluation via local to global feature aggregation. *IEEE Transactions on Image Processing*, 26(10):4923–4936, Oct 2017. ISSN 1057-7149. doi: 10.1109/TIP.2017.2725584. 58

[258] Jiachen Yang, Huifang Xu, Yang Zhao, Hehan Liu, and Wen Lu. Stereoscopic image quality assessment combining statistical features and binocular theory. *Pattern Recognition Letters*, 2018. ISSN 0167-8655. doi: https://doi.org/10.1016/j.patrec.2018.10.012. URL http://www.sciencedirect.com/science/article/pii/S0167865518308158. 58

[259] C. Li, A. C. Bovik, and X. Wu. Blind image quality assessment using a general regression neural network. *IEEE Transactions on Neural Networks*, 22(5):793–799, May 2011. ISSN 1045-9227. doi: 10.1109/TNN.2011.2120620. 58

[260] M. Karimi, M. Nejati, S. M. R. Soroushmehr, S. Samavi, N. Karimi, and K. Najarian. Blind stereo quality assessment based on learned features from binocular combined images. *IEEE Transactions on Multimedia*, 19(11):2475–2489, Nov 2017. ISSN 1520-9210. doi: 10.1109/TMM.2017.2699082. 58

[261] Zhongkang Lu, W. Lin, X. Yang, EePing Ong, and Susu Yao. Modeling visual attention's modulatory aftereffects on visual sensitivity and quality evaluation. *IEEE Transactions on Image Processing*, 14(11):1928–1942, Nov. 2005. ISSN 1057-7149. doi: 10.1109/TIP.2005.854478. 59

[262] Ulrich Engelke, Daniel P Darcy, Grant H Mulliken, Sebastian Bosse, Maria G Martini, Sebastian Arndt, Jan-Niklas Antons, Kit Yan Chan, Naeem Ramzan, and Kjell Brunnström. Psychophysiology-based QoE assessment: A survey. *IEEE Journal of Selected Topics in Signal Processing*, 11(1):6–21, 2017. 61

[263] F. M. L. Ribeiro, J. F. L. de Oliveira, A. G. Ciancio, E. A. B. da Silva, C. R. D. Estrada, L. G. C. Tavares, J. N. Gois, A. Said, and M. C. Martelotte. Quality of experience in a stereoscopic multiview environment. *IEEE Transactions on Multimedia*, 20(1):1–14, Jan. 2018. ISSN 1520-9210. doi: 10.1109/TMM.2017.2714425. 61

[264] Jian Ding, Stanley A Klein, and Dennis M Levi. Binocular combination of phase and contrast explained by a gain-control and gain-enhancement model. *Journal of vision*, 13(2):13–13, 2013. 61

[265] Jiawei Zhou, Mark A Georgeson, and Robert F Hess. Linear binocular combination of responses to contrast modulation: Contrast-weighted summation in first-and second-order vision. *Journal of vision*, 14(13):24–24, 2014. 61

[266] Fan Zhang, Wenfei Jiang, Florent Autrusseau, and Weisi Lin. Exploring V1 by modeling the perceptual quality of images. *Journal of Vision*, 14(1):26–26, 01 2014. ISSN 1534-7362. doi: 10.1167/14.1.26. URL https://doi.org/10.1167/14.1.26. 61

[267] Shinji Takase, Shinji Yukumatsu, and Kazuo Bingushi. Local binocular fusion is involved in global binocular rivalry. *Vision Research*, 48(17):1798 – 1803, 2008. ISSN 0042-6989. doi:

https://doi.org/10.1016/j.visres.2008.05.010. URL http://www.sciencedirect.com/science/article/pii/S0042698908002769. 61

[268] K. Lee and S. Lee. 3d perception based quality pooling: Stereopsis, binocular rivalry, and binocular suppression. *IEEE Journal of Selected Topics in Signal Processing*, 9(3):533–545, April 2015. ISSN 1932-4553. doi: 10.1109/JSTSP.2015.2393296. 61

[269] Yang Yao, Liquan Shen, and Ping An. Bivariate analysis of 3D structure for stereoscopic image quality assessment. *Signal Processing: Image Communication*, 65:128 – 140, 2018. ISSN 0923-5965. doi: https://doi.org/10.1016/j.image.2018.02.014. URL http://www.sciencedirect.com/science/article/pii/S0923596518301772. 61

[270] S. Ki, S. Bae, M. Kim, and H. Ko. Learning-based just-noticeable-quantization-distortion modeling for perceptual video coding. *IEEE Transactions on Image Processing*, 27(7):3178–3193, July 2018. ISSN 1057-7149. doi: 10.1109/TIP.2018.2818439. 61

[271] N. K. Kottayil, I. Cheng, G. Valenzise, and F. Dufaux. Learning local distortion visibility from image quality. In *25th IEEE International Conference on Image Processing (ICIP)*, pages 281–285, Oct 2018. doi: 10.1109/ICIP.2018.8451732. 61

[272] H. Hadizadeh, A. R. Heravi, I. V. Bajic, and P. Karami. A Perceptual Distinguishability Predictor For JND-Noise-Contaminated Images. *IEEE Transactions on Image Processing*, 28(5): 2242–2256, May 2019. ISSN 1057-7149. doi: 10.1109/TIP.2018.2883893. 62

[273] H. Duan, G. Zhai, X. Min, Y. Zhu, Y. Fang, and X. Yang. Perceptual quality assessment of omnidirectional images. In *2018 IEEE International Symposium on Circuits and Systems (ISCAS)*, pages 1–5, May 2018. 62

[274] M. Huang, Q. Shen, Z. Ma, A. C. Bovik, P. Gupta, R. Zhou, and X. Cao. Modeling the perceptual quality of immersive images rendered on head mounted displays: Resolution and compression. *IEEE Transactions on Image Processing*, 27(12):6039–6050, Dec 2018. ISSN 1057-7149. doi: 10.1109/TIP.2018.2865089. 62

[275] Ziheng Zhang, Yanyu Xu, Jingyi Yu, and Shenghua Gao. Saliency detection in 360° videos. In *European Conference on Computer Vision (ECCV)*, pages 488–503, September 2018. 62

[276] Jing Ling, Kao Zhang, Yingxue Zhang, Daiqin Yang, and Zhenzhong Chen. A saliency prediction model on 360 degree images using color dictionary based sparse representation. *Signal Processing: Image Communication*, 69:60 – 68, 2018. ISSN 0923-5965. doi: https://doi.org/10.1016/j.image.2018.03.007. URL http://www.sciencedirect.com/science/article/pii/S0923596518302418. 62

[277] R. Schatz, A. Sackl, C. Timmerer, and B. Gardlo. Towards subjective quality of experience assessment for omnidirectional video streaming. In *Ninth International Conference on Quality of Multimedia Experience (QoMEX)*, pages 1–6, May 2017. doi: 10.1109/QoMEX.2017.7965657. 62

[278] M. Xu, C. Li, Z. Chen, Z. Wang, and Z. Guan. Assessing visual quality of omnidirectional videos. *IEEE Transactions on Circuits and Systems for Video Technology*, pages 1–1, 2018. ISSN 1051-8215. doi: 10.1109/TCSVT.2018.2886277. 62

Part II

Included Papers

Chapter 6

Paper I: On the Performance of 3D Just Noticeable Difference Models

Y. Fan[1,2], M.-C. Larabi[1], F. A. Cheikh[2], C. Fernandez-Maloigne[1]

[1]XLIM, University of Poitiers, France
[2]Norwegian Colour and Visual Computing Laboratory, NTNU-Gjøvik, Norway

In *IEEE International Conference on Image Processing (ICIP)*, pages 1017–1021, September 2016.

Abstract

The just noticeable difference (JND) notion reflects the maximum tolerable distortion. It has been extensively used for the optimization of 2D applications. For stereoscopic 3D (S3D) content, this notion is different since it relies on different mechanisms linked to our binocular vision. Unlike 2D, 3D-JND models appeared recently and the related literature is rather limited. These models can be used for the sake of compression and quality assessment improvement for S3D content. In this paper, we propose a deep and comparative study of the existing 3D-JND models. Additionally, in order to analyze their performance, the 3D-JND models have been integrated in recent metric dedicated to stereoscopic image quality assessment (SIQA). The results are reported on two widely used S3D image databases.

Index terms– Just noticeable difference (JND), stereoscopic 3D (S3D), 3D-JND, stereoscopic image quality assessment (SIQA).

6.1 Introduction

In the recent years, S3D technology has changed the user's viewing experience. It provides the viewer with new sensations of immersion. However, the advent of this technology introduced some technical challenges such as compression and quality assessment, to name a few.

Besides, to ensure quality and comfort at every step of the S3D application, it is important to understand and account for the different perceptual processes of the human visual system (HVS). For decades, the scientific community has acquired a deep knowledge about 2D perception. Several models have been successfully developed and exploited like the just noticeable difference (JND). The latter informs about thresholds, depending on the luminance, contrast and spatial frequency of the local image regions, beyond which a distortion will be visible. 2D-JND models are generally developed based on specific characteristics of MV, which does not fit with the complexity of 3D perception requiring specific models accounting for both monocular and binocular depth cues.

To date, only a few 3D-JND models can be found in the literature [1–5] and additional research efforts are needed to achieve a more complete and efficient model. The aim of such a model is to improve the performance of 3D applications (compression, watermarking, quality assessment, ...). To the best of our knowledge, there is no comprehensive survey of the 3D-JND models. In this paper, we propose an exhaustive literature review of the 3D-JND models. Each model is briefly described by giving its rationale and main components in addition to providing information about the targeted applications, the pros and cons. In order to quantitatively compare the different models, we adopted a quality assessment framework including a 3D-JND block. Every model has been tested separately on two widely used 3D image databases. The performance has been measured thanks to state-of-the-art measures namely Pearson and Spearman correlations and RMSE. The remainder of the paper is organized as follows. In Sect. 6.2, we review the existing 3D-JND models separately. The models are compared and analyzed in Sect. 6.3. Sect. 6.4 presents the experimental results of performance comparison used in SIQA. This paper ends with some conclusions and future work.

6.2 3D-JND models

6.2.1 Just Noticeable Difference in Depth (JNDD)

De Silva *et al.* proposed the just noticeable difference in depth (JNDD) model [1, 6] describing the threshold for depth changes that human can perceive on a 3D display. The JNDD model was developed based on Psychophysical Experiments (PE) by using different simulated depth values. As described in [1], the sensitivity of the human eyes to depth difference is mainly dependent on the viewing distance and the displayed depth level of the image. The viewing distance provides the location of the fixation point, i.e., the screen. The mathematical JNDD model [6] is described by:

$$JNDD = 10^{0.94 \times log_{10}(v) - 2.25} + K_w \cdot |dp|\,, \tag{6.1}$$

where K_w is the Weber Constant determined by the experiment. v and dp denote the viewing distance and the simulated depth level (from 0 to 255), respectively. Since the viewing distance rarely changes in real situations, the JNDD is measured by only considering the displayed depth level in PE. The experimental validation of the JNDD has been performed using 2D-plus-depth videos. Two identical objects are first displayed at the same depth level, and then the depth level of one object is changed gradually. The subjects are asked to inform about depth change between the two objects when perceived. Various initial depth levels of the two objects have been investigated and the final threshold is obtained by averaging the JNDD values of all subjects.

6.2.2 Binocular Just Noticeable Difference (BJND)

Another 3D-JND model, called binocular JND (BJND) has been introduced by Zhao *et al.* [2]. It indicates the threshold in inter-difference between the left and right views that human can recognize. This model investigates the properties of the binocular vision in response to asymmetric noise in a pair of stereoscopic image based on the HVS characteristics such as luminance adaptation (LA) and contrast masking (CM). The latter HVS characteristics are often used in 3D-JND modeling. It should be noted that there are two BJND thresholds (left and right) for each stereo-pair, because the BJND of one view indicates the maximum distortions that can be introduced in this view without evoking binocularly visible differences, given the distortions in the corresponding region of the other view. Like this, $BJND_{(l|r)}$ [2] is defined by:

$$\begin{aligned} BJND_{l|r}(i,j,d) &= BJND_{l|r}(bg_{r|l}(i,j+d), eh_{r|l}(i,j+d), n_{r|l}(i,j+d)) \\ &= A_C(bg_{r|l}(i,j+d), eh_{r|l}(i,j+d)) \\ &\times \left[1 - (\frac{n_{r|l}(i,j+d)}{A_C(bg_{r|l}(i,j+d), eh_{r|l}(i,j+d))})^{\gamma}\right]^{\frac{1}{\gamma}}, \end{aligned} \tag{6.2}$$

where $l|r$ refers to left or right, (i,j) denotes the pixel coordinate, and d is the horizontal disparity value at pixel (i,j). The parameter γ controls the influence of noise in the other view and is set to 1.25. $bg_{r|l}(i,j)$ and $eh_{r|l}(i,j)$ respectively indicate the average background luminance (ABL) and edge height. Note that $0 \leq n_r \leq A_C$ and if there is no noise in the other view, $BJND_{(l|r)}$ can be reduced to A_C. The experimental results showed that human perceives the distortion when viewing the stereo images if this distortion in one view is more than the BJND value.

6.2.3 Joint Just Noticeable Difference (JJND)

In a different fashion, Li *et al.* [3] proposed the joint just noticeable difference (JJND) model based on the idea that human has different perceptions of objects with different depths. Unlike the JNDD and BJND models, the JJND model was developed with a 2D-JND model [7], namely nonlinear additively masking model (NAMM), which accounts for LA and texture masking (TxM).

Firstly, the JND thresholds of one image (e.g., left image) are calculated by NAMM. For the other image (e.g., right image), disparity estimation is performed in order to classify the image pixels into two categories: occluded and non-occluded pixels [8]. Occluded regions, often appearing at the edges of objects, are very strong monocular clues and participate to depth perception. Based on the aforementioned classification, the JJND of one view is proportional to the JND threshold of the other view, where the coefficients are defined as: 1) depth-dependent value β for non-occluded pixel and 2) a fixed value α for occluded pixels. The JJND of the right image [3] can be formalized as follows:

$$JJND_r(i,j) = \begin{cases} \alpha \cdot JND(i,j), & if(i,j) \in occlusion \\ \beta(i,j) \cdot JND(i,j), & otherwise \end{cases}, \tag{6.3}$$

where JND is the visibility threshold for left image determined by [9] and α set to 0.8. The JJND has been validated thanks to subjective experiments.

6.2.4 Stereo Just Noticeable Difference (SJND)

The abovementioned 3D-JND models were dedicated to stereoscopic images, while the one discussed here focuses on stereoscopic video. Based on the free energy theory in brain theory and neuroscience, HVS adaptively excludes the disorder tendency information in a continued movement scene, and try to focus on the definite content of the perceived image [10]. This phenomenon can be modeled as the temporal masking (TM) caused by motion when watching a video. Therefore, Qi *et al.* [4, 11] developed the stereo just noticeable difference (SJND) model by considering both intra-view and inter-view masking effects in addition to LA and spatial masking (SP). The intra-view masking includes binocular masking (BM), whereas intra-view masking refers to TM.

Firstly, for one of the views, a JND image is determined by integrating LA and SP [12]. Based on it, the temporal JND (TJND) is derived for each view's sequence. For a pair of stereoscopic images, the final TJND is computed by combining the TJND of left and right frames using weights set to 0.375 and 0.625 respectively according to the used asymmetry between views. Next, views are decomposed into occluded and non-occluded regions that result in the binocular rivalry (BR) and binocular fusion (BF), respectively. To model the BM, different inter-view JND ($IJND$) thresholds are calculated based on left and right views according to occluded and non-occluded regions. The IJND of the occluded regions is combined with TJND to obtain the $TIJND$. The $IJND$ of the non-occluded regions can be computed by considering the LA and SP as well as the left and right view's consistency of luminance. By combining the $TIJND$ and $IJND$ for non-occluded regions, SJND [4]is defined as:

$$SJND(i,j,t) = (TIJND(i,j,t))^{\mu} + (IJND_n(i,j,t))^{1-\mu}, \tag{6.4}$$

where μ manages the tradeoff between $TIJND$ and $IJND_n$ ($\mu = 0.6$ in [11] for the best performance. SJND was used for SVQA showed a very good performance.

Table 6.1 – Comparison between the existing 3D-JND models.

	JNDD	BJND	JJND	SJND	HJND
Inputs	Viewing distance, depth level	ABL, image luminance and noise amplitude	ABL, image luminance, disparity and depth maps	ABL, image luminance, disparity map, and temporal information	ABL, image luminance, depth map
Masking effect	DI	LA and CM	LA, TxM and DI	LA, CM, TM and BM	LA, TxM, DI and GD
3D format	2D-plus-depth	LR views	LR views	LR frames	DIBR, MVD
Model validation	Theoretical results vs results derived from PE	Noise detection probability in S3D image	Comparison with 2D-JND in terms of subjective QA	A metric using this model vs SVQA metrics in terms of subjective VQ	This model vs JJND in terms of both compression efficiency and subjective VQ
Complexity	-	**	*	*	****
Pros	Different stereoscopic displays	Several 3D video formats	Binocular vision properties	Multiple masking effects	Depth contrast information, GD
Cons	Limits 3D video format, influence of depth image quality	Zero disparity, stereo matching algorithm impact	Unreliable results for low disparity, lack of comparison with 3D models	Difficult to design the PE for model validation directly, Many parameters	Highly depending on DIBR techniques, specially designed for MVD format

6.2.5 Hybrid Just Noticeable Difference (HJND)

It has been demonstrated that depth perception is influenced not only by depth intensity (DI) but also by depth contrast (DC). In light of this, Zhong *et al.* [5] proposed the hybrid just noticeable difference (HJND) model. The HJND model is de-signed to measure JND threshold especially for multi-view video plus depth (MVD) [13]. It considers DI, DC and the geometric distortion (GD) and is developed on the top the 2D-JND model described in [13]. The HVS is more sensitive to closer objects than deeper ones, and the regions with inconsecutive depth or higher DC attract more attention. Based on these considerations, a depth saliency model [14] was used to quantify the combined action of DI and DC for 3D video. The geometric distortion in synthetic views, which derives from one of depth image-based rendering (DIBR) techniques [15], is introduced by the quantization distortion decoded depth map and measured by the Hausdorff distance [16]. The HJND threshold [5] can be calculated as follows:

$$HJND(i,j) = \varepsilon \cdot JND(i,j) \cdot \omega^{N(S_d(i,j) \cdot G(i,j))}, \omega \in (0,1), \tag{6.5}$$

where $JND(i,j)$ denotes the 2D-JND threshold at coordinate (i,j). The parameters ε and ω set empirically to 1.4 and 0.15, respectively. S_d and G indicate depth saliency image and GD image respectively. The symbol $N(\cdot)$ is used for a normalization function.

6.3 Comparison of 3D-JND models

The overall comparison between the previously described 3D-JND models is given in Table 6.1 We compare in this section these models by analyzing their various aspects.

6.3.1 Pros and cons

The JNDD model has been designed thanks to PE on stereoscopic displays. This model can be extended to a variety of stereoscopic display types. However, it can only measure the visibility threshold with limited depth levels, thus could not satisfy the desirable depth range for real application. For instance, the JNDD is not suitable for estimating the tolerable depth difference in virtual view rendering. In addition, this model is only compatible with 2D-plus-depth representation of 3D content. Besides, depth images with poor quality may lead to inaccurate JND thresholds. Compared to JNDD model, the BJND model is closer to human binocular perception. Further-more, the BJND model can use both texture-plus-depth and LR views formats. However, the BJND model was generated based on experiments that ignore the effect of disparity, which makes it less suitable for real-world stereoscopic images. In addition, the disparity estimation error using stereo matching algorithm may decrease the reliability of the BJND estimation.

The JJND model copes with the issue of the zero disparity happening with the BJND model. Moreover, this model mimics the BF and BR by computing different JND thresholds for occluded and non-occluded regions separately. The performance of JJND can be reduced for a pair of S3D images with low disparity image i.e., weak depth perception. Besides, even though JJND has been given as more effective than 2D-JND model, it has not been compared with the 3D-JND models. The SJND model is theoretically the most reliable among these 3D-JND models since it takes into consideration four masking effects so that it completely model the HVS characteristics. However, it is difficult to run a subjective validation of this model because there are several factors from different masking effects. Tuning the parameters is somewhat complicated and may lead to very different results, in addition to the necessary adjustment to the used dataset. As described previously about the SJND model, the non-occluded regions lead to BF, whereas occluded regions cause BR. In fact, binocular rivalry can occur on non-occluded regions when a large inter-difference exists between left view non-occluded pixels and its corresponding pixels in the right view. The relationship between BF and BR should be better explored to model human binocular vision.

In contrast to JJND, the HJND model has taken the depth contrast into account besides depth intensity. Additionally, considering GD make this model more reliable. However, the HJND is specifically developed out for MVD format, and the estimation based on LR views format may not be correct. The accuracy of this model is highly depending on the rendered images derived from DIBR algorithm.

6.3.2 Applications embedding 3D-JND

To improve the compression efficiency, De silva *et al.* [6] proposed a depth map preprocessing algorithm based on the JJND model to remove depth detail not noticed by the viewer. In addition, the JJND model was used to enhance the depth sensation in [17, 18]. The principle is to increase depth differences between objects such that it exceeds the JNDD. Inspired by this model, Nur *et al.* [19] investigated

the sensitivity of the HVS to depth details under different illumination conditions.

Jung *et al.* [20] applied the BJND model in sharpness enhancement of stereo images, and the reliability of the BJND has been evaluated by considering the accuracy of the stereo match ng technique. From a different perspective, three SIQA metrics [21–23] have been proposed based on the BJND model. It should be noted that in this paper the performance comparison between 3D-JND models has been performed by using the method proposed by *et al.* [23]. Meanwhile, the same authors proposed an asymmetric stereoscopic coding method [24] based on BJND and depth level. This method employs the BJND model to measure the minimum distortions in one view that generate 3D perceptual difference, and then uses the depth information to adjust the resolution. In 3D video compression, a new MacroBlock level rate control method based on BJND model has been proposed in [25]. The visual perception factor measured by BJND was used to adjust the MB level bit allocation. An efficient mode decision approach using BJND has been developed for MVC [26] reducing significantly the runtime with a negligible increase of bit rate.

Recently, Wang *et al.* [27] carried out a stereo images watermarking method based on the JJND model. This approach validates the authenticity and integrity of stereo images by localizing the tampered regions. The SJND model was used for SVQA, whereas the HJND model has been applied to improve 3D coding efficiency for MVD. So far there is no application in other domains for SJND and HJND models since they have been proposed recently.

6.4 Experiments

With the aim to compare the performance of the aforementioned 3D-JND models in a more quantitative way, we adopted the quality framework described in [23]. The BJND block described in this method is sequentially replaced by the described 3D-JND models to estimate the overall 3D quality. The JNDD model has been discarded from this evaluation because of its dependency to the psychophysical conditions that cannot be controlled here. This metric is based on the assumption that 3D human perception is dominated by the view that contains more information. The overall quality score of each region is modulated based on a 3D-JND model for non-occluded pixels and the JND model for occluded pixels. In our experiments, the metric measures the quality for left and right views separately. So, for HJND model, the reference views are used to generate the virtual views using DIBR [28]. Occlusion detection was achieved by crosschecking the pair of disparity maps [29].

Various databases are publicly available for 3D QA. We propose to evaluate the performance of the 3D-JND models on two databases providing subjective scores: LIVE 3D IQA databases (phase II) [**?**] and Waterloo IVC 3D IQ databases (phase I) [**?**]. The former database contains the ground truth disparity and depth maps, and for the latter database, this information can be obtained from Middlebury 2005 Stereo Database [30]. LIVE 3D IQA database is composed of 8 reference and 360 distorted stereo pairs with symmetric and asymmetric distortions, including JPEG, JPEG2000, additive white Gaussian noise (WN), Gaussian blur (GB) and fast-fading. An example of the images

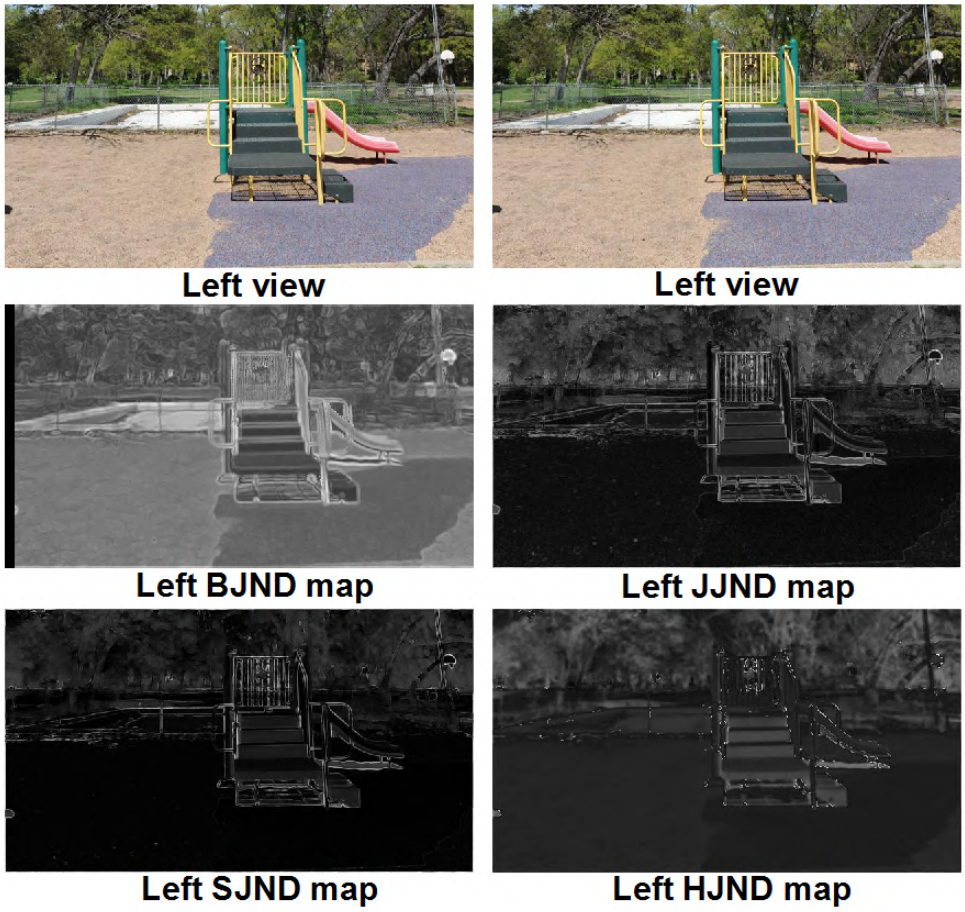


Figure 6.1 – Example of a pair of stereoscopic images from LIVE 3D IQA database and the corresponding 3D-JND maps.

from this database is given in Fig. 6.1 gives an example of stereo-pair from LIVE 3D IQA together with the corresponding BJND, JJND, SJND and HJND maps. One can notice the variability of the 3D-JND maps. Waterloo IVC 3D database phase I consists of 6 reference stereo pairs collected from the Middlebury 2005 Stereo Datasets, and 324 distorted stereo pairs. It includes symmetric and asymmetric distortions i.e., WN, GB and JPEG. The performance of the quality metric including the 3D-JND models have been evaluated using three performance measures: the Linear Correlation Coefficient (LCC), the Spearman Rank Order Correlation Coefficient (SROCC) and RMSE.

Table 6.2 shows the performance of the metrics using 3D-JND models on LIVE 3D IQA database. These results demonstrate that SJND outperforms the others models for both symmetric and symmetric distortions. This can be explained by the fact that the SJND model considers various characteristics of the binocular vision, which undoubtedly correspond better to the human quality judgment. The BJND mod-el is similar to the JJND in terms of the performance. On the other hand, the HJND model achieved the worst performance among all 3D-JND models because this model was initially designed for 3D representation with MVD format. Nevertheless, the overall performance is relatively

Table 6.2 – Performance comparison between 3D-JND models used in SIQA on LIVE 3D IQA database (phase II), AS and S denote asymmetric and symmetric distortions respectively.

Model	LCC			SROCC			RMSE		
	AS	S	ALL	AS	S	ALL	AS	S	ALL
BJND	0.714	0.811	0.762	0.680	0.804	0.746	7.09	7.31	7.30
JJND	0.701	0.816	0.755	0.652	0.809	0.734	7.24	7.22	7.40
SJND	**0.740**	**0.850**	**0.800**	**0.701**	**0.842**	**0.780**	**6.82**	**6.59**	**6.91**
HJND	0.674	0.800	0.737	0.633	0.792	0.712	7.49	7.49	7.63

Table 6.3 – Performance comparison between 3D-JND models used in SIQA on Waterloo IVC 3D IQ database (phase I).

Model	LCC			SROCC			RMSE		
	AS	S	ALL	AS	S	ALL	AS	S	ALL
BJND	0.635	0.736	0.673	0.577	0.690	0.638	13.2	10.1	12.8
JJND	0.613	0.716	0.660	0.616	0.723	0.663	13.5	10.4	13.0
SJND	**0.657**	**0.792**	**0.710**	**0.664**	**0.773**	**0.713**	**12.9**	**9.20**	**12.2**
HJND	0.652	0.744	0.685	0.647	0.735	0.674	13.0	10.0	12.6

low. In addition to the performance comparison mentioned above, we provide in Fig. 6.2 the scatter distributions of DMOS versus predicted scores obtained with the different 3D-JND models, and the non-linear fitting curves.

Table 6.3 corresponds to the performance comparison on Waterloo IVC 3D IQ database. According to these results, the metric using SJND gives the best performance, and this is particularly remarkable on the symmetric distortions. It should be noted that the metric using HJND model achieved better performance that of the metrics using BJND or JJND model. Since the overall disparity of the S3D images in this database is generally larger than that in LIVE 3D IQA data-base, the performance of the HJND may be improved owing to the consideration of depth saliency. Compared to LIVE 3D IQA database, the performance results on Waterloo IVC 3D IQ database are lower due to the high distortions levels (e.g., large amount of white Gaussian noise) in this database.

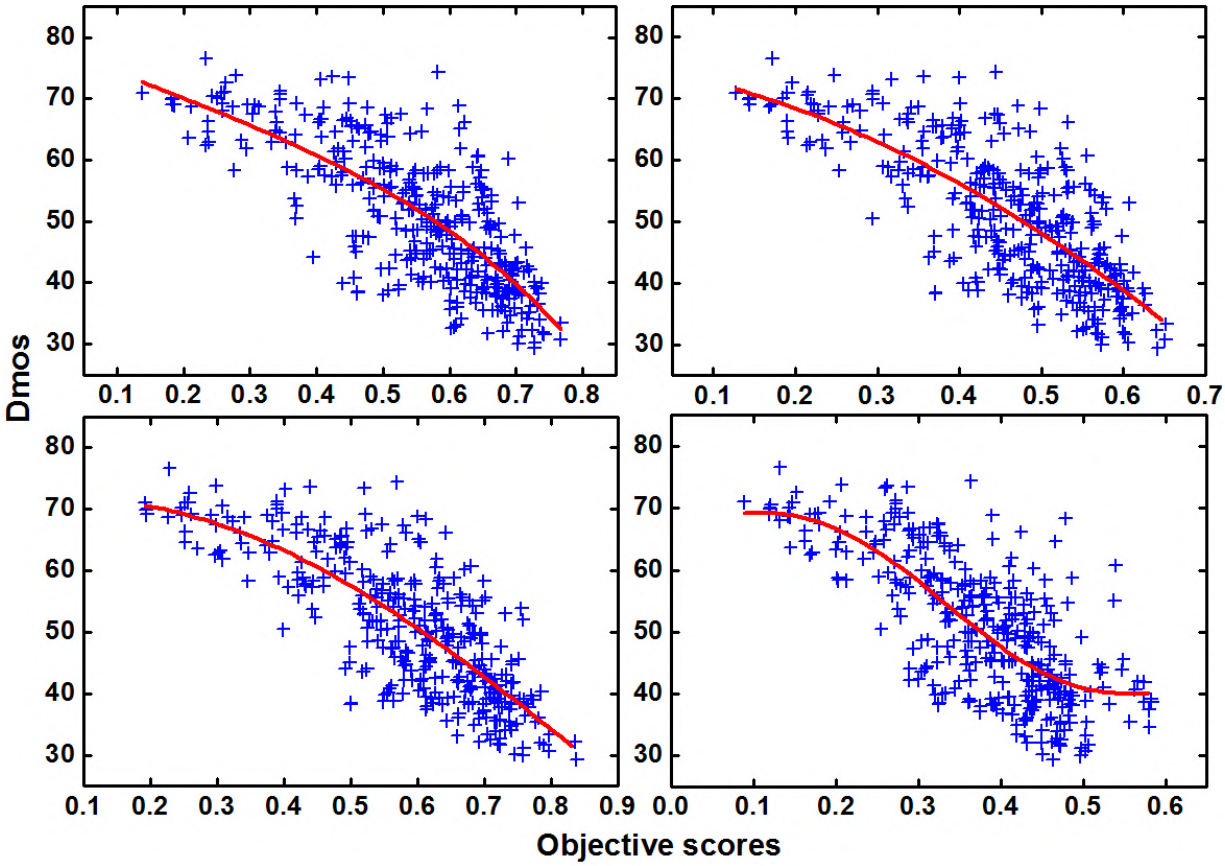


Figure 6.2 – Scatter distributions of DMOS versus predicted scores obtained by the BJND model (top left), JJND model (top right), SJND model (bottom left) and HJND model (bottom right) on LIVE 3D database. Red curve represents the non-linear fitting.

6.5 Conclusion

In this paper, we reviewed the literature of 3D-JND models and analyzed their performance by giving the pros and cons in addition to the targeted applications. In order to provide a more quantitative evaluation, we used the different models in a SIQA framework. This helped in investigating the capabilities of these models by comparing the performances on two widely used databases. The experimental results showed that using the SJND model achieved the best performance thanks to various visual masking characteristics considered in this model. In other words, the promising performance can be achieved when using a reliable 3D-JND model, which mimics the binocular vision properties and depth information as much as possible. This study allowed to determine the most important features, that will help in the construction of a more robust and adapted 3D-JND model.

Bibliography

[1] D Varuna SX De Silva, Warnakulasuriya Anil Chandana Fernando, Stewart T Worrall, SLP Yasakethu, and Ahmet M Kondoz. Just noticeable difference in depth model for stereoscopic 3d displays. In *IEEE International Conference on Multimedia and Expo (ICME)*, pages 1219–1224, 2010. 92

[2] Yin Zhao, Zhenzhong Chen, Ce Zhu, Yap-Peng Tan, and Lu Yu. Binocular just-noticeable-difference model for stereoscopic images. *IEEE Signal Processing Letters*, 18(1):19–22, 2011. 92, 93

[3] Xiaoming Li, Yue Wang, Debin Zhao, Tingting Jiang, and Nan Zhang. Joint just noticeable difference model based on depth perception for stereoscopic images. In *IEEE International Conference on Visual Communications and Image Processing (VCIP)*, pages 1–4, 2011. 92, 93, 94

[4] Feng Qi, Tingting Jiang, Xiaopeng Fan, Siwei Ma, and Debin Zhao. Stereoscopic video quality assessment based on stereo just-noticeable difference model. In *20th IEEE International Conference on Image Processing (ICIP)*, pages 34–38, 2013. 92, 94

[5] Rui Zhong, Ruimin Hu, Zhongyuan Wang, and Shizheng Wang. 3d hybrid just noticeable distortion modeling for depth image-based rendering. *Multimedia Tools and Applications*, 74(23): 10457–10478, 2015. 92, 95

[6] D Varuna SX De Silva, Erhan Ekmekcioglu, Warnakulasuriya Anil Chandana Fernando, and Stewart T Worrall. Display dependent preprocessing of depth maps based on just noticeable depth difference modeling. *IEEE Journal of Selected Topics in Signal Processing*, 5(2):335–351, 2011. 92, 96

[7] XK Yang, WS Ling, ZK Lu, Ee Ping Ong, and SS Yao. Just noticeable distortion model and its applications in video coding. *Signal Processing: Image Communication*, 20(7):662–680, 2005. 93

[8] Jian Sun, Yin Li, Sing Bing Kang, and Heung-Yeung Shum. Symmetric stereo matching for occlusion handling. In *IEEE International Conference on Computer Vision and Pattern Recognition (CVPR)*, volume 2, pages 399–406. IEEE, 2005. 94

[9] Xiaoming Li, Debin Zhao, Xiangyang Ji, Qiang Wang, and Wen Gao. A fast inter frame prediction algorithm for multi-view video coding. In *IEEE International Conference on Image Processing (ICIP)*, volume 3, pages III–417, 2007. 94

[10] Guangtao Zhai, Xiaolin Wu, Xiaokang Yang, Weisi Lin, and Wenjun Zhang. A psychovisual quality metric in free-energy principle. *IEEE Transactions on Image Processing*, 21(1):41–52, 2012. 94

[11] Feng Qi, Debin Zhao, Xiaopeng Fan, and Tingting Jiang. Stereoscopic video quality assessment based on visual attention and just-noticeable difference models. *Signal, Image and Video Processing*, 10(4):737–744, 2016. 94

[12] Chun-Hsien Chou and Yun-Chin Li. A perceptually tuned subband image coder based on the measure of just-noticeable-distortion profile. *IEEE Transactions on circuits and systems for video technology*, 5(6):467–476, 1995. 94

[13] Philipp Merkle, Aljoscha Smolic, Karsten Muller, and Thomas Wiegand. Multi-view video plus depth representation and coding. In *IEEE International Conference on Image Processing (ICIP)*, volume 1, pages I–201, 2007. 95

[14] Yun Zhang, Gangyi Jiang, Mei Yu, Ken Chen, and Qionghai Dai. Stereoscopic visual attention-based regional bit allocation optimization for multiview video coding. *EURASIP Journal on Advances in Signal Processing*, 2010:60, 2010. 95

[15] Harry Shum and Sing Bing Kang. Review of image-based rendering techniques. In *IEEE International Conference on Visual Communications and Image Processing (VCIP)*, volume 4067, pages 2–14, 2000. 95

[16] Nicolas Aspert, Diego Santa-Cruz, and Touradj Ebrahimi. Mesh: Measuring errors between surfaces using the hausdorff distance. In *IEEE International Conference on Multimedia and Expo (ICME)*, volume 1, pages 705–708, 2002. 95

[17] Seung-Won Jung and Sung-Jea Ko. Depth enhancement considering just noticeable difference in depth. *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences*, 95(3):673–675, 2012. 96

[18] Seung-Won Jung. A modified model of the just noticeable depth difference and its application to depth sensation enhancement. *IEEE Transactions on Image Processing*, 22(10):3892–3903, 2013. 96

[19] Gokce Nur, V De Silva, H Kodikara Arachchi, A Kondoz, W Fernando, Miguel O Martínez-Rach, and S Dogan. Sensitivity of the hvs for binocular disparity cue in 3d displays under different ambient illumination conditions. In *2012 IEEE International Conference on Consumer Electronics (ICCE)*, pages 459–460, 2012. 96

[20] Seung-Won Jung, Jae-Yun Jeong, and Sung-Jea Ko. Sharpness enhancement of stereo images using binocular just-noticeable difference. *IEEE Transactions on Image Processing*, 21(3):1191–1199, 2012. 97

[21] Xu Wang, Sam Kwong, and Yun Zhang. Considering binocular spatial sensitivity in stereoscopic image quality assessment. In *IEEE International Conference on Visual Communications and Image Processing (VCIP)*, pages 1–4, 2011. 97

[22] Walid Hachicha, Azeddine Beghdadi, and Faouzi Alaya Cheikh. Stereo image quality assessment using a binocular just noticeable difference model. In *IEEE International Conference on Image Processing (ICIP)*, pages 113–117, 2013. 97

[23] Sid Ahmed Fezza, Mohamed-chaker Larabi, and Kamel Mohamed Faraoun. Stereoscopic image quality metric based on local entropy and binocular just noticeable difference. In *IEEE International Conference on Image Processing (ICIP)*, pages 2002–2006, 2014. 97

[24] Sid Ahmed Fezza, Mohamed-Chaker Larabi, and Kamel Mohamed Faraoun. Asymmetric coding using binocular just noticeable difference and depth information for stereoscopic 3d. In *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 880–884, 2014. 97

[25] Gaofeng Zhu, Mei Yu, Gangyi Jiang, Zongju Peng, Feng Shao, Fen Chen, and Yo-Sung Ho. A novel macroblock level rate control method for stereo video coding. *The Scientific World Journal*, 2014, 2014. 97

[26] Yapei Zhu, Mei Yu, Xin Jin, Zongju Peng, Feng Shao, and Gangyi Jiang. Fast mode decision algorithm for multiview video coding based on binocular just noticeable difference. *Journal of Computers*, 9(10):2428–2434, 2014. 97

[27] Jing Wang, Mei Yu, Ting Luo, Feng Shao, Zongju Peng, and Gangyi Jiang. Joint just noticeable distortion based stereo image watermarking method with self-recovery. *Future Information Engineering (2*, 49:51, 2014. 97

[28] D Varuna SX De Silva, Warnakulasuriya Anil Chandana Fernando, and H Kodikara Arachchi. A new mode selection technique for coding depth maps of 3d video. In *IEEE International Conference on Acoustics Speech and Signal Processing (ICASSP)*, pages 686–689. IEEE, 2010. 97

[29] Geoffrey Egnal and Richard P Wildes. Detecting binocular half-occlusions: Empirical comparisons of five approaches. *IEEE Transactions on pattern analysis and machine intelligence*, 24(8):1127–1133, 2002. 97

[30] Daniel Scharstein and Chris Pal. Learning conditional random fields for stereo. In *IEEE International Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1–8, 2007. 97

Chapter 7

Paper II: A Survey of Stereoscopic 3D Just Noticeable Difference Models

Y. Fan[1,2], M.-C. Larabi[1], F. A. Cheikh[2], C. Fernandez-Maloigne[1]

[1]XLIM UMR CNRS 7252, University of Poitiers, France
[2]Norwegian Colour and Visual Computing Lab, Norwegian University of Science and Technology, Gjøvik, Norway

in *IEEE Access*, vol. 7, pp. 8621-8645, 2019.

Abstract

Just noticeable difference (JND) for stereoscopic 3D content reflects the maximum tolerable distortion, it corresponds to the visibility threshold of the asymmetric distortions in the left and right contents. The 3D-JND models can be used to improve the efficiency of the 3D compression or the 3D quality assessment. Compared to 2D-JND models, the 3D-JND models appeared recently and related literature is rather limited. In this paper, we give a deep and comprehensive study of the pixel-based 3D-JND models. To our best knowledge, this is the first review on 3D-JND models. Each model is briefly described by giving its rationale and main components in addition to providing exhaustive information about the targeted application, the pros, and cons. Moreover, we present the characteristics of the human visual system (HVS) presented in these models. In addition, we analyze and compare the 3D-JND models thoroughly using qualitative and quantitative performance evaluation based on Middlebury stereo datasets. Besides, we measure the JND thresholds of the asymmetric distortion

based on psychophysical experiments and compare these experimental results to the estimates from the 3D-JND models in order to evaluate the accuracy of each model.

Index terms– Human visual system, just noticeable difference (JND), 3D compression, 3D-JND models, 3D quality assessment

7.1 Introduction

The digital era has allowed simplifying the spread of Stereoscopic 3D (S3D) technologies in different application domains (*e.g.,* 3D-Cinema, 3D-TV) in recent decades. The most important aspect is that S3D can provide viewers with favorable immersion and natural sensation thanks to both binocular and monocular depth cues. However, there is a noticeable decrease in the attractiveness of S3D technology during the last few years. This is due to the complexity of such a content and the undesirable effect that it may generate from a perceptual point of view. S3D brought many technical challenges in the field of image and video processing linked to quality assessment, enhancement, and compression. Specifically, the main challenges are evaluating and optimizing the S3D imaging system with respect to storage capacity and quality of the user's experience (QoE).

To do so, it is important to understand and explore the different perceptual processes of the human visual system (HVS). For decades, the scientific community has exhaustively studied two-dimensional (2D) perception. Several properties and models of the HVS have been successfully exploited like the just noticeable difference (JND) models [1, 2]. These models refer to thresholds, depending on luminance, contrast, and temporal/spatial frequency of the local regions in the image, beyond which a distortion is visible. In other words, a given distortion cannot be perceived by the HVS if it is lower than the JND threshold. Therefore, JND models have been widely applied in visual signal processing, especially in compression and image processing [3, 4].

Over the last decades, numerous 2D-JND models have been developed either in transform domain [5–9], or in pixel domain [10–16]. Comprehensive reviews on 2D-JND models have been recently done in [17, 18]. 2D-JND models are generally proposed based on specific characteristics of monocular vision, which does not fit with the complexity of 3D perception requiring specific models accounting for both monocular and binocular depth cues.

To date, only a few 3D-JND models have been proposed because of the complex processes to be modeled [19–25]. Additional research efforts are undoubtedly needed to achieve a more accurate and efficient modeling that can effectively improve the performance of S3D applications (*e.g.,* compression, quality assessment, watermarking...). To the best of our knowledge, no review exists for the comparison of 3D-JND models in the framework of image quality assessment (IQA) [26].

In this paper, we propose a comprehensive survey of 3D-JND models. Since most of the existing 3D-JND models are computed in the pixel domain, we focus this survey on pixel-based 3D-JND models. Each model is briefly described by giving its rationale and main components in addition to providing exhaustive information about the targeted applications, the pros, and cons. The paper also provides a

brief review of visual masking effects considered in these models. Furthermore, we present a thorough comparative analysis between the 3D-JND models using qualitative and quantitative performance evaluation. This study aims at comparing the distortion masking ability of the 3D-JND models using the widely used Middlebury stereo datasets[27–29], and evaluating the accuracy of these models using psychophysical experiments.

In summary, the major contributions of this paper include:

- An exhaustive review of the 3D-JND models;
- Creation of a dataset composed of asymmetrically distorted S3D images using 2D texture images from ETHZ Synthesizability Dataset [30];
- An extensive experimental comparison with qualitative and quantitative performance evaluation of the 3D-JND models.

The remainder of the paper is organized as follows. In Section 7.2, we describe the main visual characteristics largely employed by 3D-JND models. Section 7.3 reviews the existing 3D-JND models separately. In Section 7.4, the models are thoroughly analyzed and compared. Section 7.5 presents the experimental results on the performance comparison of the 3D-JND models in terms of masking ability and accuracy using two different datasets. This paper ends with some conclusions and discussion of open issues in Section 7.6.

7.2 Visual characteristics for 3D-JND models

Over the last decade, HVS has been studied based on physiological and psychophysical experiments [31]. HVS models are widely used in image/video processing [32], since such models can simplify and mimic the behaviors of the so complex HVS system. For instance, 3D-JND models, aiming to determine whether the distortion is undetectable by the HVS in a given block, can be used to improve the coding efficiency (CE) for S3D image/video. Therefore, understanding and studying the HVS mechanisms of the HVS are critical for developing a more reliable 3D-JND model. In general, 3D-JND models from the literature account for the HVS sensitivity and VM effects.

In this section, we explain the factors affecting the HVS sensitivity related to S3D content. Most 3D-JND models are developed by combining some of the factors including spatial contrast sensitivity, luminance adaptation, contrast masking, binocular masking, temporal masking, and depth masking.

7.2.1 Spatial and temporal contrast sensitivity

The luminance contrast sensitivity (CS) of the HVS describes the ability to perceive the various frequencies of stimuli with different luminance contrasts [33, 34]. This sensitivity for a given target can be determined by measuring the minimum contrast necessary for an observer to detect the target. Accordingly, the CS depends on the spatial frequency of the visual stimuli [35]. Several psychophysical

experiments measured the CS by determining the minimum contrast to make a sine-grating of a given spatial frequency visible in an image [36, 37]. The relationship between the CS and the spatial frequency of the grating in the image is typically modeled by the achromatic contrast sensitivity function (CSF) having a band-pass behavior [38, 39]. In addition to achromatic CSF models, chromatic CSF models having a low-pass behavior were proposed in [40, 41]. Moreover, some spatial-temporal CSF models have taken the temporal CS into account [42, 43]. Recently, Rousson *et al.* [44, 45] proposed a CSF for observing stereoscopic content on S3D display. Moreover, some 2D-JND models were developed using CSF [6–8].

7.2.2 Masking effects

The JND thresholds for S3D content depend not only on the spatial-temporal CS but also on visual masking effects (MEs). The latter characteristics are often used in 3D-JND modeling. The visual masking (VM), a perceptual phenomenon, describes the visibility reduction (masking effect) of one stimulus (*e.g.,* the target) to human eyes in the presence of another (the masker) where these stimuli are coincident in space and simultaneous in time [46]. For 2D content, the masking effect (ME) is modeled by using spatial frequency [47], orientation [48], motion (commonly in video) [49] of both image signals. For S3D, the disparity/depth should be considered in VM [50].

7.2.2.1 Luminance adaptation

According to [51], the HVS has the ability to quickly adjust to the levels of background light in order to distinguish objects. This ability is known as luminance adaptation (LA). It is related to background luminance masking (LM). As described in [5], human eyes are more sensitive to luminance variation/contrast than absolute luminance intensity. In other words, the salience of an object in an image could be more influenced by the difference between its luminance and the luminance of its adjacent background than by its own absolute luminance. LA allows adjusting the sensitivity of the HVS in response to the relative luminance variations. LA can be measured in an increment threshold experiment [51] that describes the just-noticeable luminance difference of a stimulus as a function of the background luminance intensity. The experimental results showed that the ratio between the just-noticeable luminance difference and background luminance, known as Weber's fraction [6, 52], is approximatively constant for a wide range of luminance intensities. The luminance contrast LC_w can be defined as:

$$LC_w = \frac{\Delta L}{L_{bg}}, \quad \Delta L = |L - L_{bg}|, \tag{7.1}$$

where L is the luminance of a test stimulus, and L_{bg} is the surrounding background luminance. For the scene with L_{bg} of high levels, LC_w remains nearly constant as L_{bg} increases. LC_w is considered in this case as Weber fraction. On the other hand, LC_w increases when L_{bg} decreases in the case of low background luminance. This describes a high visibility threshold of luminance contrast in dark

regions of the scene. In [5, 10], the authors estimated LA (*i.e.,* visibility threshold of LM) of an image pixel in the pixel domain as follows:

$$LA_{CY}(i,j) = \begin{cases} c_1 \times (1 - \sqrt{\frac{\overline{L_{bg}}(i,j)}{127}}) + c_3, & if \quad \overline{L_{bg}}(i,j) \leq 127 \\ c_2 \times (\overline{L_{bg}}(i,j) - 127) + c_3, & otherwise \end{cases}, \tag{7.2}$$

where c_1, c_2, and c_3 are constants, and are set to 17, 3/128, and 3 respectively for a viewing distance of six times the targeted image height [5, 10]. It should be noted that the value of c_1 is proportional to the viewing distance. $\overline{L_{bg}}(i,j)$ is the average background luminance at pixel of coordinate (i,j), and is computed by:

$$\overline{L_{bg}}(i,j) = \frac{1}{32} \sum_{x=1}^{5} \sum_{y=1}^{5} I(i-3+x, j-3+y) \times B(x,y), \tag{7.3}$$

where $I(i,j)$ is the luminance intensity at pixel (i,j), and the kernel of low-pass filter B is represented as:

$$B = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 2 & 2 & 2 & 1 \\ 1 & 2 & 0 & 2 & 1 \\ 1 & 2 & 2 & 2 & 1 \\ 1 & 1 & 1 & 1 & 1 \end{bmatrix} \tag{7.4}$$

In addition to the method described above, Zhao *et al.* measured the visibility threshold of LA based on psychophysical experiments [20]. It was conducted using binocular patterns (corresponding to S3D images) that are asymmetrically distorted by noise, as shown in Fig. 7.1(a). During reading, the para-fovea could process the information within 5° of visual angle of its fixation point, while the fovea processes the information located within 2° around the fixation point [53, 54]. The fovea and para-fovea in human eye contribute jointly to the perception of a fixated region and its surrounding regions perceived in the range of para-fovea. As shown in Fig. 7.1(b), the visual stimulus in an image is modeled as 2° × 2° square (called R_3) corresponding the fovea, and 5° × 5° square (called R_2) covering the para-fovea. The aim of the psychophysical experiment of Fig. 7.1(a) is to determine the amplitude of the maximum noise LA_{Z_r} (*i.e.,* visibility threshold of LA) injected in one view (*e.g.,* right view) without evoking binocularly perceptible difference due to LM, under a background luminance L_{bg_l} in this view (*e.g.,* left view), for a given noise with amplitude n_l injected in other view (*e.g.,* left view). More specifically, given the background luminance L_{bg_r} and noise amplitude n_l in the left view, an observer adjusted the noise amplitude in the right view n_r to binocularly detect the just noticeable noise. The amplitude of the aforementioned noise on a given pixel in the right image,

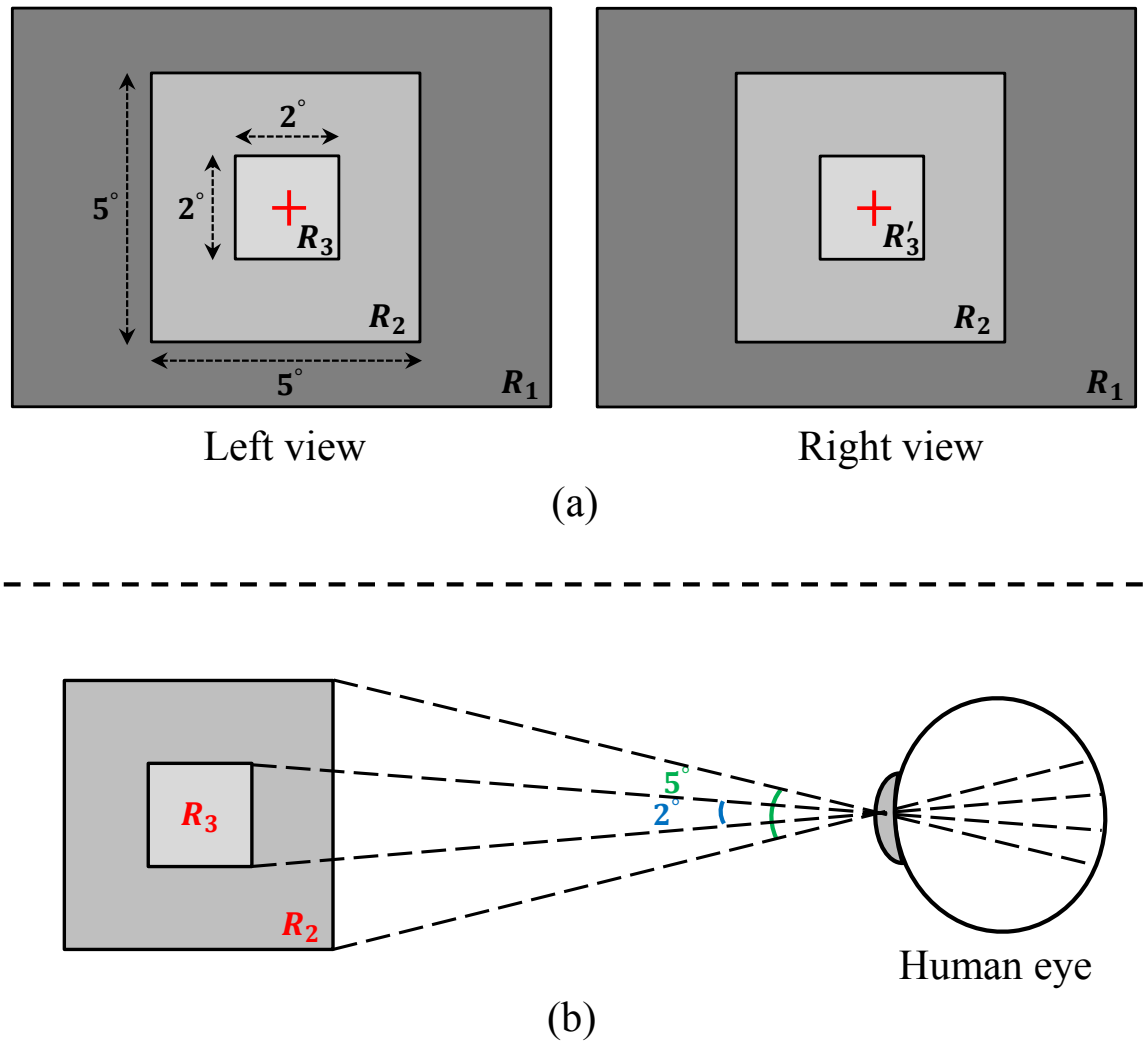


Figure 7.1 – (a) Binocular patterns used for modeling luminance adaptation (LA), (b) Schematic illustration of regions covered by the fovea and the para-fovea. Note that R_1, R_2, and $R_3/R_3^{'}$ correspond to three regions of the retinal image: the peri-fovea, the para-fovea covered by a square with $5^\circ \times 5^\circ$ of visual angle, and the fovea covered by $2^\circ \times 2^\circ$ square. In this case, the level of average background luminance corresponds to that of background luminance since the squares are uniform. The luminance intensity of R_1 is set to 112. The luminance in R_2 represents the background luminance L_{bg}. The luminance levels of R_3 and $R_3^{'}$ are different, and are represented as $L_{bg} \pm n_l$ and $L_{bg} \pm n_r$, respectively, where n_l and n_r denote the amplitude of the bipolar patterns noise injected in the left and right views respectively.

namely $LA_{Z_r}(i,j)$, is computed by:

$$LA_{Z_r}(i,j) = A_{max}(L_{bgl}(i,j+d)) \times \left[1 - (\frac{n_l(i,j+d)}{A_{max}(L_{bgl}(i,j+d))})^{\gamma}\right]^{\frac{1}{\gamma}}, \tag{7.5}$$

where λ, set to 1.25, allows adjusting the noise influence in left view. d is the disparity value at pixel (i,j) corresponding to the horizontal shift of the pixel between right to left view. It should be noted

that $L_{bg_r}(i,j)$ corresponds to the average background luminance of the pixel (i,j), which is determined by averaging the luminance intensity in the 5×5 window. $A_{max}(L_{bg_l})$ denotes the visibility threshold of right view LA if there is no noise in the left view. $A_{max}(L_{bg})$ is calculated by the following formula:

$$A_{max}(i,j) = \begin{cases} a \times [L_{bg}(i,j)^2 - 96L_{bg}(i,j)] + 8, & if\ 0 \leq L_{bg}(i,j) < 48 \\ b \times [L_{bg}(i,j)^2 - 32L_{bg}(i,j)] + 1.7, & if\ 48 \leq L_{bg}(i,j) \leq 255 \end{cases} \tag{7.6}$$

where $a = 2.7 \times 10^{-3}, b = 1.0 \times 10^{-4}$. $LA_r(i,j)$ becomes maximum, namely $A_{max}(i,j+d)$ when $n_l(i,j+d) = 0$.

7.2.2.2 Contrast masking

Contrast masking (CM) describes the VM effects in presence of two or more stimuli if these stimuli are of similar or same contrast/spatial non-uniformity (*e.g.,* spatial frequency, orientation) [55]. CM is also known as spatial masking. CM explains the fact that the presence of one stimulus reduces the ability of a subject to detect a targeted stimulus. For instance, HVS could tolerate more noises in textured regions than smooth regions since the spatial frequencies in noise and textured regions are similar.

According to previous studies, the visibility threshold of CM can be defined as a function of the average background luminance $\overline{L_{bg}}(i,j)$ and the amplitude of luminance edge (namely, edge height) $Eh(i,j)$, which refers to the contrast degree. For a viewing distance of six times of the targeted image height, Chou and Li [5] computed the visibility threshold related to contrast masking $CM_C(i,j)$ as follows:

$$CM_C(i,j) = 0.01\overline{L_{bg}}(i,j) \times [0.01G_m(i,j) - 1] + 0.115G_m(i,j) + c_4, \tag{7.7}$$

where c_4 adjusts the average amplitude of $CM_C(i,j)$, and is set to 0.5 in [5]. $G_m(i,j)$ denotes the maximum gradient at pixel (i,j) over four directions, and is computed as follows:

$$G_m(i,j) = \max_{s=1,2,3,4} \{grad_s(i,j)\}, \tag{7.8}$$

with

$$grad_s(i,j) = \frac{1}{16} \sum_{x=1}^{5} \sum_{y=1}^{5} I(i-3+x, j-3+y) \times g_s(x,y), \tag{7.9}$$

where $g_s(x,y)$ are kernels corresponding to four directional hight-pass filters. These four kernels are

defined in equation 7.10 and 7.11:

$$g_1 = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 1 & 3 & 8 & 3 & 1 \\ 0 & 0 & 0 & 0 & 0 \\ -1 & -3 & -8 & -3 & -1 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}, g_2 = \begin{bmatrix} 0 & 0 & 1 & 0 & 0 \\ 0 & 8 & 3 & 0 & 0 \\ 1 & 3 & 0 & -3 & -1 \\ 0 & 0 & -3 & -8 & 0 \\ 0 & 0 & -1 & 0 & 0 \end{bmatrix} \tag{7.10}$$

$$g_3 = \begin{bmatrix} 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 3 & 8 & 0 \\ -1 & -3 & 0 & 3 & 1 \\ 0 & -8 & -3 & 0 & 0 \\ 0 & 0 & -1 & 0 & 0 \end{bmatrix}, g_4 = \begin{bmatrix} 0 & 1 & 0 & -1 & 0 \\ 0 & 3 & 0 & -3 & 0 \\ 0 & 8 & 0 & -8 & 1 \\ 0 & 3 & 0 & -3 & 0 \\ 0 & 1 & 0 & -1 & 0 \end{bmatrix} \tag{7.11}$$

Since HVS is more sensitive to the distortion around edge regions than that in textured regions, CM in edge and textured regions should be considered separately. Yang *et al.* [10] found that Chou and Li approach overestimates the visibility threshold of CM for edge regions. Thus, they used the Canny detector to decrease the thresholds for edge regions, and divided CM into texture masking (TxM) and edge masking (EM). Note that we focus only on the luminance component here. For a viewing distance of six times the targeted image height, Yang *et al.* calculated the visibility threshold of $CM_Y(i,j)$ by:

$$CM_Y(i,j) = 0.117 \times W_{ed} \times G_m(i,j), \tag{7.12}$$

where $G_m(i,j)$ describes the maximal weighted average of gradients for the pixel (i,j). W_{ed} denotes the edge-related weight of the pixel (i,j), and is defined as:

$$W_{ed}(i,j) = Ed(i,j) * h_{lp}, \tag{7.13}$$

where Ed is the edge map estimated by Canny's detector [56] with a threshold of 0.5. $*$ represents the convolution operator, and h_{lp} is a $k \times k$ Gaussian low-pass filter having σ as a standard deviation. In [10], σ and k are set to 0.8 and 7, respectively.

Similarly, Liu *et al.* [12] employed the image decomposition method [57] to decompose the targeted image into structural and textural regions that lead to EM with Canny's detector and TxM, respectively. Therefore the visibility thresholds of CM due to edge and texture are described by:

$$CM_L(i,j) = 0.117 \times (w_e \cdot CM_e(i,j) + w_t \cdot CM_t(i,j)), \tag{7.14}$$

where $w_e = 1$ and $w_t = 3$ are the weights for edge masking (CM_e) and texture masking (CM_t) respectively. This means that the CM effect is stronger in textured regions than edge regions. CM estimation proposed by Chou and Li (see Eqs. 7.7, 7.8 and 7.9) was used to calculate CM_e and CM_t

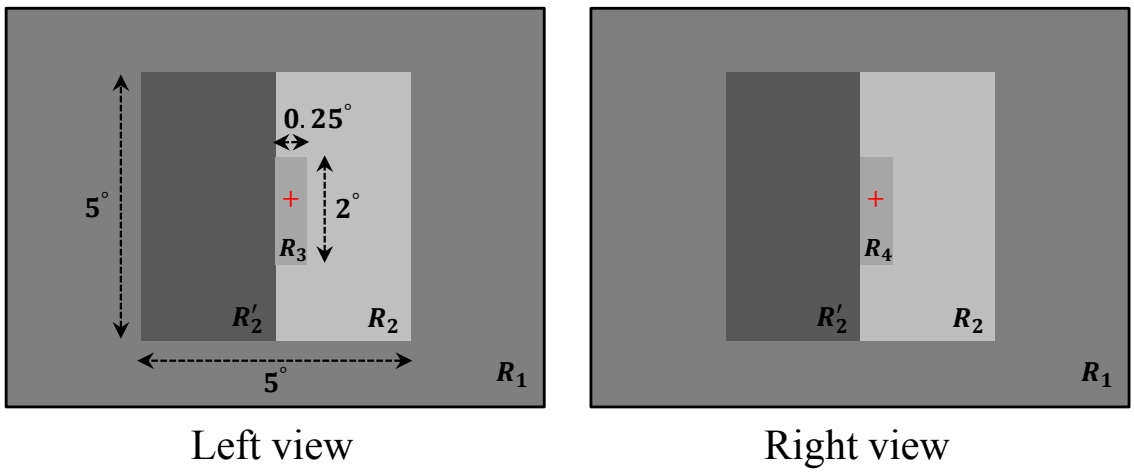


Figure 7.2 – Binocular patterns used in the experiment for modeling contrast masking. Note that R_1, $R_2/R_2^{'}$, and R_3/R_4 correspond to three regions of a human retinal image: the peri-fovea, the para-fovea covered by a square with $5^\circ \times 5^\circ$ of visual angle, and the slice in fovea region with 2° height and 0.25° width. The luminance intensity of R_1 is set to 112. The background in patterns consists of 2 regions: R_2 with luminance L_{bg} and $R_2^{'}$ with luminance of $L_{bg} - Eh$, where Eh represents the edge height. The luminance levels of R_3 and R_4 are different, and equal to $L_{bg} \pm n_l$ and $L_{bg} \pm n_r$, respectively. n_l and n_r denote the amplitude of the bipolar patterns noise injected in the left and right views, respectively.

for both structural and textural images, respectively.

Zhao *et al.* [20] estimated the visibility threshold of CM using binocular patterns shown in Fig. 7.2 Similar to the psychophysical experiment used in LA, the subjects are asked to focus on R_3 and R_4, and adjust the luminance of the noise in the right view, n_r until the noise around the edges is binocularly just detected, L_{bg} and n_l being fixed. The just noticeable noise pair $\{n_l,\ n_l\}$ are then recorded. In [20], the authors conducted several experiments to determine different noise pairs $\{n_l,\ n_l\}$ under different L_{bg} or n_l. Thus, the visibility threshold due to CM of the right view, which depends on L_{bg_l} and the left image I_l, is expressed by:

$$CM_{Z_r}(i,j) = A_{max}(L_{bg_l}(i, j+d)) + F(L_{bg_l}(i, j+d)) \times Eh(I_l(i, j+d)), \tag{7.15}$$

where d is the disparity, A_{max} is estimated by Eq. 7.6, and I_l denotes the left image. F is a fitting function according to the average background luminance of one view, L_{bg_l}, is experimentally described as:

$$F(i,j) = -10^{-6} \times \left[0.7L_{bg}(i,j)^2 + 32L_{bg}(i,j)\right] + 0.07. \tag{7.16}$$

The edge height $Eh(i,j)$ of one pixel in image I is calculated using the following formula:

$$Eh(i,j) = \sqrt{E_h^2(i,j)) + E_v^2(i,j))}, \tag{7.17}$$

where

$$E_k(i,j) = \frac{1}{24}\sum_{h=1}^{5}\sum_{v=1}^{5} I(i-3+h, j-3+v) \times G_k(h,v),\ \ k = h, v, \tag{7.18}$$

$$G_h = \begin{bmatrix} -1 & -2 & 0 & 2 & 1 \\ -2 & -3 & 0 & 3 & 2 \\ -3 & -5 & 0 & 5 & 3 \\ -2 & -3 & 0 & 3 & 2 \\ -1 & -2 & 0 & 2 & 1 \end{bmatrix}, G_v = \begin{bmatrix} 1 & 2 & 3 & 2 & 1 \\ 2 & 3 & 5 & 3 & 2 \\ 0 & 0 & 0 & 0 & 0 \\ -2 & -3 & -5 & -3 & -2 \\ -1 & -2 & -3 & -2 & -1 \end{bmatrix} \tag{7.19}$$

7.2.2.3 Binocular masking

Binocular masking (BM) describes the interocular interaction/masking in the case of two dissimilar stimuli presented to both eyes [58, 59]. The limited distortion in one view is influenced/masked by the other so that the two views can be successfully fused to a 3D image. This visual phenomenon is known as the binocular fusion (BF) [60]. The BM reveals that the HVS can tolerate a certain limited asymmetric distortion in one view that does not impair 3D perception. For instance, the subject perceives a stereo pair where the blur is introduced in the right image while the left image is kept unchanged. The fused 3D image is slightly blurred since the blur effect is reduced by the left image. Zhao *et al.* [20] modeled the BM using LA and CM as described previously. In addition, Qi *et al.* [23] conducted a psychophysical experiment similar to one of Fig. 7.1 in order to determine the visibility threshold of the right view $BM_r(i,j)$ relative to left one due to BM, which is approximately described as:

$$BM_r(i,j) = \begin{cases} 15 \times (1 - \sqrt{\frac{\overline{L_{bg_l}}(i,j)}{127}}) + 5.08, & if\ \ \overline{L_{bg_l}}(i,j) \leq 127 \\ 0.04 \times (\overline{L_{bg_l}}(i,j) - 127) + 5.08, & otherwise \end{cases}, \tag{7.20}$$

where $\overline{L_{bg_l}}$ is the average background luminance that is calculated by using Eq. 7.3 and 7.4. The BM described above is similar to the LA shown in Eq. 7.2, but the visibility threshold of one view is calculated based on the luminance intensity of the other view.

7.2.2.4 Temporal masking

The visual MEs mentioned above are dedicated to images, while the one discussed here focuses on video. Based on the free energy principle, HVS adaptively conceals the disorder tendency information in a continued movement scene, and tries to focus on the definite content of the input image [61]. This phenomenon can be modeled as the temporal masking (TM) caused by temporal discontinuities in intensity, such as motion when watching a video [62, 63]. Yang *et al.* indicated that TM is proportional to motion [64]. Inspired by [65], Zhou *et al.* [22] estimated the visibility threshold of TM using the

temporal JND (TJND) model described as follows:

$$TJND_Z(i,j,t) = \begin{cases} max\left\{\tau, \frac{H}{2} \times e^{\frac{-0.15}{2\pi}\times[\Delta(i,j,t)+255]} + \tau\right\}, & if \Delta(i,j,t) \leq 0 \\ max\left\{\tau, \frac{K}{2} \times e^{\frac{-0.15}{2\pi}\times[255-\Delta(i,j,t)]} + \tau\right\}, & otherwise \end{cases}, \tag{7.21}$$

where

$$\Delta(i,j,t) = \frac{I(i,j,t) - I(i,j,t-1) + \overline{L_{bg}}(i,j,t) - \overline{L_{bg}}(i,j,t-1)}{2}, \tag{7.22}$$

$TJND_Z(i,j,t)$ is the TJND threshold of a pixel (i,j) of a given frame of multi-view plus depth video. $I(i,j,t)$ and $\overline{L_{bg}(i,j,t)}$ denote the luminance and the average background luminance of the pixel (i,j) respectively. τ, H, and K are set to 8, 3.2, and 0.8, respectively. $\Delta(i,j,t)$ represents the luminance difference of the inter-frame. Larger $\Delta(i,j,t)$ values result in higher TM thresholds. $H > K$ reveals that the changes from high to low luminance can bring more TM than the changes from low to high luminance. Similarity, Qi *et al.* [23] estimated TM with the following formula:

$$TJND_Q(i,j,t) = max\left\{f_1(i,j,t), f_2(i,j,t)\right\}, \tag{7.23}$$

where

$$f_1(i,j,t) = max\left\{abs(CM_C(i,j,t) - CM_C(i,j,t-1)), \Delta\overline{CM_C}\right\}, \tag{7.24}$$

$$f_2(i,j,t) = max\left\{abs(LA(i,j,t) - LA(i,j,t-1)), \Delta\overline{LA}\right\}, \tag{7.25}$$

$CM_C(i,j,t)$ and $LA(i,j,t)$ are the visibility thresholds of CM and LA at pixel (i,j) in the frame t $(t \geq 2)$, respectively. Eqs. 7.2 and 7.7 were used to calculate the $CM^C(i,j,t)$ and $LA(i,j,t)$ respectively. $\Delta\overline{CM_C}$ and $\Delta\overline{LA}$ denote respectively the average difference between two adjacent frames of all $\overline{CM_C}$ and $\overline{LA}$ of the whole video:

$$\Delta\overline{ME} = \frac{1}{N}\sum_{t=2}^{N}\left[ME(i,j,t) - ME(i,j,t-1)\right], \tag{7.26}$$

where ME represents LA or CM_C, and N is the number of frames. TM thresholds for the left and right views are calculated separately.

7.2.2.5 Depth masking

In addition to 2D VM effects, binocular depth masking (DM) have been studied by De Silva *et al.* [19, 66], who demonstrated that the subject cannot perceive sufficiently small depth changes on the scene. Moreover, the studies in indicated that the quality of the S3D video (with color plus depth representation) hardly changes with the compression of the depth map [67, 68]. In this circumstance, De Silva *et al.* first derived the visibility threshold relative to DM, which is known as the just noticeable difference in depth (JNDD). As described in [19], the JNDD threshold is mainly dependent on the

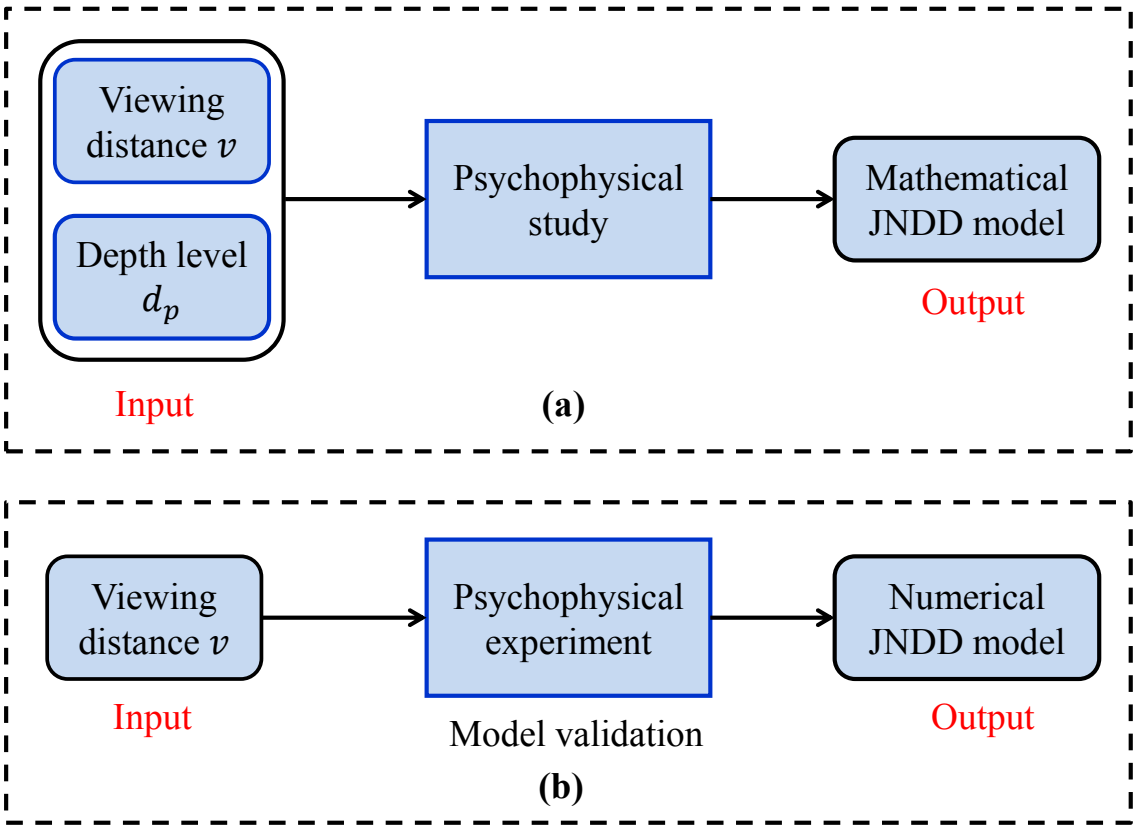


Figure 7.3 – Frameworks of the JNDD model for: (a) real-world 3D perception, and (b) S3D display.

viewing distance and the displayed depth level of the image. Based on the existing psychophysical models, a mathematical JNDD model for real-world viewing scenarios is defined as follows:

$$JNDD = 10^{[0.94\times\log_{10}(\nu)-2.25]} + K_w \times |dp|\,, \tag{7.27}$$

where K_w is the Weber constant and experimentally set to 0.03. dp is the simulated depth level with meter unit, while v denotes the distance between the subject's eyes and the fixation point *i.e.,* the screen.

As shown in Fig. 7.3(a), JNDD is estimated according to v and K_w. In fact, the JNDD in Eq. 7.25 can split in two parts: first, the visibility thresholds $JNDD_{d=0}$ when the simulated depth is equal to zero. Then the visibility thresholds $JNDD_{|d|>0}$ in the case of nonzero disparity. $JNDD_{d=0}$ and $JNDD_{|d|>0}$ are described by :

$$JNDD_{d=0} = 10^{[0.94\times\log_{10}(\nu)-2.25]} \tag{7.28}$$

and

$$JNDD_{|d|>0} = 0.03 * |dp|\,. \tag{7.29}$$

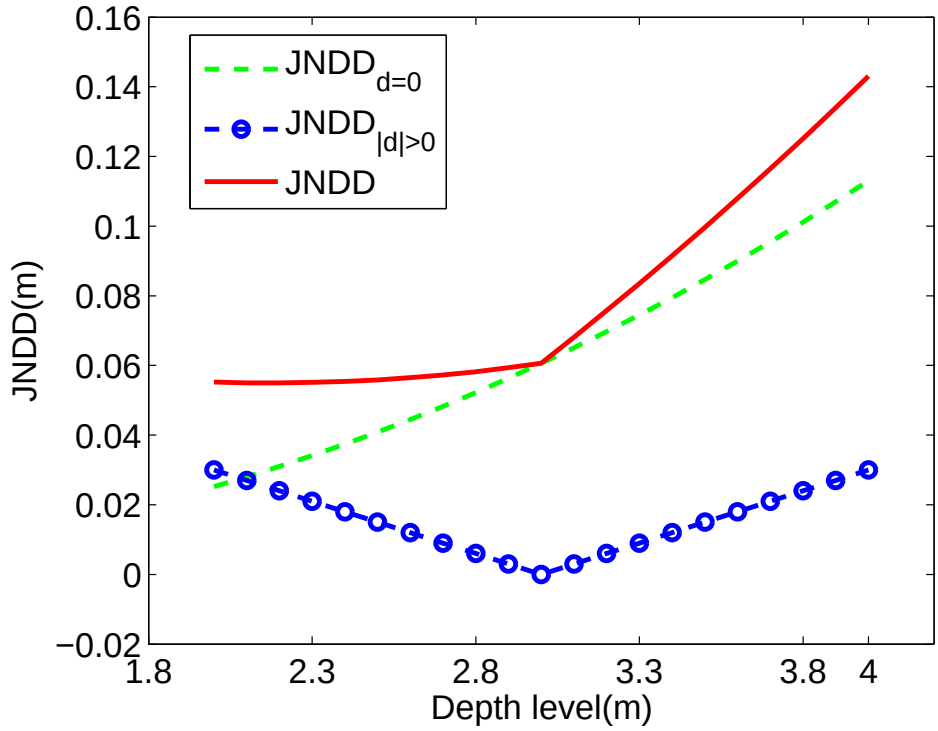


Figure 7.4 – JNDD thresholds $JNDD$, $JNDD_{d=0}$ and $JNDD_{|d|>0}$ according to simulated distance for real-world 3D perception. The viewing distance is set to 3 m.

According to Eqs. 7.25, 7.26, and 7.27, the JNDD curves corresponding to $JNDD$, $JNDD_{d=0}$, $JNDD_{|d|>0}$ are depicted in Fig. 7.4. One can notice that the linear summation between green and blue curves derive the red curve (JNDD thresholds).

7.3 3D-JND models

In this section, we give a brief introduction of the existing 3D-JND models. Specifically, each model is described with its framework as well as its mathematical expression. It should be noted that all 3D-JND presented models measure the achromatic JND thresholds. In other words, only JND thresholds of the luminance component of the color image are taken into account.

7.3.1 JNDD model

A JNDD model is addressed firstly in [19, 66, 68], which indicates that a human subject could not perceive depth changes below the JNDD threshold. The visibility thresholds due to DM (described in Section 7.2.2.5) could not be applied for S3D displays. This is due to the fact that the viewing distance rarely changes when a subject watches a S3D image/video on a S3D display. Therefore, De Silva *et al.* ignored the viewing distance, and only considered the depth level in JNDD estimation (see

Fig. 7.3(b)). They conducted a psychophysical experiment to validate the $JNDD_{|d|>0}$ (as shown in Fig. 7.4), and to measure the JNDD thresholds on a S3D display using 2D-plus-depth videos.

As described in [19, 66], the simulated depth level is 8 bits, where 0 and 255 denote the farthest and the nearest positions apart from the subject, respectively. Objects on the display with a depth value of 128 have zero disparity. The plane with zero disparity, called zero parallax plane, is the co-planar with the display plane. In the psychophysical experiment, two identical (left and right) objects were first displayed at the same depth level, namely initial depth level, and then the depth level of one object is changed gradually. The subjects were asked to inform about depth changes between the two objects when perceived. Various initial depth levels of the two objects have been investigated and the final threshold is obtained by averaging the JNDD values of all subjects. By analyzing the JNDD values according to different initial depth levels, the JNDD threshold $JNDD_{num}$, for a given initial depth value dp_i, is modeled as follows:

$$JNDD_{num}(i,j) = \begin{cases} 21, & if\ 0 \leq dp_i(i,j) < 64 \\ 19, & if\ 64 \leq dp_i(i,j) < 128 \\ 18, & if\ 128 \leq dp_i(i,j) < 192 \\ 20, & if\ 192 \leq dp_i(i,j) < 225 \end{cases} \tag{7.30}$$

where $dp_i(i,j)$ is the depth value (in pixels) of the original depth map at the pixel coordinate (i,j). The JNDD thresholds in Eq. 7.28 correspond to the symmetrical shape of the $JNDD_{|d|>0}$ in Fig. 7.4 except the zero disparity level (128). Moreover, according to the experimental results, the expert's subjects are more sensitive to depth changes than the non-expert ones.

7.3.2 BJND model

Meanwhile, another 3D-JND model, namely binocular JND (BJND) was proposed by Zhao *et al.* [20]. It reveals the threshold in inter-difference between the left and right views that human can recognize. The BJND model investigates the properties of the binocular vision in response to asymmetric noise in a stereo pair based on the VM effects of the HVS. These considered in this model consists of LA (see Section 7.2.2.1) and CM (see Section 7.2.2.2). Eqs. 7.5 and 7.15 are used to calculate the visibility thresholds related to LA (LA_{Z_r}) and CM (CM_{Z_r}), respectively. Fig. 7.5 illustrates the framework of calculating the BJND thresholds of one view of the stereopair. It is worth noting that there are two (left and right) BJND thresholds for each stereopair, since the BJND of one view indicates the maximum distortions that can be introduced in this view without evoking binocularly visible differences, given the distortions in the corresponding pixels of the other view. Like this, BJND of the left or right view

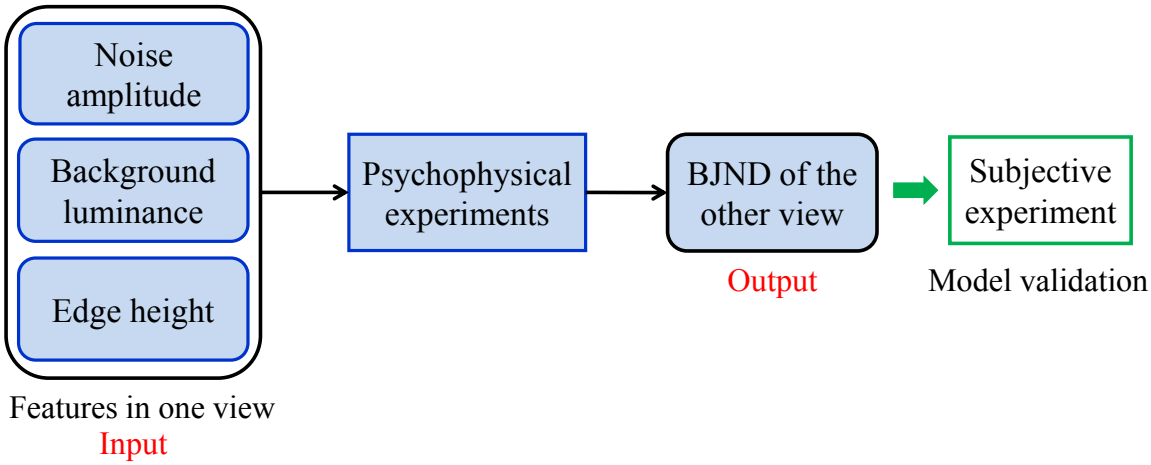


Figure 7.5 – Framework for calculating the BJND of a single view of a stereo pair.

$BJND_{l|r}$ is defined by:

$$\begin{aligned} BJND_{l|r}(i,j) &= BJND_{l|r}(L_{bg_{r|l}}(i,j-d_{l|r}), Eh_{r|l}(i,j-d_{l|r}), n_{r|l}(i,j-d_{l|r})) \\ &= CM_Z(L_{bg_{r|l}}(i,j-d_{l|r}), Eh_{r|l}(i,j-d_{l|r})) \\ &\times \left[1-(\frac{n_{r|l}(i,j-d_{l|r})}{CM_Z(L_{bg_{r|l}}(i,j-d_{l|r}), Eh_{r|l}(i,j-d_{l|r}))})^{\gamma}\right]^{\frac{1}{\gamma}} \end{aligned} \tag{7.31}$$

where $l \mid r$ represents left or right, and d is the horizontal disparity value at pixel (i,j). The disparity values of the left view (d_l) are positive while those of the right view (d_r) are negative. $L_{bg}(i,j)$ indicates the average background luminance at pixel (i,j) that is estimated by averaging the luminance intensity in the 5×5 surrounding region. $Eh(i,j)$ refers to the edge height that is estimated using Eqs. 7.17, 7.18, and 7.19. CM_Z denotes the visibility thresholds of the CM computed by Eq. 7.15. $n(i,j)$ is the luminance difference between the original and distorted images at pixel (i,j), *e.g.,* (noise amplitude). Note that $0 \leq n_{r|l} \leq CM_{Z_{r|l}}$, and BJND of one view $BJND_{l|r}$ can be reduced to $CM_{Z_{r|l}}$ if there is no noise in the other view. The BJND model was validated by means of subjective experiments [20]. The experimental results showed that human perceives the noise when viewing the stereo images if and only if this noise in one view is higher than the BJND value.

7.3.3 JJND model

In addition to LA and CM, the binocular depth cue is proposed to be considered for the design of this 3D-JND model. Since monocular and binocular cells in V1 area have different receptive fields [69], it is reasonable to calculate the JND thresholds for monocular and binocular regions separately. The monocular region in one view refers to 1) the pixels not having corresponding pixels in the other view due to the occlusion effect; or/and 2) disparity-shifted pixels within image borders. Thus the

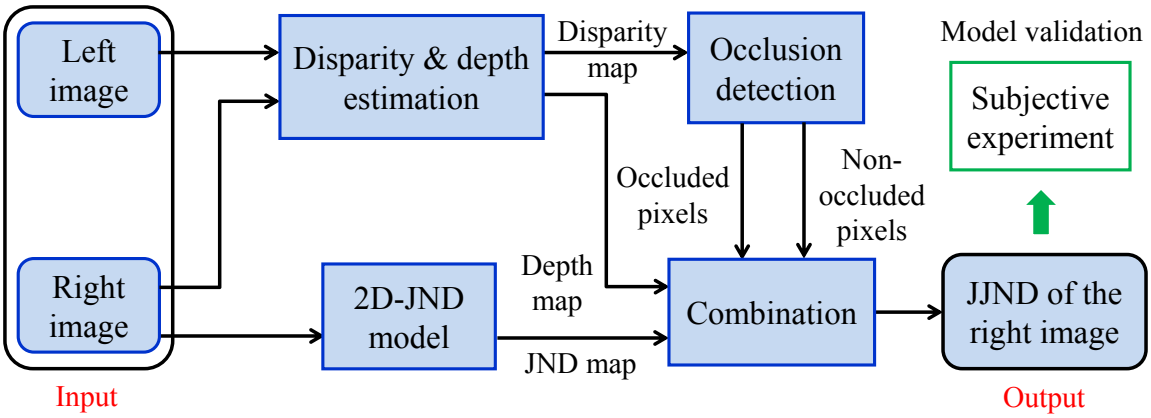


Figure 7.6 – Framework for calculating the JJND of the right view of a stereo pair.

monocular region is known as an occluded region (OR) or non-corresponding region (NCR), and the OR/NCR is only seen by one eye. In contrast, the binocular region in one view is called non-occluded region (NOR) or corresponding region (CR), and the NOR/CR can be perceived by both eyes correspondingly.

Accordingly, Li *et al.* estimated the JND thresholds of both OR and NOR, and thus proposed the joint JND (JJND) model based on the idea that a human subject has different perceptions of objects with different depths [21]. Unlike the JNDD and BJND models, the JJND model was developed with a 2D-JND model, namely non-linear additively masking model (NAMM) [10], which accounts for LA and CM. As shown in Fig. 7.6, the JND thresholds of one image (*e.g.,* left image) are calculated using NAMM. This JND threshold $JND_{Y_l}(i,j)$ of a pixel (i,j) in the left image is defined by:

$$JND_{Y_l}(i,j) = LA_{CY}(i,j) + CM_Y(i,j) - C \times min\left\{LA_{CY}(i,j), CM_Y(i,j)\right\}, \tag{7.32}$$

where C is a constant used to adjust the inter-effect between $LA_{CY}(i,j)$ and $CM_Y(i,j)$. The latter are calculated using Eqs. 7.2 and 7.12, respectively. C is within the $[0,1]$ range, and set to 0.3 in [10].

For the other image (*e.g.,* right image), disparity estimation is firstly performed in order to classify the image pixels into two classes: occluded and non-occluded pixels [70, 71]. The OR, often appearing at the objects' edges or the image borders, represents very strong monocular clues and any distortion in this region is easy to be noticed compared to NOR. Besides, the depth map is derived according to the disparity map and viewing distance. Based on the aforementioned classification, the JJND of the right view is proportional to its 2D-JND thresholds estimated by NAMM, where the coefficients are defined as 1) a fixed value α_{OR} for occluded pixels; and 2) depth-dependent value β_{NOR} for non-occluded pixels (see [21]). The JJND of the right image is formalized as follows:

$$JJND_r(i,j) = \begin{cases} \alpha_{or} \times JND(i,j), & if (i,j) \in OR \\ \beta_{dp} \times JND(i,j), & otherwise \end{cases}, \tag{7.33}$$

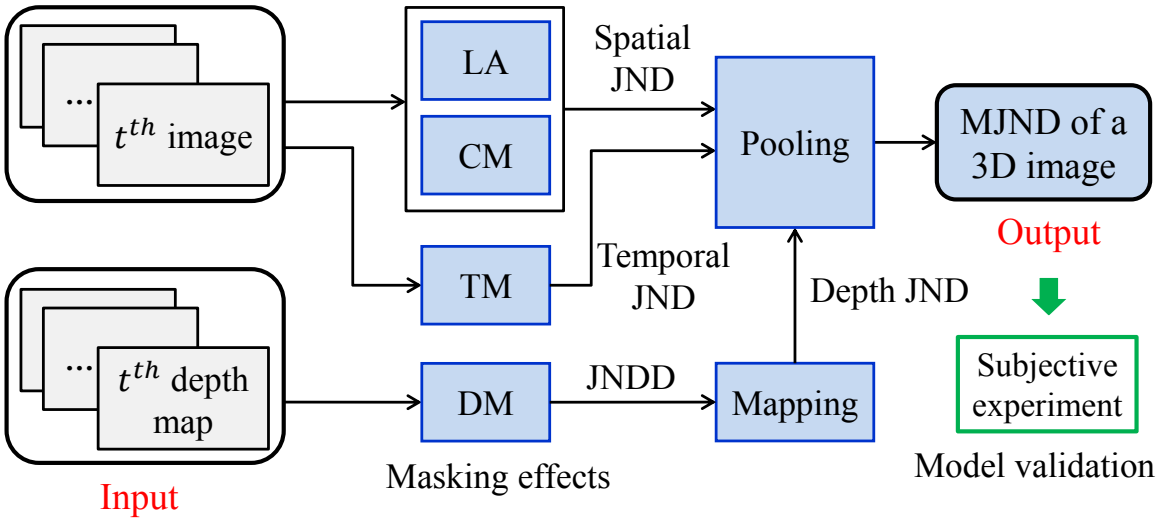


Figure 7.7 – Framework for calculating the MJND of one image from multi-view videos.

where $JND(i,j)$ is the visibility threshold for right image at pixel (i,j). $\alpha_{or} = 0.8$ is used to limit the JND thresholds for OR. The effectiveness of the JJND model was demonstrated using subjective quality evaluations. Specifically, the qualities of the noise-injected S3D images are compared between using JJND and 2D-JND [10]. The experimental results showed that the S3D images receiving JJND noise tolerate more noise than with 2D-JND, in the case of nearly same perceptual quality.

7.3.4 MJND model

Even though the JJND accounts for binocular depth cues, the reliability of this model can be reduced for the stereopairs with low average depth value or uniform depth map. To avoid this constraint, Zhou *et al.* [22] designed a JND in the Multi-view case (MJND) by combining spatial JND (SPJND), TJND and depth JND (DPJND). As shown in Fig. 7.7, the MJND model is defined as:

$$MJND(i,j,t) = [SPJND(i,j,t)]^{w_1} \times [TJND_Z(i,j,t)]^{w_2} \times [DPJND(i,j,t)]^{w_3}, \tag{7.34}$$

where $MJND(i,j,t)$ is the JND threshold at pixel (i,j) at the tth 3D frame (image plus depth map). w_1, w_2 and w_3, are used to control the contribution of SPJND, TJND and DPJND respectively, are set to 1. $SPJND$ denotes the JND thresholds for both LA and CM, and is calculated using a 2D-JND model [5] defined as follows:

$$SPJND(i,j,t) = max\{LA_{CY}(i,j,t), CM_C(i,j,t)\}, \tag{7.35}$$

where $LA_{CY}(i,j,t)$ and $CM_C(i,j,t)$ are estimated based on Eqs. 7.2 and 7.7. In Eq. 7.2, c_1, c_2, and c_3 are set to 14, 3/128, and 2, respectively. c_4 in Eq. 7.7 is set to 1/4. In addition, $TJND_Z(i,j,t)$ is determined using Eq. 7.21. Zhou *et al.* estimated the DPJND thresholds based on the JNDD model

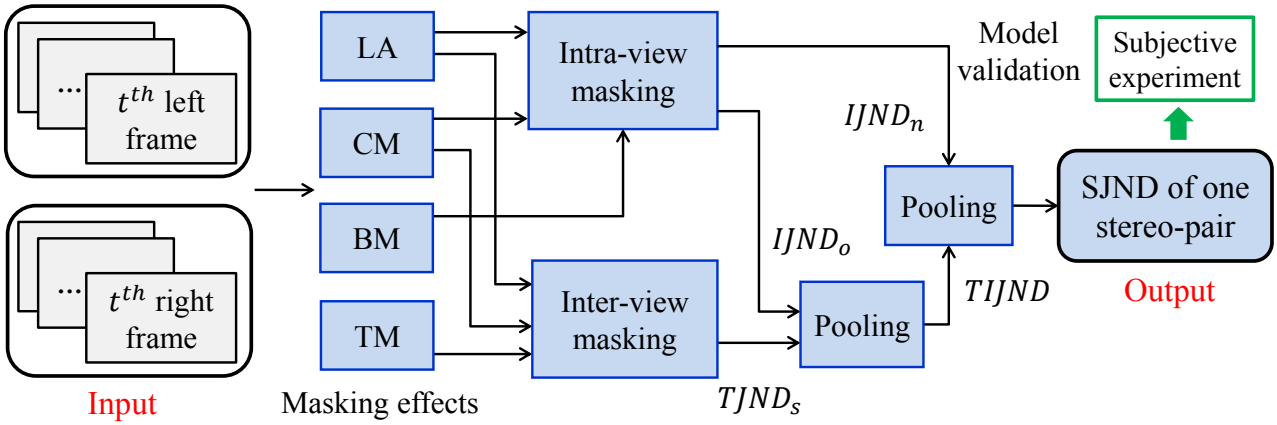


Figure 7.8 – Framework for calculating the SJND of a stereo pair.

proposed in [66]. Thus, the $DPJND(i,j,t)$ is defined by:

$$DPJND(i,j,t) = 1 + \frac{JNDD_{num}(i,j,t)}{256}, \tag{7.36}$$

where $JNDD_{num}$ denotes the numerical JND thresholds computed by Eq. 7.30. The performance of MJND was validated based on subjective experiments. Compared to using the spatial-temporal JND (STJND) [22] or the foveated JND (FJND) [13], the noise-injected 3D video distorted using MJND can tolerate much more noise for the same perceptual quality. Furthermore, the multi-view coding (MVC) [72] using the MJND model achieves better perceptual quality than using the joint multi-view model [73] for the same bit rate.

Inspired by the MJND, Liu *et al.* [74] proposed a new multi-view JNDD (MJNDD) model used to improve the joint multi-view video coding (JMVC). The MJNDD model combines STJND with an adapted JNDD model, which segment the texture frame into background regions (BR) and foreground regions (FR). Recently, Shi *et al.* [75] developed a new 3D-JND model, which considers the depth information and visual saliency in addition to LA, CM, and TM.

7.3.5 SJND model

Qi *et al.* [23, 76] developed the stereo JND (SJND) model for 3D video with the stereo interleaving format [77] (*i.e.,* left and right frames). The SJND model takes into account both intra-view and inter-view MEs in addition to LA and CM. The intra-view masking includes BM, whereas inter-view masking refers to TM.

As shown in Fig. 7.8, for one of the left and right frames, the visibility thresholds for intra-view ME (namely $TJND_L/TJND_R$) are determined by integrating LA, CM, and TM. $TJND_L/TJND_R$ is calculated according to Eq. 7.23. For a pair of stereoscopic frames, the stereo TJND ($TJND_s$) is

computed as follows:

$$TJND_s(i,j,t) = \frac{3}{8} \times [TJND_L(i,j,t)] + \frac{5}{8} \times [TJND_R(i,j,t)], \tag{7.37}$$

The weights for left and right views are used to determine the asymmetry between views [78]. Besides, the views are decomposed into NOR and OR involved respectively in the binocular fusion (BF) [60] and the binocular rivalry (BR) [79]. The human brain can fuse the left and right views into a single mental image when the stimuli in both views are similar. However, if the stimuli are sufficiently different, our brain fails to merge both views resulting in BR phenomena. To model the BM, different intra-view JND (IJND) thresholds are computed based on left and right views according to OR and NOR. The occluded pixels appear on the edge of foreground objects. Therefore, The IJND for non-occluded pixels $IJND_o$ only accounts for CM, and is defined as:

$$IJND_o(i,j,t) = r(t) \times CM_l(i,j,t) + [1 - r(t)] \times CM_r(i,j,t), \tag{7.38}$$

where $r(t)$ is a random value in the range $[0,1]$. Since OR is detected for a random moment, $r(t)$ varies according to time. $CM_{l/r}$ thresholds are calculated based on Eq. 7.7. $IJND_o$ is not based on experiments measuring BM effect, thus $IJND_o$ and $TJND_s$ above should be combined to consider the VM effects of both inter-frame and intra-frame. Accordingly, the new model called $TIJND$ is described as follows:

$$ITJND(i,j,t) = w_t \times TJND(i,j,t) + w_b \times IJND_o(i,j,t), \tag{7.39}$$

where w_t and w_b are the weights used to balance the importance of inter-frame and intra-frame JNDs, respectively. Since BM appears less on OR than NOR, BM effect should be considered less than TM for NOR. Thus w_t and w_b are set to 0.9 and 0.1, respectively. In contrast, LM, CM, and BM are taken into account for NOR. The visibility threshold of the intra-view masking for a non-occluded pixel, namely $IJND_n$ is represented by:

$$IJND_n(i,j,t) = max\left\{f_1(i,j,t), f_2(i,j,t), BM(i,j,t)\right\}, n \in l, r, \tag{7.40}$$

where f_1 and f_2 are calculated based on Eqs. 7.24 and 7.25, respectively. $BM(i,j,t)$ refers to the luminance visibility of one view relative to the other view in the tth frame of the video. By using Eq. 7.20, $BM(i,j,t)$ can be determined. $IJND_n$ of a stereo pair is computed by averaging the $IJND_n$ values for left and right views ($IJND_l$ and $IJND_r$). By integrating $IJND_n$ with $ITJND$, the SJND threshold of a stereo pair is defined as:

$$SJND(i,j,t) = [TIJND(i,j,t)]^{\mu} \times [IJND_n(i,j,t)]^{(1-\mu)}, \tag{7.41}$$

where μ manages the tradeoff between $TIJND$ and $IJND_n$, and is set to 0.5 in [23] or 0.6 in [76]. The effectiveness of SJND for stereoscopic video quality assessment (SVQA) was demonstrated thanks to subjective experiments.

7.3.6 HJND model

It has been demonstrated that depth perception is influenced not only by depth intensity (DI) but also by depth contrast (DC). In light of this, Zhong *et al.* first proposed a 3D image JND model combining 2D-JND with depth saliency taking DI and DC into account [80]. Moreover, the serious geometric distortion (GD) in synthesized views attracts visual attention leading to smaller JND thresholds. Therefore, based on their previous work and a 2D-JND model [10], Zhong *et al.* [24] recently developed a hybrid JND (HJND) model, which considers GD in addition to DI and DC.

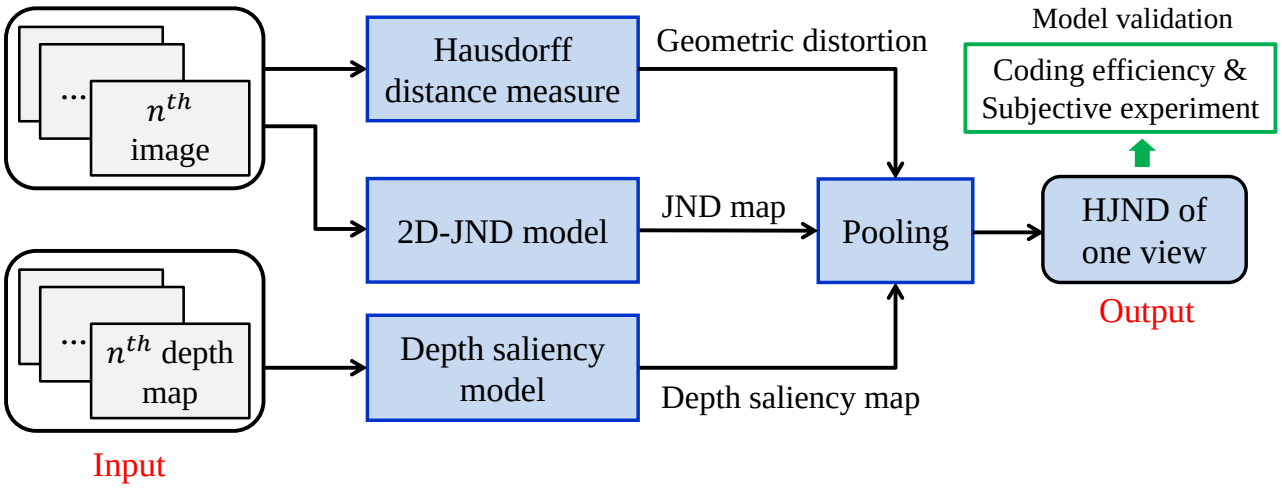


Figure 7.9 – Framework for calculating HJND for a single view.

HVS is more sensitive to closer objects than deeper ones, and the regions with inconsecutive depth or higher DC attract more attention. Thus, depth saliency is influenced by DI, depth intensity contrast and depth orientation contrast. Based on these considerations, a depth saliency model [81] was used to quantify the combined action of DI and DC for 3D video. For the nth and $(n+1)^{th}$ views, the disparity map can be estimated by the stereo matching algorithm. In order to obtain the depth map, the first step is to translate the disparity value $disp$ into depth value dp by:

$$dp(i,j) = \frac{b \times f}{disp(i,j)}, \ disp(i,j) \neq 0, \tag{7.42}$$

where b and f denote the baseline distance between the adjacent cameras and focal length of the camera, respectively. $disp(i,j)$ is the disparity of the pixel at coordinate (i,j). The intersection of two adjacent cameras creates a zero disparity plane, and this zero disparity corresponds to the 3D display.

Deep objects refer to positive disparity, while the pop-out objects have a negative disparity. Next, the depth value is quantized as an 8 bits value, where 0 means the farthest object and 255 denotes the nearest one. Nearer objects are obviously the most salient to the observers. Thereby the depth value $dp(i,j)$ is mapped to the range [0, 255] through the non-linear quantization, defined as follows:

$$dp_m(i,j) = \left\lfloor 255 \times \frac{dp_{min}}{dp(i,j)} \times \frac{dp_{max} - dp(i,j)}{dp_{max} - dp_{min}} + 0.5 \right\rfloor, \tag{7.43}$$

where $\lfloor v \rfloor$ denotes the integer less than or equal to v. dp_{max} and dp_{min} represent the maximum and minimum values of depth, respectively. dp_m is the depth map used to determine depth saliency map S_d. The detail of the S_d estimation is described in [24]. The GD in synthetic views, created by the depth image-based rendering (DIBR) technique [82, 83], is related to the quality of the distorted depth map, and measured by the Hausdorff distance [84]. The latter calculates the geometric distance between the surfaces of the synthesized view and that of the original one. Besides, Yang *et al.* [10] proposed a 2D-JND model expressed by Eq. 7.32. Combining the depth saliency map S_d, GD image G, and the 2D-JND map JND_Y, the HJND threshold of one view can be calculated as follows:

$$HJND(i,j) = \varepsilon \times JND_Y(i,j) \times \omega^{N(S_d(i,j)\cdot G(i,j))}, \tag{7.44}$$

where the parameters ε and ω are empirically set to 1.4 and 0.15, respectively. The symbol $N(\cdot)$ represents a unity-based normalization function that brings all values into the range $[0, 1]$. To validate the effectiveness of the HJND model, it was integrated into the MVC encoding framework to remove the perceptual redundancy. Compared to the standard JMVC scheme and the joint multi-view video plus depth scheme using JJND, the JMVC using HJND can save more bit-budget while providing a better perceptual quality.

7.3.7 DJND model

As described in the HJND model, the HVS is more sensitive to nearby objects than far away objects in the scene. In the real world, the focused areas have higher resolution on the retina while the other areas are blurred by the HVS [85], namely depth of focus (DOF) blur effect [86]. However, conventional 3D displays cannot reproduce the DOF blur effect. In this case, the viewer focuses on the whole scene, which does not correspond to human depth perception. Moreover, this behavior may result in visual fatigue. The described above 3D-JND models have not considered the DOF blur effect. Since FR are more sensitive by HVS than BR, the JND thresholds of FR and BR should be calculated differently. Thereby, Xue *et al.* [25] proposed a disparity-based JND (DJND) model by combining LA, CM with disparity information used to simulate the DOF blur effect.

Fig. 7.10 shows the framework for calculating the DJND of the left view of a stereo pair. First, the visibility thresholds of LA of the left view (LJND) is estimated according to Eq. 7.2. In order to distinguish thresholds for FR and BR, the LJND is filtered by a Gaussian low-pass filter simulating

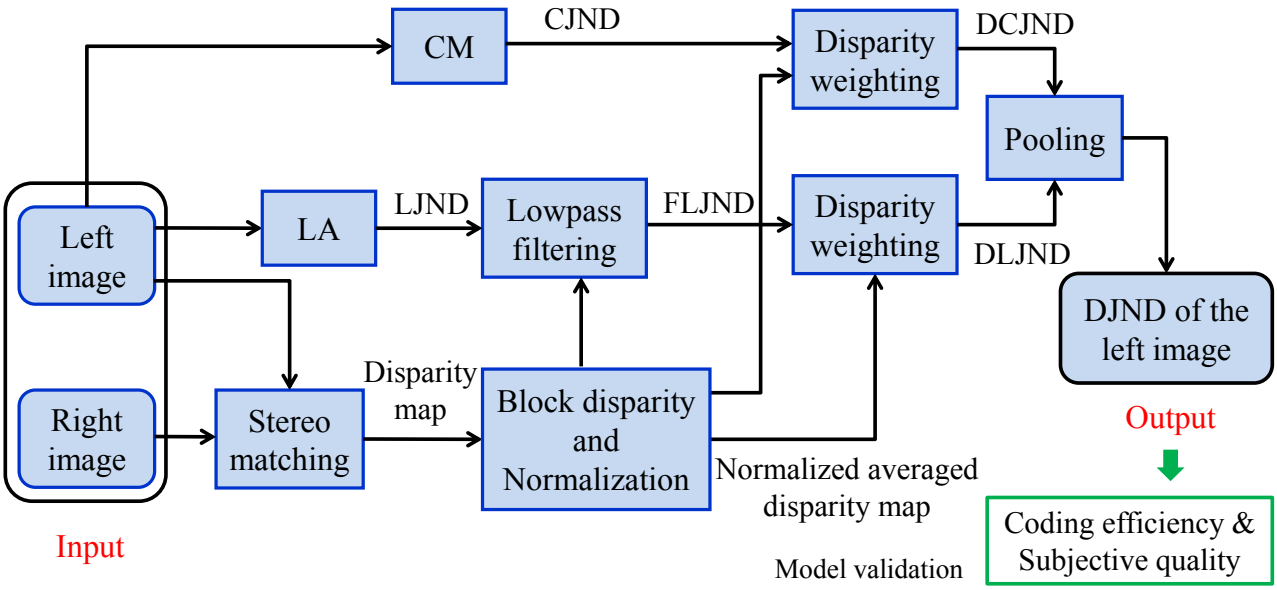


Figure 7.10 – Framework for calculating the DJND of the left views.

the DOF blur effect. The standard deviation of this filter $\sigma(i,j)$ is adaptively calculated based on average disparity values of the disparity image as follows:

$$\sigma(i,j) = \left[\psi + e^{-\alpha \times N(\overline{Disp(i,j)}) - \beta}\right]^2, \tag{7.45}$$

with

$$\overline{Disp(i,j)} = \frac{1}{25}\sum_{x=-2}^{2}\sum_{y=-2}^{2} Disp(i+x, j+y), \tag{7.46}$$

where $N(\overline{Disp(i,j)})$ denotes the normalized average disparity value of a 5×5 block centered at pixel (i,j). The constants α and β are set to 10 and 0.6, respectively. ψ is a constant and defined as 0.117. Then, the filtered LJND is calculated by:

$$FLJND(i,j) = \frac{1}{G}\sum_{x=-2}^{2}\sum_{y=-2}^{2}\left\{\left[e^{-\frac{x^2+y^2}{2\times\varepsilon^2(i,j)}}\right] \times LJND(i+x, j+y)\right\}, \tag{7.47}$$

where G is a bidimensional Gaussian function. Next, the disparity information is used to weight the FLJND by the following negative exponential function:

$$DLJND(i,j) = e^{-2\times\overline{Disp(i,j)}} \times FLJND(i,j) + \delta, \tag{7.48}$$

where δ is a constant and set to 3. The region with a larger disparity (*e.g.,* FR) has lower DLJND thresholds than that with smaller disparity (*e.g.,* BR). Besides, the disparity-based CJND is estimated

by:

$$DCJND(i,j) = e^{-2\times\overline{Disp(i,j)}} \times CJND(i,j), \tag{7.49}$$

$CJND(i,j)$ corresponds to CM_Y that is calculated based on Eq. 7.12. Finally, the DJND of the left view is obtained by combining DLJND with DCJND using NAMM as follows:

$$\begin{aligned} DJND(i,j) = & DLJND(i,j) + DCJND(i,j) \\ & - \phi \times min\left\{DLJND(i,j), DCJND(i,j)\right\}, \end{aligned} \tag{7.50}$$

where ϕ is used to adjust the overlapping effect of LA and CM, and set to 0.3. Similar to HJND model, the DJND is applied to MVC in order to evaluate its performance. The DJND model was validated based on two aspects: 1) the DJND-based MVC outperforms the conventional JMVC in terms of both subjective quality of FR and visual comfort in 3D videos, 2) compared with the conventional JMVC and JMVC using 2D-JND [64], the DJND-based MVC saves more coding bit-budget without the degradation of the perceived quality. This is because MVC using DJND preserves the details in the salient regions and it reduces the redundancies in the other regions.

In addition to the previously mentioned 3D-JND models, Zhang *et al.* [87] proposed a foveated stereoscopic JND model and then applied it to improve the 3D video CE. Moreover, Wang *et al.* [88] developed a 3D just noticeable distortion model for asymmetrical coding. Recently, Du *et al.* [89] studied the effect of texture complexity on the JND threshold for asymmetrically encoding S3D images based on subjective experiments.

7.4 Comparison of 3D-JND models

In this section, we compare the previously described 3D-JND models by analyzing various aspects. The overall comparison between these models is given in Table 7.1. The summary of some important notations and abbreviations used in Table 7.1 is given in Table 7.2.

For each 3D-JND model, the Inputs, the MEs, the 3D content format and the process of model validation have been presented in the previous section. Thereby, we mainly compare in this section these models in terms of their complexity, pros, and cons, as well as applications.

7.4.1 Complexity

To compare the complexity between 3D-JND models, we evaluated not only the MEs and features considered in each model, but also the computational runtime of each model for S3D images. It is worth noting that JJND has been discarded from this evaluation because of its dependency on psychophysical conditions that cannot be controlled here. In order to calculate the runtime of the 3D-JND model, we employed four S3D images with LR images format from the open Middlebury stereo database [90]. This set contains "Teddy" with resolution 450×375 [91], "Art" and "Moebius" with resolution 463×370 [27, 28], and "Baby2" with resolution 1240×1110 [27, 28]. The ground-truth

Table 7.1 – Comparison between the described 3D-JND models.

	JNDD	**BJND**	**JJND**	**MJND**	**SJND**	**HJND**	**DJND**
Inputs	VD, DpM	LCs NAM	LC, DsM, DpM	LCs, DpM, TI	LCs, DsM, TI	LCs, DpM	LC, DsM
VM & Features	DM	LA, CM	LA, CM, DI	LA, CM, TM, DM	LA, CM, TM, BM	LA, CM, DI, DC, GD	LA, CM, DI, DOF
3D format	2D + depth	LR images	LR images	MVD	LR frames	DIBR, MVD	LR images
Model validation	Theoretical results vs. results derived from PE	Noise detection probability in S3D images	Comparison with 2D-JND in terms of SQ	MJND-based MVC vs. JMVM-based MVC in terms of CE and PQ	A metric using SJND vs. SVQA metrics in terms of SQ	MVC with HJND vs. MVC with JJND vs. JMVC in terms of CE and SQ	MVC with DJND vs. MVC with 2D-JND vs. JMVC in terms of CE,PQ and VC
Complexity	_____	***	*	*	**	*****	**
Pros	Extension for various 3D displays	Suitable to several 3D formats	Binocular vision properties	Multiple MEs	Multiple MEs	Considering DC and GD	Several 3D formats, VC improvement
Cons	Limit 3D format, influence of depth image quality	Disparity effect ignoring, SMA impact on JND accuracy	Accuracy decrease for low or uniform disparities, lack of comparison with 3D-JND models	Accuracy decrease for large depth range	Difficult to design the PE for model validation, many parameters in the model	Highly depending on DIBR techniques, specially designed for MVD format	Accuracy decrease for 3D image with small depth difference between FR and BR
Application	Depth sensation enhancement, 3D QoE enhancement, 3D video coding, S3DW	Sharpness /contrast enhancement, 3D video coding, SIQA, S3DW, S3D IR	S3DW	MVC and 3D-HEVC	SVQA	MVC	MVC

Table 7.2 – Important notations and abbreviations used in Table 1.

BM	binocular masking	LR	left and right
BR	background regions	MVC	multi-view video coding
CE	coding efficiency	MVD	multi-view video plus depth
CM	contrast masking	NAM	noise amplitude map
DC	depth contrast	PE	psychophysical experiment
DI	depth intensity	PQ	perceived quality
DIBR	depth image-based rendering	QA	quality assessment
DM	depth masking	SMA	stereo matching algorithm
DpM	depth map	SIQA	stereoscopic image QA
DsM	disparity map	SQ	subjective quality
FR	foreground regions	SVQA	stereoscopic video QA
GD	geometric distortion	S3DW	S3D watermarking
IR	image retargeting	TI	temporal information
JMVC	joint multi-view video coding	TM	temporal masking
JMVM	joint multi-view video model	VC	visual comfort
LA	luminance adaptation	VD	viewing distance
LC	luminance component	VM	visual masking

disparities in this database were used for this evaluation. JND thresholds were estimated using the right view for all 3D-JND models except SJND because of its definition. Besides, the TM effect has not been considered in MJND and SJND for S3D images As both MJND and SJND were designed for S3D videos, it is not fair to compare other 3D-JND models with MJND and SJND including TM effect in terms of computational runtime. The runtime (in second) per image for each 3D-JND model is shown in Table 7.3. Considering runtime, MEs and features (in Table 7.1), the complexity for each 3D-JND model is reported in Table 7.1 using stars. The greater the number of stars is, the higher its complexity is, and vice versa.

Note that the experiments are performed by using MATLAB code on a computer (Inter Core i7-2630 QM Processor at 2.00 GHz, 4GB RAM, Windows 7). As shown in Table 7.3, HJND consumes the longest time among all models due to the process of DIBR and estimation of GD per block. Even though DJND, JJND and MJND accounted for disparity/depth information, DJND is lower than two other models due to the consideration of DOF blur effect. In addition, MJND and SJND use more MEs than, they are faster than BJND. This is due to the fact that MJND and SJND were designed based on a conventional 2D-JND model, and BJND was developed integrating the noise amplitude with LA and CM.

Table 7.3 – Computational runtime (in second) of the described 3D-JND models.

	BJND	**JJND**	**MJND**	**SJND**	**HJND**	**DJND**
Teddy	1.41	0.15	0.04	0.30	9.68	0.60
Art	1.33	0.16	0.04	0.31	10.02	0.60
Moebius	1.37	0.17	0.04	0.29	9.51	0.58
Baby2	11.74	1.09	0.35	2.28	76.05	4.82
Average	**3.96**	**0.39**	**0.12**	**0.80**	**26.32**	**1.65**

7.4.2 Pros and Cons

In this section, we assess the 3D-JND models in terms of their pros and cons. The JNDD has been designed thanks to psychophysical experiments on stereoscopic 3D displays. This model can be extended to various types of S3D displays [66], such as auto-stereoscopic display and passive stereoscopic display [92]. However, it can only measure the visibility threshold with limited depth levels, not satisfying the desired depth range for real application. For instance, JNDD is not suitable for estimating the tolerable depth difference in virtual view rendering [93, 94]. Furthermore, this model is only compatible with the 2D-plus-depth representation of 3D content, and its accuracy depends on the quality of the depth image. Hence, a depth image with poor quality may lead to inaccurate JND thresholds.

Compared to the JNDD, BJND is closer to human binocular perception. Moreover, it can use 2D/color-plus-depth and LR formats. However, this model was designed based on PE using binocular patterns with zero disparity. In other words, BJND ignored the effect of disparity of the visual stimuli on visibility thresholds, which makes it less suitable for real-life stereoscopic images. To avoid this constraint, Kim *et al.* [95] conducted PE to measure the binocular visibility thresholds with different disparities under various amplitudes of the asymmetric noises and background luminance levels. However, they have not studied the impact of the disparity on JND estimation for CM. In addition, the disparity estimation error issue from stereo matching algorithm may decrease the reliability of the BJND estimation. Finally, BJND did not explore the visibility threshold for different types of asymmetric noises (*e.g.,* Gaussian/Poisson noise).

JJND model copes with the issue of disparity ignoring in BJND. This model mimics BF and BR by computing different JND thresholds for OR and NOR, separately. However, the performance of JJND can be reduced for a pair of S3D images having uniform disparity maps or/and low disparity *i.e.,* weak depth perception. Even though it was reported that JJND is more effective than 2D-JND models, the authors did not make any comparison with other 3D-JND models.

MJND and SJND are the most reliable among these 3D-JND models since they take into account both 2D and 3D MEs so that they completely model the stereoscopic HVS characteristics. Since depth values in MJND vary in a very small range, the accuracy of the model may be decreased for S3D images with a larger depth range. For SJND, a subjective validation is difficult because there are

several factors from different MEs. Tunning the parameters is somewhat complicated and may result in very different results, in addition to the necessary adjustment to the used dataset. As described previously for SJND, the NOR leads to BF, whereas OR leads to BR. In fact, BR can occur on NOR when a large inter-difference exists between left view non-occluded pixels and the corresponding pixels in the right view. The relationship between BF and BR should be better explored to model the human binocular vision.

In contrast to JJND, HJND has taken DC into account in addition to DI. Considering GD makes this model more reliable. However, HJND using GD is specifically developed for multi-view video plus depth (MVD) format, and the estimation based on LR views format may not be correct. The accuracy of this model is highly depending on the rendered images obtained using DIBR.

DJND can estimate the visibility thresholds for S3D video with LR or MVD formats. As reported by the authors, using this model in MVC can increase the VC in the S3D display. However, DJND is less effective for S3D images with small depth difference between FR and BR. In other words, this model performs well if FR and BR have large depth difference.

7.4.3 Applications embedding 3D-JND models

In order to improve the compression efficiency of 3D videos, De Silva *et al.* proposed a depth map preprocessing algorithm based on JNDD to remove depth details that are imperceptible by viewers [66]. Similarity, Ding *et al.* [96] recently developed a depth map preprocessing method using JNDD to improve the 3D extension of the high efficiency video coding (HEVC) standard. Bai *et al.* applied JNDD in H.265/HEVC for color image coding by adjusting the quantization parameter (QP) [97]. JNDD has also been employed in depth sensation enhancement [98–100]. The principle is to increase the depth difference between objects such that it exceeds the JNDD. In addition, Lee *et al.* proposed a stereoscopic watermarking method for DIBR using JNDD [101]. More recently, it has been used in visual presence measurement [102] and 3D QoE (*e.g,* VC and depth sensation) enhancement [103].

Over the past few years, the BJND has been applied in several domains. First, Jung *et al.* applied it in sharpness enhancement of S3D images, and the reliability of BJND has been evaluated by considering the accuracy of the stereo matching algorithm. Second, Sdiri *et al.* [104] recently proposed a contrast enhancement method for stereo endoscopic images combining both local image activity and depth information with BJND. The latter was used to control the inter-view enhancement and avoid visual fatigue. Second, BJND was used in 3D video coding or compression. For instance, Fezza *et al.* [105] proposed a non-uniform asymmetric coding method for S3D video based on BJND and depth level. This method employs BJND to measure the minimum distortion in one view that generates 3D perceptual difference, and then uses depth information to adjust the resolution. Meanwhile, Zhu *et al.* [106] developed a fast mode decision approach using BJND to improve the efficiency of MVC. For S3D compression, a new macroblock level rate control method based on BJND model has been proposed in [107]. The visual perception factor measured by BJND was used to adjust the macroblock level bit

allocation. From a different perspective, BJND was used in several works related to stereoscopic image quality assessment [108–113]. The main idea is to use the 2D-JND and BJND to model the visual sensitivity for OR and NOR respectively, and then monocular/binocular visual sensitivity is employed to weight image quality [114]. Besides, Zhou *et al.* [115] proposed a S3D watermarking scheme based on BJND with the aim to guide the watermark embedding. Finally, Shao *et al.* recently carried out a seam carving method for S3D image retargeting combining the 3D visual attention model with BJND [116].

Wang *et al.* developed a S3D watermarking method using JJND [117]. This method validated the authenticity and integrity of stereoscopic images by localizing the tampered regions. MJND model has been used in order to improve the efficiency of 3D-HEVC [75] and MVD video coding [74]. SJND was used for SVQA, whereas HJND and DJND have been applied to improve 3D CE for MVD. To date, there is no application in other domains for these three models since they have been proposed recently.

7.5 Experimental results

n this section, extensive experiments are carried out to compare the performance of the described 3D-JND models. On the one hand, we evaluate the performance using Middlebury stereo database [90] consisting of real-world S3D images. On the other hand, the accuracy estimation of each 3D-JND model for S3D images is compared using psychophysical experiments.

7.5.1 Performance evaluation on the Middlebury stereo database

To compare the efficiency of previously described 3D-JND models, we performed an experimental quantitative evaluation as well as a qualitative demonstration using the Middlebury stereo database. As shown in Fig. 7.11, twenty S3D images from 2005 stereo datasets [27, 28], 2006 stereo datasets [27, 28] and 2014 stereo datasets [29] were chosen for the experimental evaluation. We used the stereo pairs with a full-size resolution from three datasets. The used resolution ranges from 1342×1100 to 1390×1100 in 2005 stereo datasets, 1240×1100 to 1372×1100 in 2006 stereo dataset and 2632×1988 to 2964×2000 in 2014 stereo dataset. These images have been selected based on the number and the "textureness" of the objects in FR and BR.

7.5.1.1 Quantitative evaluation and comparison

Inspired by [15], we propose to evaluate the distortion masking ability as a performance of the 3D-JND models. The distortion tolerance ability (DTA) is estimated in terms of energy of the JND map of one view as follows:

$$DTA = \frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} \left[JND_{3D}(i,j) \right]^2 , \tag{7.51}$$



Figure 7.11 – Right views of the S3D image set from Middlebury databases.

where DTA denotes the JND energy of the 3D-JND map (*i.e.,* JND_{3D}) of the left/right view. H and W are the image height and width respectively. In fact, the DTA value corresponds to the mean square error (MSE) between original and test images with a maximal degradation. To compute DTA, the 3D-JND map of the right image is considered as JND_{3D} for the whole 3D-JND models except SJND. The latter calculates the JND thresholds of the stereo pair [23, 76].

Table 7.4 shows the distortion tolerance ability of different 3D-JND models. It can be observed that

Table 7.4 – Distortion tolerance ability comparison of 3D-JND models. The best result for each image is highlighted in **boldface**, while the second-best result is shown in *italic*.

Image name	BJND	JJND	MJND	SJND	HJND	DJND
Art	8.143	54.323	29.995	*59.179*	**73.609**	35.411
Books	15.363	36.425	19.592	*40.224*	**42.300**	21.467
Dolls	8.704	*59.187*	30.260	58.713	**77.235**	25.969
Laundry	9.972	35.744	18.896	*42.610*	**46.444**	24.755
Moebius	7.758	*50.469*	21.003	44.669	**54.910**	30.571
Reindeer	7.163	*104.748*	44.623	83.879	**116.702**	57.102
Aloe	15.987	32.127	14.755	*34.038*	33.030	**35.944**
Baby2	16.241	32.188	15.620	*32.573*	**38.140**	30.362
Flowers	14.109	*58.438*	25.986	47.643	**70.535**	56.364
Jadeplant	8.969	*86.622*	38.710	70.508	**104.081**	67.885
Bowling2	12.087	25.378	14.061	*32.072*	**36.324**	20.740
Cloth1	13.013	22.342	10.152	**29.494**	*28.968*	19.281
Midd2	13.928	38.469	19.957	*41.160*	**47.906**	34.007
Plastic	13.138	31.493	17.015	*35.990*	**41.717**	19.425
Rocks2	10.142	32.962	15.286	*35.557*	**38.859**	26.620
Wood1	8.073	26.441	11.283	*28.221*	**31.835**	24.291
Motorcycle	11.068	58.599	30.579	*58.742*	**80.058**	37.431
Piano	9.305	*82.391*	39.199	75.196	**102.741**	44.118
Pipes	9.159	*87.985*	44.078	80.267	**115.784**	58.727
Playroom	10.701	*75.550*	38.174	68.923	**100.328**	42.382
Average	11.151	*51.594*	24.961	49.982	**64.075**	35.643

HJND and JJND achieve the best and second-best performance in terms of distortion tolerance ability among all models. This is mainly due to the fact that HJND and JJND thresholds depend highly on disparity/depth values, having a great effect on distortion masking. Higher average disparity value results in higher HJND/JJND energy that corresponds to stronger distortion masking ability. Even though SJND does not take into account disparity/depth information, its masking ability for is close to JJND thanks to the one of consideration of both left and right views. DJND shows lower masking ability than JJND even though both models are developed based on the same 2D-JND model [10]. This could be explained by the DOF blur effect considered in DJND which reduces JND thresholds of FR. MJND takes DM effect into account, where small depth change implies low JND energy. BJND yields the worst performance in terms of distortion masking ability, because the BJND ignores the disparity/depth for 3D visibility thresholds.

Furthermore, we calculated the average disparity level from the ground-truth disparity maps of the Middlebury stereo database, and then revealed the relationship between the average disparity level

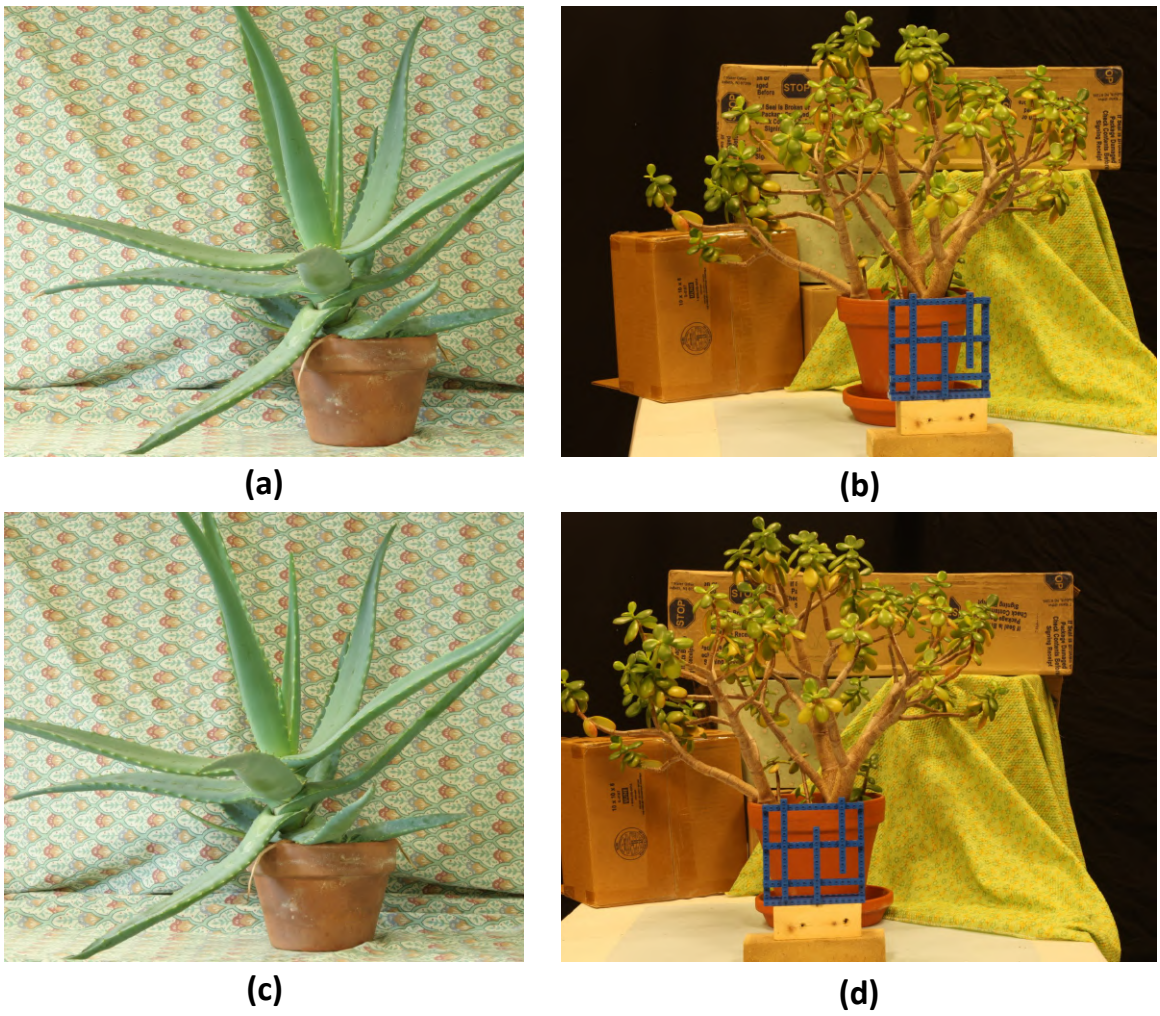


Figure 7.12 – "Aloe" and "Jadeplant" stereo pairs. (a) left view of "Aloe", (b) left view of "Jadeplant", (c) right view of "Aloe", (d) right view of "Jadeplant"

and the JND energy. Higher the average disparity level lead to stronger distortion tolerance ability for same luminance intensity and luminance contrast. For instance, "Aloe" with an average disparity of 72.44 has lower 3D-JND energy than "Jadeplant" with an average disparity of 270.98. It is worth noting that BJND energy of "Aloe" stereo pair is higher than that of "Jadeplant" stereo pair due to the lack of consideration of DM. Fig. 7.12 shows that the horizontal shift/disparity between left and right images in "Jadeplant" is larger compared to "Aloe".

Besides, Fig. 7.13 depicts the plots of the average 3D-JND energies and of the average disparity levels. It can be observed that the average JND energy is approximatively proportional to the average disparity value. More specifically, the visibility threshold of the distortion in S3D image increases as the disparity amplitude increases in the case of similar luminance intensity and luminance contrast. This is consistent with the conclusion drawn in [95].

As the distortion in edge regions is more sensitive to HVS than non-edge regions, the 3D-JND model

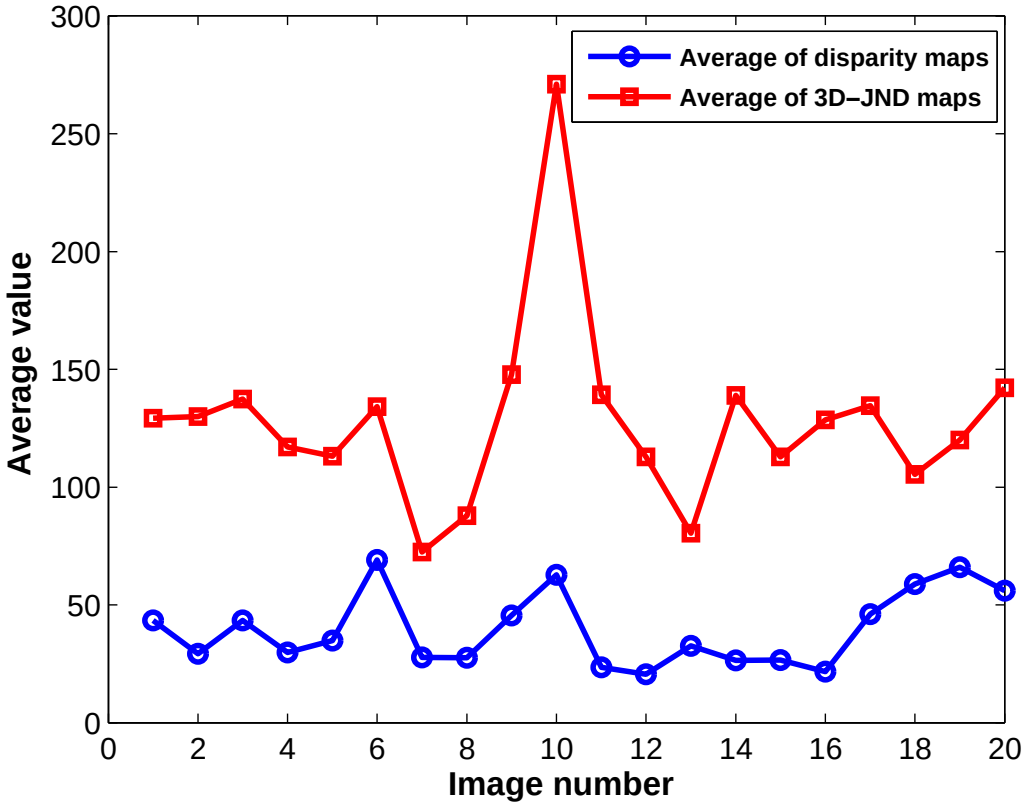


Figure 7.13 – Plots of the average 3D-JND energies and of the disparity values.

yielding high JND thresholds for edge region is efficient for 3D compression. Therefore, we further explore the relative strength of 3D-JND for edge pixels using different 3D-JND models. Firstly, we divide image pixels into two regions: edge regions R_E and non-edge regions R_{NE}. To achieve this, we use the method proposed in [12]. This method can accurately detect edge pixels and deal with the issue of the confusion between textural and edge regions. Then, the 3D-JND map JND_{3D} of the right view is estimated based on the 3D-JND model. Finally, the relative strength of the 3D-JND for edge pixels r_E, the percentage of distortion (*e.g.,* noise) injected to the edge regions, is calculated by:

$$r_E = \frac{\frac{1}{N_E}\sum_{p\in R_E} JND_{3D}(p)}{\frac{1}{N_E}\sum_{p\in R_E} JND_{3D}(p) + \frac{1}{N_{NE}}\sum_{p\in R_{NE}} JND_{3D}(p)}. \tag{7.52}$$

Higher r_E corresponds to higher distortions for edge regions given the same level of 3D-JND thresholds. The r_E values for the different 3D-JND models in the right view are given in Table 7.5. Here we assume that left view is not distorted in order to create the asymmetrically distorted stereopairs. The relative strength of the 3D-JND in the left view can also be calculated in a similar manner. MJND performs best in terms of distortion masking ability for edge pixels. BJND achieves better performance in

Table 7.5 – The relative strength of 3D-JND for edge pixels using different 3D-JND models. The values below are expressed as a percentage, higher value means that the 3D-JND model can mask more noise in edge regions. The best result for each image is highlighted in **boldface**, while the second-best result is shown in *italic*.

Image name	BJND	JJND	MJND	SJND	HJND	DJND
Art	**59.343**	53.491	*56.426*	55.360	53.162	50.386
Books	54.530	52.493	**68.602**	*62.891*	53.350	51.021
Dolls	*56.214*	53.512	**56.535**	54.808	54.075	51.982
Laundry	*58.902*	52.481	**64.035**	58.737	51.692	50.434
Moebius	**59.497**	54.046	*57.658*	55.862	54.590	50.328
Reindeer	**59.778**	50.138	*54.207*	53.940	49.802	50.027
Aloe	52.462	50.449	**57.332**	*54.708*	52.132	49.692
Baby2	49.147	52.627	**65.314**	*60.776*	53.665	53.501
Flowers	**55.662**	45.248	49.940	*51.773*	48.547	47.326
Jadeplant	**56.801**	46.578	53.009	*54.100*	49.051	45.046
Bowling2	43.966	57.414	61.788	*58.174*	**59.840**	53.918
Cloth1	53.397	*56.081*	53.435	52.073	**56.343**	51.415
Midd2	58.133	55.330	**65.161**	*62.359*	57.017	47.398
Plastic	58.511	51.891	**62.762**	*60.678*	49.736	50.392
Rocks2	55.439	55.398	**60.924**	*57.257*	55.938	53.048
Wood1	**60.501**	50.204	*55.684*	53.933	48.340	49.071
Motorcycle	**60.386**	51.409	*57.389*	56.032	51.728	49.328
Piano	**56.209**	45.537	51.902	*52.472*	47.363	47.779
Pipes	**61.785**	42.814	*52.680*	52.157	44.178	42.735
Playroom	**58.743**	46.054	*53.271*	52.826	46.164	46.261
Average	*56.470*	51.160	**57.903**	56.046	51.836	49.554

contrast to JJND and HJND since it estimates higher CM around edge regions than non-edge regions. SJND performs quite similarly to BJND. In fact, MJND and SJND based on 2D-JND estimated by Chou and Li [5] have higher masking ability for edge regions than HJND, JJND, and DJND relying on the 2D-JND model of Yang *et al.* [10]. This is due to the fact that Yang's 2D-JND model estimates lower CM thresholds for edge pixels, whereas Chou's 2D-JND model considers that the CM thresholds for edge and texture regions are the same. In general, by considering the results in both Table 7.4 and 7.5, SJND results in the highest distortion ability among all the 3D-JND models.

7.5.1.2 Qualitative evaluation and comparison

In this section, we provide a qualitative comparison of the six 3D-JND models based on the analysis of the JND profile/map of a stereo pair. The JND thresholds of the right image of a stereo pair were computed using the different 3D-JND models. Dark and bright regions of the JND map indicate

regions having low and high JND visibility thresholds, respectively. For a test stereo pair, we used a method proposed in [70] to detect occluded pixels of the left and right videos. Moreover, the hole in disparity/depth maps was filled by using an efficient algorithm proposed by Jain *et al.* [118] in order to obtain the accurate 3D-JND thresholds. This algorithm can accurately fill the holes of the disparity map based on both color and disparity information of the stereo pair.

The 3D-JND maps of "Art", "Plastic", and "Piano" are shown in Fig. 7.15, 7.17 and 7.19, respectively. For "Art", the four circles with different disparity values indicate different JND thresholds for all 3D-JND models except BJND. The farther the circle is from the observer, the brighter the circle is, and the higher the JND thresholds are. In other words, the farthest circle has the highest distortion masking ability among all circles. This is especially highlighted in the DJND map (see Fig. 7.15(f)), because this model distinguishes the FR from the BR with different visibility thresholds. As shown in Fig. 7.15(f), JND thresholds of the overall FR are lower than those of the overall background ones. One can also notice that the leftmost object of the DJND map is brighter than that of the HJND map even if DJND and HJND are developed on the top of Yang's 2D-JND model [10].

The comparison between Fig. 7.17(e) and Fig. 7.17(f), focusing on the middle yellow object in Fig. 7.16, also demonstrates the decrease of the DJND thresholds in FR. As shown in Fig. 7.15(b) and (e), the Yang's 2D-JND model based JJND and HJND maps are similar. However, the JJND map shows a ghosting effect (*e.g.,* around the sculpture) on the occluded pixels (see Fig. 7.14). This is due to the fact that JJND distinguishes the visibility of the occluded and non-occluded regions in terms of visibility thresholds. The BJND map also shows the ghosting effect on the occluded pixels in Fig. 7.15(a). It should be noted that the black bands of the BJND maps (Fig. 7.15(a), Fig. 7.17(a) and Fig. 7.19(a)) represent "unknown JND thresholds". The width of this band depends on the maximal ground-truth disparity. In addition, the edge around the middle yellow object in Fig. 7.17(b) is darker than the one in Fig. 7.17(e). HJND model not only depends on DI, but also on DC. The latter corresponding to depth variation around the edge of this object (see Fig. 7.16), and attracting more visual attention, results in a decrease of the distortion masking ability. This conclusion can be demonstrated by referring to the variation of the JND thresholds around the edge of the lampshade in Fig. 7.19(e).

For BJND, the fact that all circles have quite similar JND thresholds in Fig. 7.15(a) demonstrates that depth cues have a limited influence on its thresholds. This is indeed consistent with ignoring the binocular disparity in the design of the model. Furthermore, its profiles (Fig. 7.15(a), Fig. 7.17(a) and Fig. 7.19(a)) exhibit higher JND thresholds at edges than textures of certain objects, such as the circles in Fig. 7.15(a) and the music book in Fig. 7.17(a).

In addition, the edges around the sculpture in Fig. 7.15(c) and (d) appear slightly brighter than the ones in Fig. 7.15(b) and (f), and the edges around the middle yellow object in Fig. 7.17(c) and (d) have slightly higher JND thresholds than the ones in Fig. 7.17(b) and (f). This can be explained by the fact that MJND and SJND estimate lower CM thresholds for edge regions than texture regions whereas the same CM thresholds are estimated for edge and texture regions in JJND and DJND. The



Figure 7.14 – "Art" stereo pair. From left to right: left and right views (top), disparity map with holes filling of the right view and occlusion map of the right view. Occluded regions appear in black.

ghosting effect illustrated in Fig. 7.15(d) indicates that the SJND map is estimated using the left and right views.

7.5.2 Performance evaluation based on psychophysical experiments

To evaluate the accuracy of each 3D-JND model, we compare the estimated JND thresholds with the JND thresholds obtained thanks to the psychophysical experiments. We first present the generation of the synthesized 3D images containing textures collected from the ETHZ dataset [30]. Then, we describe the experimental setup used in the subjective measurement of the visibility threshold of the asymmetric distortion. In addition, we further explain how to estimate the visibility thresholds of the synthesized 3D images using previously described 3D-JND models. Finally, we evaluate the 3D-JND models' accuracy by comparing their estimated JND results with the JND data from the psychophysical experiments.



Figure 7.15 – JND profiles of "Art" stereo pair obtained using different 3D-JND models. (a) BJND map, (b) JJND map, (c) MJND map, (d) SJND map, (e) HJND map, (f) DJND map.

7.5.2.1 Selection of the texture images

In order to generate 3D images to be used in psychophysical experiments, we selected the texture images from ETHZ Synthesizability dataset [30] as the patches of the 3D images. The main idea

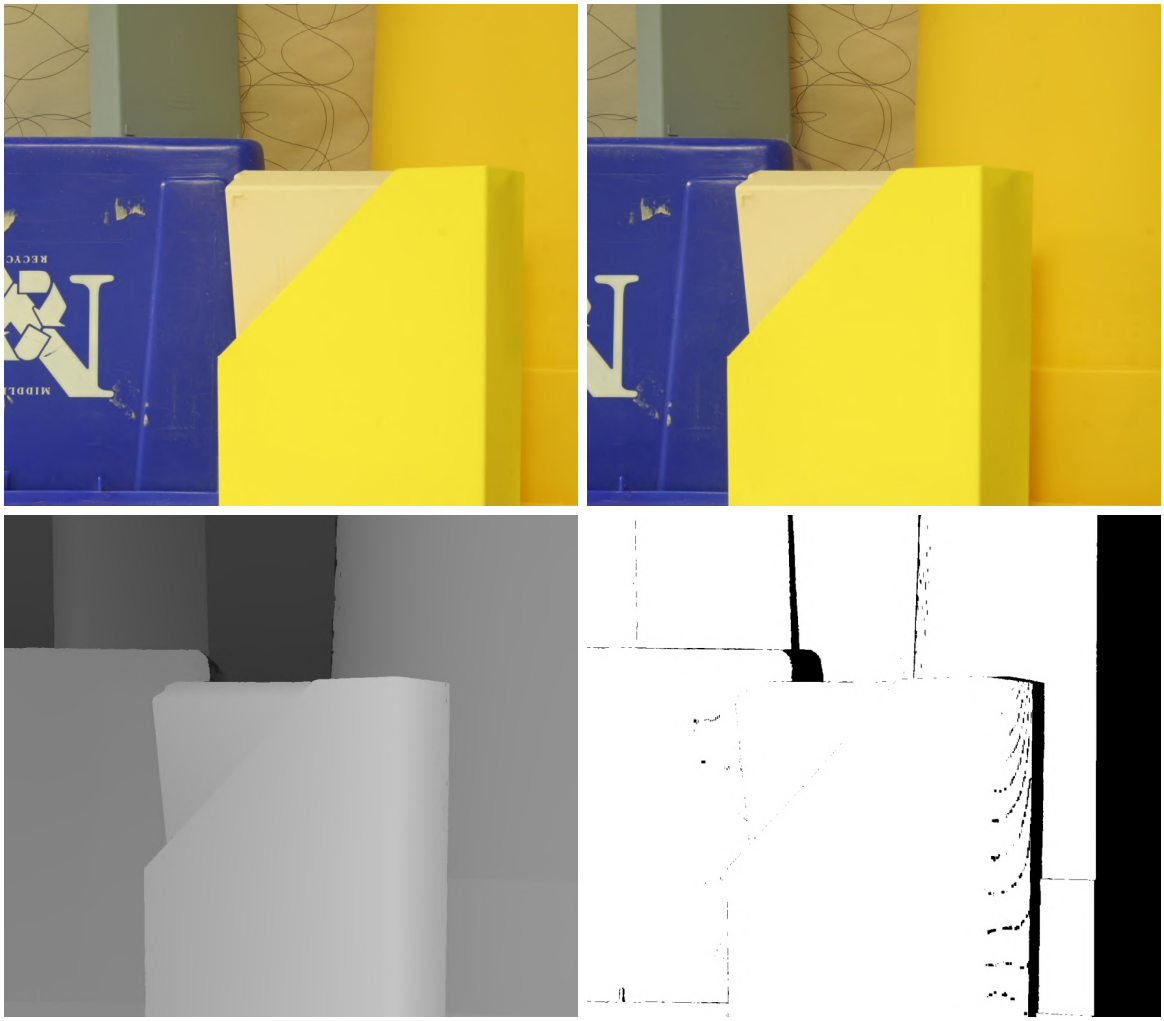


Figure 7.16 – "Plastic" stereo pair. From left to right: left and right views (top), disparity map with holes filling of the right view and occlusion map of the right view. Occluded regions appear in black.

is to synthesize the 3D images consisting of different texture images. In addition, we explore the relationship between 3D-JND thresholds, the "textureness" of the texture image provided in ETHZ dataset [30] and the average of the 2D-JND thresholds. The latter is calculated by averaging the JND values obtained based on the 2D-JND model described in [12]. The "textureness" score indicates the texture strength of the image. As shown in Fig. 7.20, the higher the "textureness" score is, the more the image is textured. Fig. 7.21 shows the textures selection used in our psychophysical experiments. Firstly, the number of texture images is chosen so as to fit in 7 classes according to their "textureness" score and the average of 2D-JND thresholds. Finally, we randomly select one texture image from each dataset. The 7 texture images shown in Fig. 7.20 are further used to synthesize the 3D images. It is worth noting that these texture images were converted to grayscale for the following experiments.

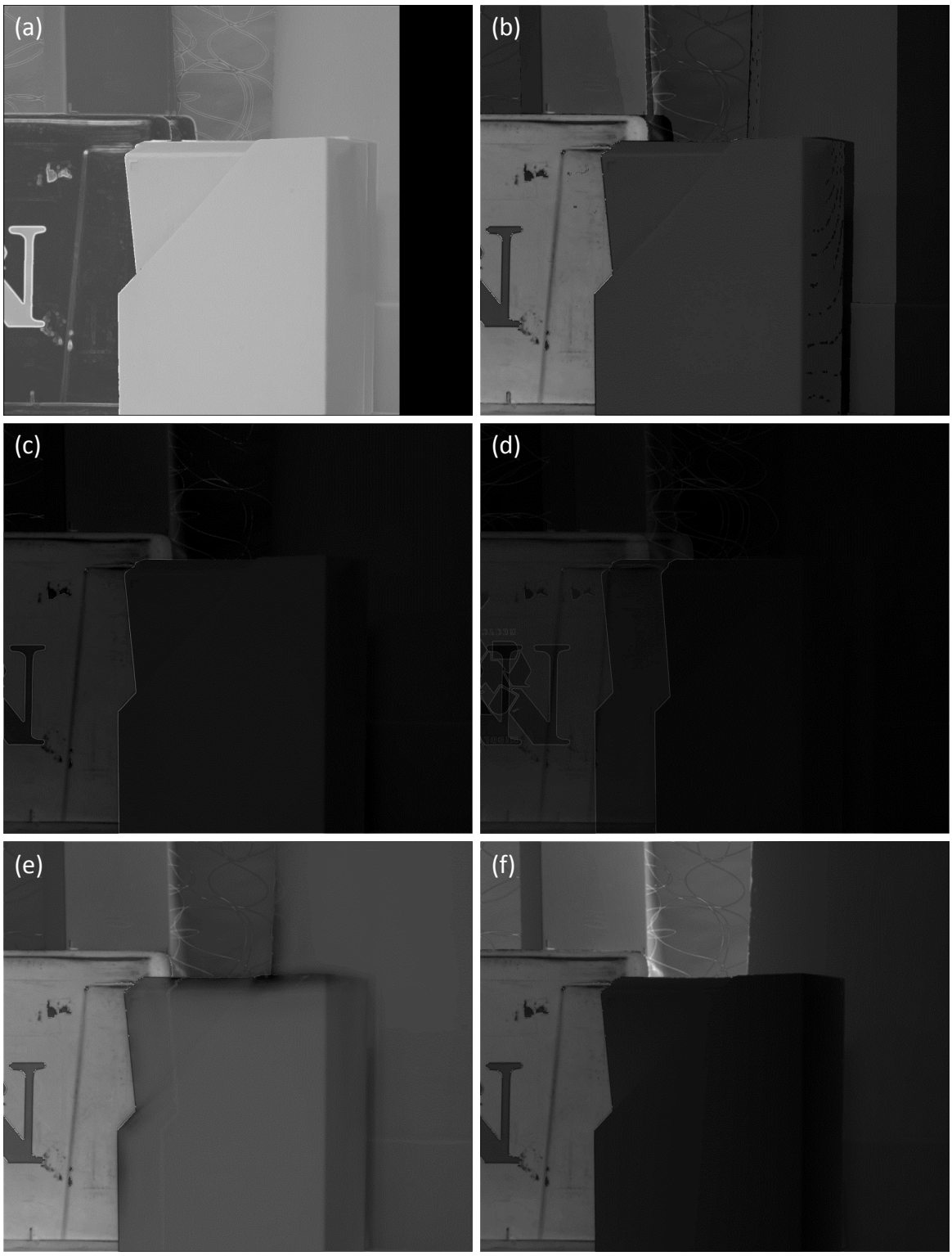


Figure 7.17 – JND profiles of "Plastic" stereo pair obtained using different 3D-JND models. (a) BJND map, (b) JJND map, (c) MJND map, (d) SJND map, (e) HJND map, (f) DJND map.



Figure 7.18 – "Piano" stereo pair. From left to right: left and right views (top), disparity map with holes filling of the right view and occlusion map of the right view. Occluded regions appear in black.

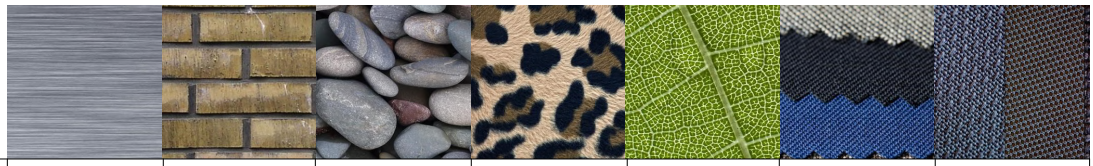


Image number	1	2	3	4	5	6	7
2D-JNDs average	4.45	5.96	6.97	9.84	10.69	12.12	14.04
Textureness	0.070	0.245	0.435	0.478	0.715	0.803	0.974

Figure 7.20 – Seven texture images used to synthesize the S3D stereo pairs. The average of the 2D-JND thresholds and "textureness" values of the texture images are given for each image.

7.5.2.2 Stimuli

To determine the visibility thresholds for the different types of distortion in our psychophysical experiment, we synthesize the asymmetrically distorted stereo pairs based on the 7 pristine texture image shown in Fig. 7.20. Fig. 7.22 illustrates an example of the 3D images presented in the psychophysical experiment. The stimulus consists of reference (left) and distorted (right) stereo pairs. The right image of the stereo pair was altered by four types of distortions, including white Gaussian noise (WN), Gaussian blur (GB), JPEG, and JPEG 2000 (JP2K), respectively. Each distortion type was applied

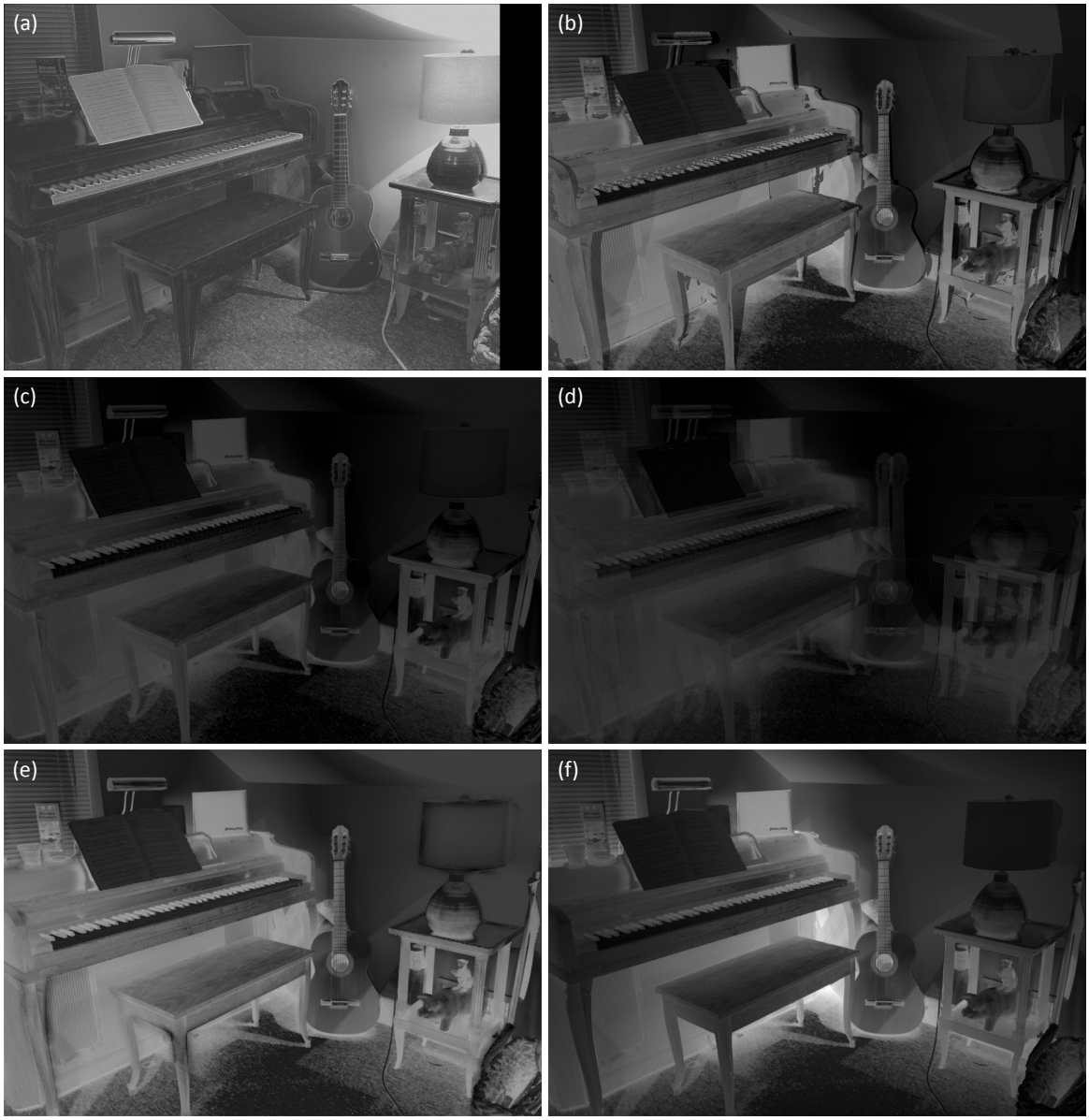


Figure 7.19 – JND profiles of "Piano" stereo pair obtained using different 3D-JND models. (a) BJND map, (b) JJND map, (c) MJND map, (d) SJND map, (e) HJND map, (f) DJND map.

using thirty distortion levels, where the control parameters of these distortions indicated in Table 7.6 were decided to ensure that the subject detects the just noticeable distortion not too early and not too late on the 3D display using stereo glasses. More specifically, the standard deviation σ_{WN} of the WN was used to control the distortion level on the intensity image. The intensity image was filtered using a rotationally symmetric 2D Gaussian kernel of size 7×7 with standard deviation σ_{GB} for GB distortion. The control parameter of the JPEG compression was the quality compression level QC_{JPEG} that determines the amount of information that is lost during compression of the MATLAB functions

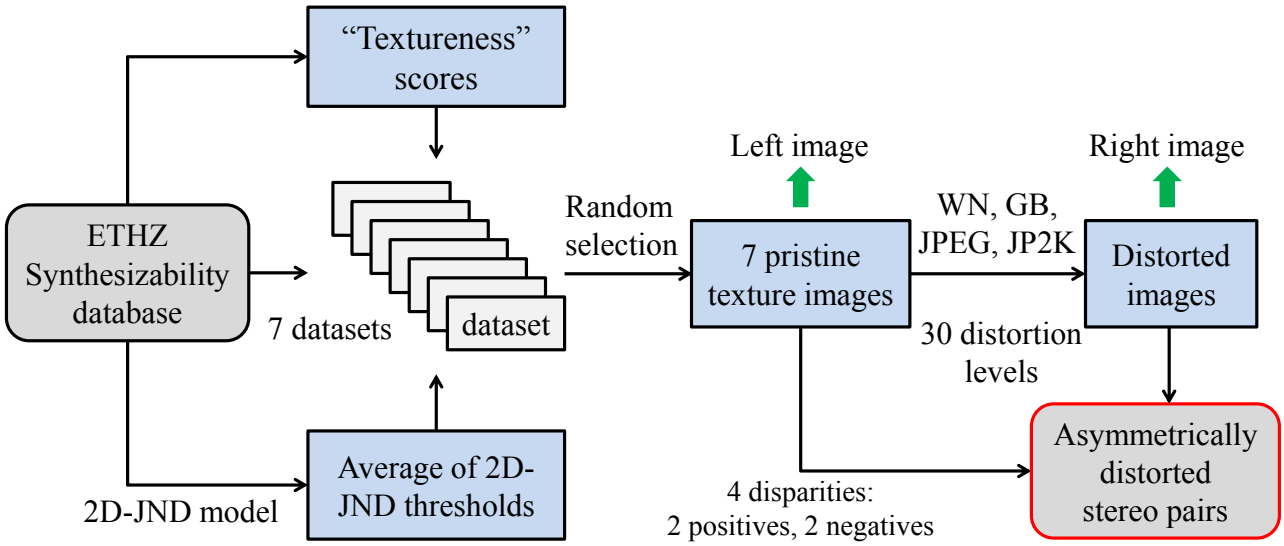


Figure 7.21 – A process of the selection of the textures used in psychophysical experiments.

"im2jpeg.m" and "jpeg2im.m" provided in [119]. Similarly, the JP2K compression was simulated using the MATLAB function "im2jpeg2k.m" and "jpeg2k2im.m" with $n = 5$ and quantization value ranging from 9.9 to 7.

Table 7.6 – Increment step and value ranges of control parameters for distortion simulation on the right texture view of the stereo pair.

Distortion	Control parameter	Increment step	Range
WN	Standard deviation of the Gaussian distribution	0.8	[0.8, 24]
GB	Standard deviation of the Gaussian low-pass filter	0.1	[0.1, 3]
JPEG	Quality compression level	0.3	[0.3, 9]
JP2K	Quantization value	−0.1	[9.9, 7]

Fig. 7.22 shows the visual stimuli presented in our psychophysical experiments. They consist of texture images with a resolution of 300 × 300 corresponding to a visual angle of 2.86° × 2.86° with the experimental condition described in Table 7.7, and a uniform background with an intensity equal to 128. In addition, we created the stimuli with different disparities in order to investigate how the binocular disparity interacts with the detection of the just noticeable distortion. More specifically, the threshold of stereoscopic acuity is approximately 2.3 minutes of arc (arcmin) [120]. In order to easily perceive 3D effects for texture image and to avoid visual fatigue effect during the experiment, we have chosen disparity values well above 2.3 arcmin with both positive and negative parallax: ±26

and ±43 as indicated in Table 7.7. The positive disparity corresponds to "inside" 3D effect and the negative one to "outside" 3D effect. In contrast to the experiment described in [95], the reference stereo pair has been used in our experimental design. This is because we considered not only the noise distortion but also the blurring and compression artifacts. As a result, a 3D image with a resolution of 1920×1080 presented in the subjective test is composed of reference and distorted stereo pairs with the parameters reported on Table 7.7. h and s are set to 390 and 200 pixels, respectively. s is the distance between the reference and the distorted stereo pairs. $s = 200$ ensures that subjects can move its eyes but not the head to detect the stimuli during the experiment. Given the disparity value $Disp$ (in pixel) of a stereo pair, w_2 was defined as $560 - Disp$, and w_1 was set to 560. In sum, a total number of 112 visual stimuli (7 texture images $\times$ 4 distortion types $\times$ 4 disparities) were presented to the subjects during the psychophysical test. For each stimulus, the level of the asymmetric distortion was increased gradually until the subject binocularly just detects the distortion.

7.5.2.3 Subject

The psychophysical experiment was conducted in the XLIM Laboratory at the University of Poitiers. Eighteen subjects, 13 males and 5 females with age ranging from 25 to 35, participated in this experiment. These subjects are composed of 9 naive participants and 9 expert participants who work in the domain of the image processing/computer vision. Each subject undergoes acuity and stereoscopic acuity test. All subjects have the visual acuity around 1.29 with normal or corrected vision, measured by Freiburg Visual Acuity Test (FrACT) with "Landolt C" setting and 1.2 m of viewing distance. Additionally, they have a stereoscopic acuity more than 70 seconds of arc, checked by the RANDOT stereo test. The subjects who used visual correction in daily life were asked to keep it during the experiment.

7.5.2.4 Apparatus

The psychophysical experiment was conducted in a diffuse lighting and noise-isolated room designed especially for subject test. The ambient illumination of the room was set to 65 lux measured by a lux-meter. A 3D display and polarized 3D glasses were used during the experiment. The display is a calibrated Hyundai TriDef S465D with 60 Hz progressive scanning at a resolution of 1920×1080, and a display area of 1.015 meters width and height 0.57 meter. It can work with 2D and S3D modes. The brightness of this display was set to 50% of the maximum. The viewing distance between the subject and the 3D display was set to 1.5 meters, which is recommended by the user's guide of this display and approximatively three times the height of the display. The detail of the experimental setup is given in Table 7.7.

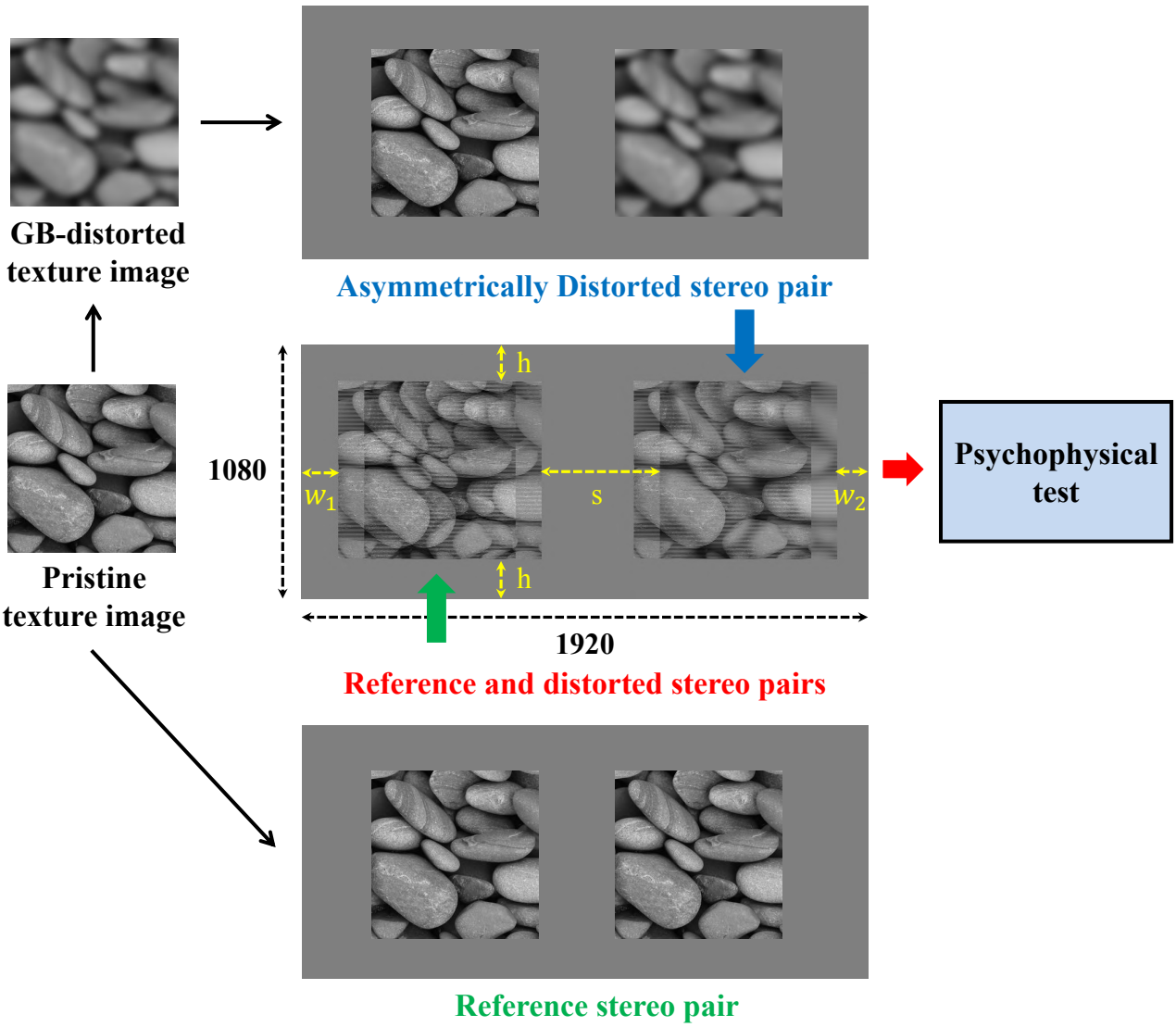


Figure 7.22 – An example illustration of the visual stimuli.

7.5.2.5 Experiment procedure

After the visual screening, the subject was informed about the objective of this experiment, and instructed on how to report the results by using the keyboard. He/She was asked to wear the 3D glasses during the whole experiment. The distorted S3D images with different distortion types and levels were presented to each subject in order to get familiar with the experiment. Once subjects confirmed their understanding of the experiment process, the experiment started.

During the experiment, the subjects compared two 3D images, and checked whether the distortion is just noticeable. The subjects pressed "space" key to continue to increase the distortion level the if the previous level is considered as invisible. Otherwise, the subjects pressed the "enter" key to report the JND result for this stimulus. To provide sufficient time to judge the just noticeable distortion, the

Table 7.7 – Viewing conditions and visual stimulus attributes of the psychophysical test. The positive and negative values correspond to the "inside" and "outside" 3D effects for texture images.

Parameter		**Value**	
Disparity of the texture image	arcmin	±26	±52
	pixel	±43	±86
	degree	±0.44	±0.87
	radian	±0.0076	±0.0151
Screen (m)	width	1.015	
	height	0.57	
Screen resolution (px)	horizontal	1920	
	vertical	1080	
Viewing distance (m)		1.5	
Viewing angle (degree)		37.38	
Background luminance (px)		128	
Ambient illumination (lx)		65	

exposure time of a stimulus is not limited to subjects. Each image of the test sequence related to the stimulus was exposed and followed by a neutral grey image with 128 intensity to avoid visual memory. By pressing the "enter" key, this ended the current sequence and a message was presented to remind the subject to move to the next stimulus. After 56 visual stimuli, the subjects were asked to take a break of 25 minutes to avoid visual fatigue. For each subject, the experiment was stopped immediately when he/she started feeling visual fatigue. The subjects can move their head freely during the test. For each visual stimulus, we assumed that the measures corresponding to the JND threshold ($(DL_{JND})_n$) from all subjects follows a Gaussian distribution. The experimental DL_{JND} value should be within the interval $[0.95DL_{JND}, 1.05DL_{JND}]$, namely confidence interval, where DL_{JND} is the distortion level corresponding to the maximum value of the histogram. To obtain more accurate JND thresholds, $0.95DL_{JND}$ was selected to show the just unnoticeable distortion in the image, whereas $1.05DL_{JND}$ was selected to show the just noticeable distortion in the image. The estimation of the experimental 3D-JND values is described in Fig. 7.23(a).

7.5.2.6 JND maps estimation using psychophysical data

In order to compare the performance of the 3D-JND models, we first estimated totally the 27 JND maps (7 texture images × 4 disparities) for each model based on the reference 3D images. It should be noted that each synthesized 3D image has only one disparity and no occluded regions. Fig. 7.24 shows an example of the JND maps of a synthesized 3D image estimated by different 3D-JND models. In addition, the MJND thresholds shown in Fig. 7.24(c) are higher than JJND thresholds shown in Fig. 7.24(b) around the edges.

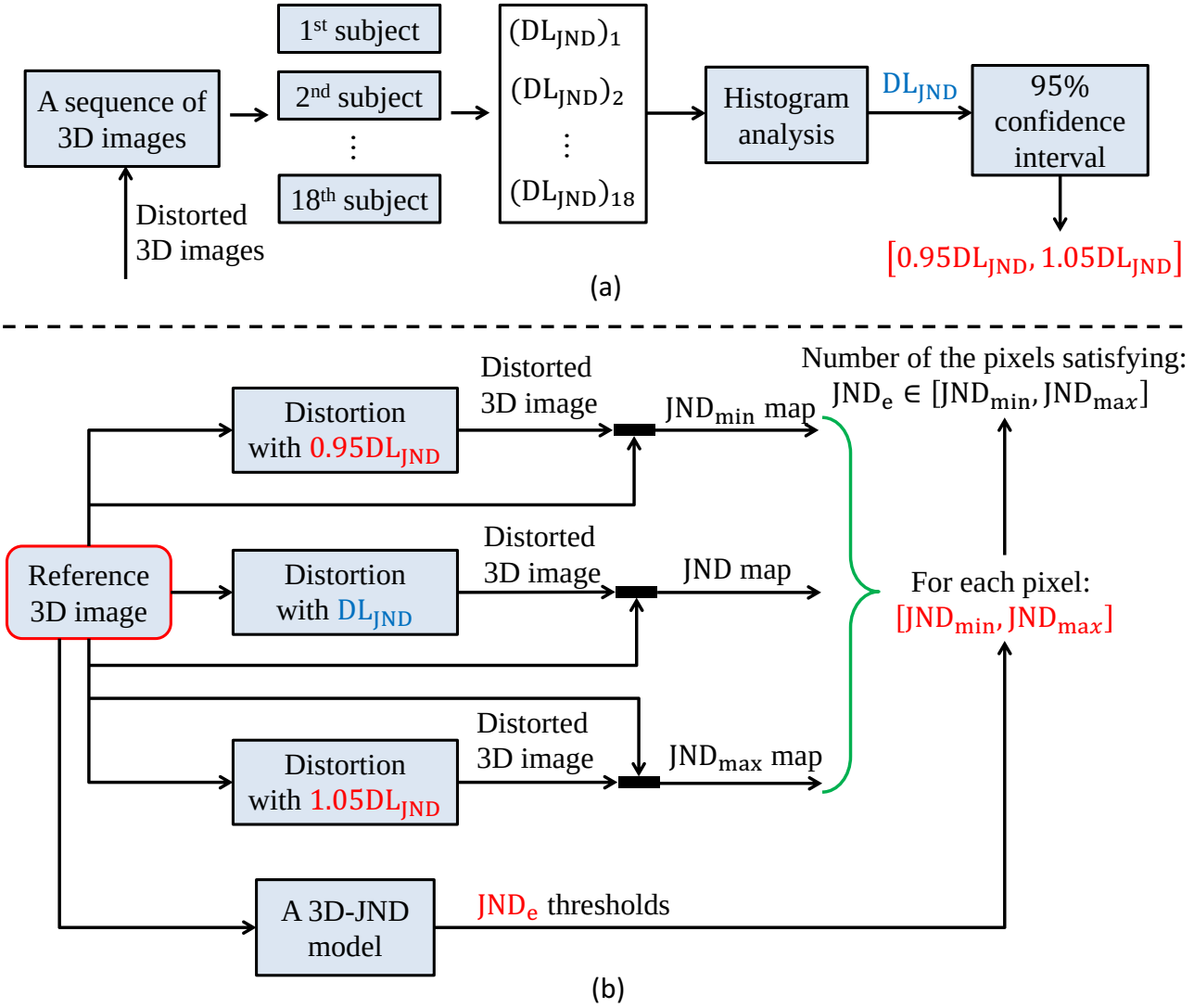


Figure 7.23 – Framework of the accuracy evaluation of the 3D-JND model. (a) Estimation of the 3D-JND interval based on psychophysical results, (b) Accuracy evaluation of the 3D-JND model using the 3D-JND interval.

7.5.2.7 JND maps estimation based on 3D-JND models

As shown in Fig. 7.23(b), we generated three distorted 3D images using $0.95DL_{JND}$, DL_{JND} and $1.05DL_{JND}$. Then, we computed the difference maps between reference and distorted images, and considered these difference maps as the JND maps. Next, we determined a JND interval namely $[JND_{min}, JND_{max}]$ for each pixel of the reference 3D image. Meanwhile, we estimated the JND map using a 3D-JND model. Finally, we checked whether the estimated JND value of each pixel is included in $[JND_{min}, JND_{max}]$, and computed the number of pixels in the JND map included in their corresponding intervals. The greater the number is, the more the 3D-JND model's accuracy is.

Table 7.8 – Psychophysical distortion level corresponding to the 3D-JND thresholds according to different texture images. The maximal value of each column is highlighted in **boldface**, while the minimal value is shown in *italic*.

Texture images number	Distortion level corresponding to 3D-JND threshold					
	WN	**GB**	**JPEG**	**JP2K**	**WN + JPEG**	**GB + JP2K**
1	*3.60*	0.95	*1.95*	*0.75*	*2.78*	*0.85*
2	6.40	0.90	3.00	1.05	4.70	0.95
3	5.20	0.85	3.15	1.50	4.18	1.18
4	8.80	1.05	3.90	1.20	6.35	1.12
5	9.00	*0.70*	**4.95**	0.85	6.98	0.78
6	10.60	**1.30**	3.75	1.75	7.17	**1.53**
7	**11.60**	0.85	3.45	**1.80**	**7.53**	1.33

7.5.2.8 Results and discussion

In this section, we investigate the effects of distortion type and disparity on the experimental visibility threshold of asymmetric distortions. Besides, the comparison of the estimation accuracy between the 3D-JND models is described. Table 7.8 shows the average distortion level corresponding to 3D-JND thresholds for different types of distortion according to different texture images. The "textureness" scores of these texture images are reported in Fig. 7.20. The just noticeable distortion level for WN is proportional to the "textureness" of the texture image. The results for JPEG is similar to those for WN except for images 6 and 7. This is because the image 6, with its low coarseness, can mask more JPEG artifact (*i.e.,* blockiness) than image 7. In contrast to WN, the just noticeable blur (JNB) for GB decreases as the "textureness" score increases. This is due to the fact that images with high texture strength have a low ability to mask the blurring. However, the JNB of image 6 is higher than that of image 5 because of the high contrast of the latter. Accordingly, the blur is easier to be detected in image 5 than in image 6. In general, the just noticeable distortion level is proportional to the "textureness" score for additive distortions as shown in the 6th column of Table 7.8. For subtractive distortion (GB+JP2K), the just noticeable blur level is inversely proportional to the "textureness" score for the texture images from image 3 to 5.

Additionally, we computed the average distortion level corresponding to 3D-JND thresholds according to four disparities in order to explore its effect on asymmetric distortion level threshold. As shown in Table 7.9, the distortion level corresponding to 3D-JND threshold increases as the absolute disparity value increases for all distortion types. In addition, the distortion level thresholds of disparity $+26$ are generally higher than those of disparity -26 for WN, GB, and JPEG. We can draw the same conclusion when comparing the results of disparity *i.e.,* ± 52 for WN, GB, and JP2K. The 3D image with positive disparity ("inside" 3D effect) is farther than that with negative disparity ("outside" 3D effect) from the subject. The larger the distance between the 3D image and the subject is, the less the distortion is visible, thus the higher the JND threshold is. As a result, the asymmetric distortion

Table 7.9 – Psychophysical distortion levels corresponding to the 3D-JND thresholds according to different disparities for each type of distortion.

Disparity (arcmin)	Distortion level corresponding to 3D-JND thresholds			
	WN	**GB**	**JPEG**	**JP2K**
−26	11.60	0.85	3.45	1.75
−52	13.40	1.30	6.00	1.90
+26	13.20	1.15	4.20	1.50
+52	15.60	1.90	6.00	1.95

Table 7.10 – Estimation accuracy (%) comparison of the 3D-JND models according to different distortion types. The larger the value is, the higher the estimation accuracy is. The best result for each distortion type is highlighted in **boldface**, while the second-best result is shown in *italic*.

Distortion type	BJND	JJND	SJND	HJND	DJND	MJND
WN	2.43	**3.19**	3.02	*3.16*	2.91	2.70
GB	**3.28**	3.05	*3.21*	3.07	3.10	3.16
JPEG	*34.77*	30.08	**36.57**	31.45	32.64	32.24
JP2K	*31.33*	29.85	**32.78**	30.40	30.54	31.01
Average	*17.95*	16.62	**18.89**	17.02	17.30	17.28

level threshold for the image with positive disparity is higher in comparison to the one with negative disparity for the same disparity magnitude. In general, the visibility threshold of the asymmetric distortion is proportional to the disparity magnitude under the same background luminance and luminance contrast. This conclusion is in agreement with the observations in [95]. Furthermore, higher depth values in the 3D image may make the asymmetric distortion more tolerable by the HVS.

Based on the psychophysical experiment results mentioned previously, we evaluated and compared the performance of the 3D-JND models in terms of estimation accuracy given in Table 7.10. Generally, SJND performs the best among all 3D-JND models, while BJND ranks second. HJND has the lowest estimation accuracy within all 3D-JND models. For GB, BJND achieves better performance than SJND. Conversely, SJND performs better than BJND for JPEG and JP2K. The estimation accuracies for JPEG and JP2K are generally higher in contrast to WN and GB for all 3D-JND models. This is due to the fact that blockiness is easier to notice by HVS than noise and blur. For WN, HJND and JJND based on Chou's 2D-JND model [5] perform better than SJND and MJND based on Yang's 2D-JND model [10]. For WN, the edge region is more sensitive than the texture region, thus the visibility threshold of the edge region should be lower than that of the texture region. Yang's model estimates lower CM thresholds for edge regions than texture regions whereas the same CM thresholds are estimated for edge and texture regions in Chou's model. Therefore, the 3D-JND models based

Table 7.11 – Estimation accuracy (%) comparison of the 3D-JND models according to four disparities. The larger the value is, the higher the estimation accuracy is. The best result for each distortion type is highlighted in **boldface**, while the second-best result is shown in *italic*.

Disparity (arcmin)	BJND	JJND	SJND	HJND	DJND	MJND
−26	*15.38*	14.63	**16.47**	14.74	15.19	14.99
−52	*15.86*	15.36	**16.68**	15.40	15.14	15.49
+26	*16.93*	15.02	**17.74**	15.69	16.23	16.15
+52	*18.30*	16.81	**19.13**	17.42	17.48	17.49
Average	*16.62*	15.45	**17.51**	15.81	16.01	16.03

Table 7.12 – Estimation accuracy (%) comparison of the 3D-JND models according to different texture images. The larger the value is, higher the estimation accuracy is. The best result for each distortion type is highlighted in **boldface**, while the second-best result is shown in *italic*.

Texture image	BJND	JJND	SJND	HJND	DJND	MJND
1	**15.88**	15.52	*15.77*	15.74	15.18	15.15
2	*18.40*	17.15	**18.97**	17.45	17.55	17.18
3	**17.56**	14.33	*17.52*	14.83	16.10	15.16
4	**22.17**	14.92	*21.80*	15.82	18.60	17.13
5	12.99	12.91	**14.02**	12.87	12.51	*13.41*
6	19.33	19.47	**21.84**	20.14	19.97	*21.05*
7	12.41	*16.18*	15.08	**16.22**	14.72	15.50
Average	*16.96*	15.78	**17.86**	16.15	16.38	16.37

on Yang's model (*i.e.,* HJND and JJND) are more accurate than those based on Chou's model (*i.e.,* SJND and MJND). In contrast, BJND, SJND and MJND perform better than the DJND, HJND, and JJND for GB. For JPEG and JP2K, SJND and BJND achieve higher estimation accuracy than the other models, and JJND performs the worst.

Table 7.11 shows the estimation accuracy comparison of the 3D-JND models according to four disparities. The results in this table demonstrate that SJND and BJND deliver the best and the second-best performance compared to the other models. DJND and MJND are quite similar in terms of estimation accuracy, because both of them account for depth information. JJND generally performs worse than other models. In addition, the comparison between the results for disparity ±52 and disparity ±26 indicates that the larger the disparity magnitude is, the more accurate the 3D-JND models are.

Table 7.12 shows the estimation accuracy of each 3D-JND model according to different texture images. SJND and BJND outperform all the other models, and thus achieve the best and second-best

performance, respectively. DJND is similar to MJND in terms of estimation accuracy. The accuracy of the 3D-JND models for texture image 4 is the highest among 7 texture images. This is mainly due to the fact that highly uniform or textured images (*e.g.,* image 1 or 7) may result in a decrease of the JND estimation accuracy.

It can be noticed that the estimation accuracies of the 3D-JND models for texture 5 and 7 are generally lower compared to the results of other texture images. As shown in Fig. 7.20, the image 5 has a large coarseness whereas image 7 has a large average contrast. Therefore, the detection of the visibility threshold of the asymmetric distortion in these two images based on psychophysical experiments is error-prone. In summary, results in Table 7.10, 7.11 and 7.12 demonstrate that SJND and BJND outperform the other 3D-JND models in terms of estimation accuracy. This is mainly due to the fact that SJND model accounts for various MEs of both monocular and binocular vision, which undoubtedly correspond better to the human quality judgment. BJND achieves slightly lower accuracy than SJND because it ignores the effect of binocular disparity in the development of this model.

7.6 Conclusion

In this paper, we presented a comprehensive review of pixel-based 3D-JND models. The visual characteristics of the HVS considered in these models have been specifically introduced. In addition, these models have been briefly described by giving their rationale and main components in addition to their application, pros, and cons. Besides, we performed an extensive experimental evaluation using Middlebury stereo database with a qualitative demonstration and a performance comparison between these models. Finally, we thoroughly compared the estimation accuracy of the 3D-JND models by using subjective results from our psychophysical experiments. Our study on 3D-JND models allow determining the important characteristics, that will help in the design of a more accurate and efficient 3D-JND model to be used in 3D quality assessment and compression.

Acknowledgment

This work was supported by the Research Council of Norway as part of the project HyPerCept no. 221073 and the region Nouvelle Aquitaine under the doctoral allocation program.

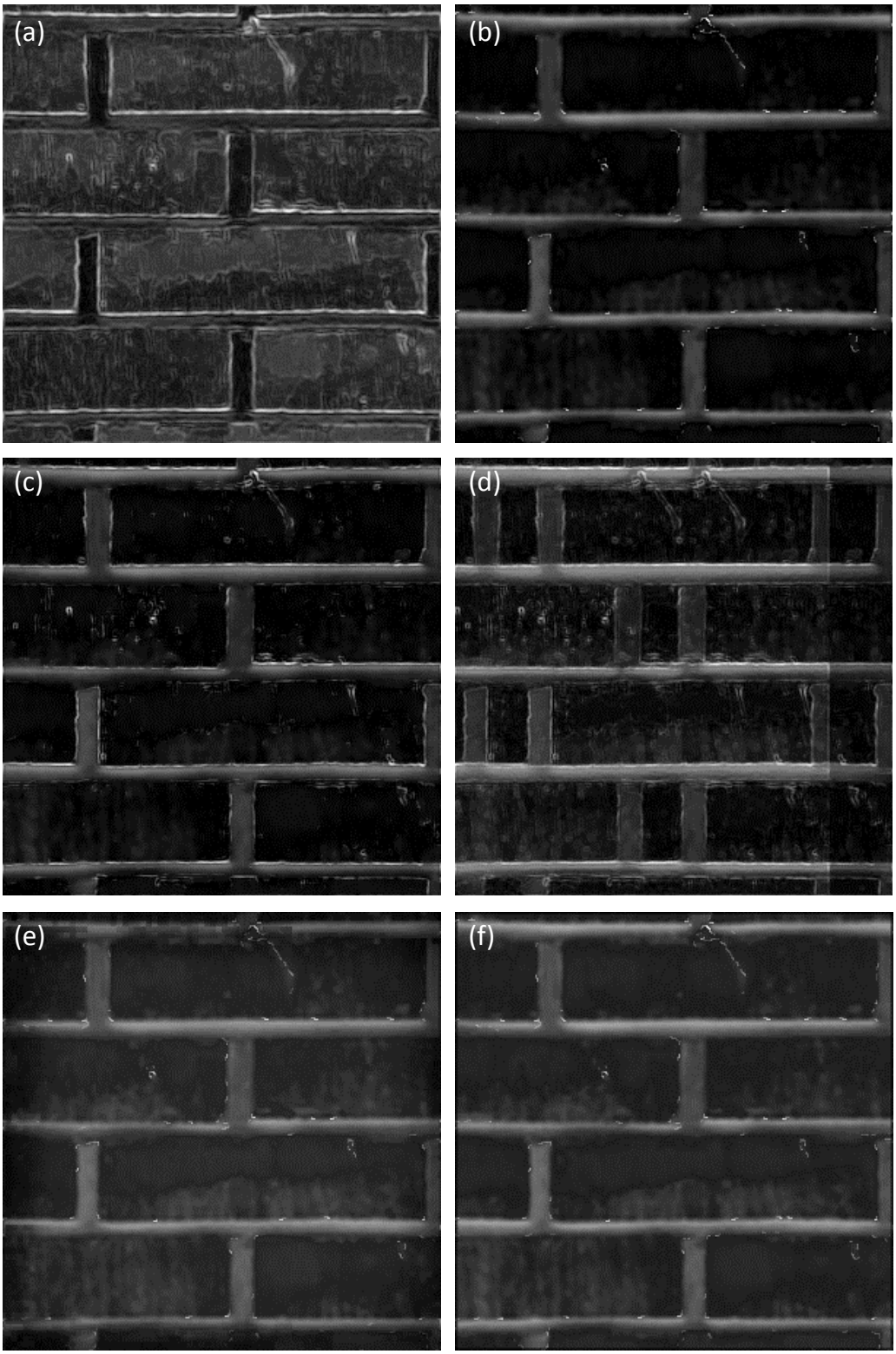


Figure 7.24 – An example illustration of the JND maps of a synthesized 3D image estimated by different 3D-JND models. (a) BJND map, (b) JJND map, (c) MJND map, (d) SJND map, (e) HJND map, (f) DJND map.

Bibliography

[1] Brian A Wandell. *Foundations of vision.* Sinauer Associates, 1995. 106

[2] Melissa K Stern and James H Johnson. Just noticeable difference. *Corsini Encyclopedia of Psychology*, 2010. 106

[3] Nikil Jayant, James Johnston, and Robert Safranek. Signal compression based on models of human perception. *Proceedings of the IEEE*, 81(10):1385–1422, 1993. 106

[4] Hong Ren Wu, Amy R Reibman, Weisi Lin, Fernando Pereira, and Sheila S Hemami. Perceptual visual signal compression and transmission. *Proceedings of the IEEE*, 101(9):2025–2043, 2013. 106

[5] Chun-Hsien Chou and Yun-Chin Li. A perceptually tuned subband image coder based on the measure of just-noticeable-distortion profile. *IEEE Transactions on circuits and systems for video technology*, 5(6):467–476, 1995. 106, 108, 109, 111, 121, 137, 151

[6] Ingo Hontsch and Lina J Karam. Adaptive image coding with perceptual distortion control. *IEEE Transactions on Image Processing*, 11(3):213–222, 2002. 106, 108

[7] XH Zhang, WS Lin, and Ping Xue. Improved estimation for just-noticeable visual distortion. *Signal Processing*, 85(4):795–808, 2005. 106, 108

[8] Xiaohui Zhang, Weisi Lin, and Ping Xue. Just-noticeable difference estimation with pixels in images. *Journal of Visual Communication and Image Representation*, 19(1):30–41, 2008. 106, 108

[9] Zhenyu Wei and King N Ngan. Spatio-temporal just noticeable distortion profile for grey scale image/video in dct domain. *IEEE Transactions on Circuits and Systems for Video Technology*, 19(3):337–346, 2009. 106

[10] XK Yang, WS Ling, ZK Lu, Ee Ping Ong, and SS Yao. Just noticeable distortion model and its applications in video coding. *Signal Processing: Image Communication*, 20(7):662–680, 2005. 106, 109, 112, 120, 121, 124, 125, 134, 137, 138, 151

[11] Yuting Jia, Weisi Lin, and Ashraf A Kassim. Estimating just-noticeable distortion for video. *IEEE Transactions on Circuits and Systems for Video Technology*, 16(7):820–829, 2006. 106

[12] Anmin Liu, Weisi Lin, Manoranjan Paul, Chenwei Deng, and Fan Zhang. Just noticeable difference for images with decomposition model for separating edge and textured regions. *IEEE Transactions on Circuits and Systems for Video Technology*, 20(11):1648–1652, 2010. 106, 112, 136, 141

[13] Zhenzhong Chen and Christine Guillemot. Perceptually-friendly H. 264/AVC video coding based on foveated just-noticeable-distortion model. *IEEE Transactions on Circuits and Systems for Video Technology*, 20(6):806–819, 2010. 106, 122

[14] Jinjian Wu, Guangming Shi, Weisi Lin, Anmin Liu, and Fei Qi. Just noticeable difference estimation for images with free-energy principle. *IEEE Transactions on Multimedia*, 15(7): 1705–1710, 2013. 106

[15] Shiqi Wang, Lin Ma, Yuming Fang, Weisi Lin, Siwei Ma, and Wen Gao. Just noticeable difference estimation for screen content images. *IEEE Transactions on Image Processing*, 25(8):3838–3851, 2016. 106, 132

[16] Jinjian Wu, Guangming Shi, Weisi Lin, and CC Jay Kuo. Enhanced just noticeable difference model with visual regularity consideration. In *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 1581–1585, 2016. 106

[17] Zhenzhong Chen and Hongyi Liu. JND modeling: approaches and applications. In *2014 19th International Conference on Digital Signal Processing*, pages 827–830. IEEE, 2014. 106

[18] Ee-Leng Tan and Woon-Seng Gan. Perceptual image coding with discrete cosine transform. In *SpringerBriefs in Electrical and Computer Engineering*, pages 21–41. Springer, 2015. 106

[19] D Varuna SX De Silva, Warnakulasuriya Anil Chandana Fernando, Stewart T Worrall, SLP Yasakethu, and Ahmet M Kondoz. Just noticeable difference in depth model for stereoscopic 3D displays. In *IEEE International Conference on Multimedia and Expo (ICME)*, pages 1219–1224, 2010. 106, 115, 117, 118

[20] Yin Zhao, Zhenzhong Chen, Ce Zhu, Yap-Peng Tan, and Lu Yu. Binocular just-noticeable-difference model for stereoscopic images. *Signal Processing Letters, IEEE*, 18(1):19–22, 2011. 106, 109, 113, 114, 118, 119

[21] Xiaoming Li, Yue Wang, Debin Zhao, Tingting Jiang, and Nan Zhang. Joint just noticeable difference model based on depth perception for stereoscopic images. In *IEEE International Conference on Visual Communications and Image Processing (VCIP)*, pages 1–4. IEEE, 2011. 106, 120

[22] Lili Zhou, Gang Wu, Yan He, Tao Wang, QingWen Chen, Xiaopeng Fan, and Wen Gao. A new just-noticeable-distortion model combined with the depth information and its application in multi-view video coding. In *Eighth International Conference on Intelligent Information Hiding and Multimedia Signal Processing (IIH-MSP)*, pages 246–251. IEEE, 2012. 106, 114, 121, 122

[23] Feng Qi, Tingting Jiang, Xiaopeng Fan, Siwei Ma, and Debin Zhao. Stereoscopic video quality assessment based on stereo just-noticeable difference model. In *20th IEEE International Conference on Image Processing (ICIP)*, 2013. 106, 114, 115, 122, 124, 133

[24] Rui Zhong, Ruimin Hu, Zhongyuan Wang, and Shizheng Wang. 3d hybrid just noticeable distortion modeling for depth image-based rendering. *Multimedia Tools and Applications*, pages 1–22, 2014. 106, 124, 125

[25] Fei Xue, Cheolkon Jung, and Joongkyu Kim. Disparity-based just-noticeable-difference model for perceptual stereoscopic video coding using depth of focus blur effect. *Displays*, 42:43–50, 2016. 106, 125

[26] Yu Fan, Mohamed-Chaker Larabi, Faouzi Alaya Cheikh, and Christine Fernandez-Maloigne. On the performance of 3D just noticeable difference models. In *IEEE International Conference on Image Processing (ICIP)*, pages 1017–1021, 2016. 106

[27] Daniel Scharstein and Chris Pal. Learning conditional random fields for stereo. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1–8, 2007. 107, 127, 132

[28] Heiko Hirschmuller and Daniel Scharstein. Evaluation of cost functions for stereo matching. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1–8, 2007. 107, 127, 132

[29] Daniel Scharstein, Heiko Hirschmüller, York Kitajima, Greg Krathwohl, Nera Nešić, Xi Wang, and Porter Westling. High-resolution stereo datasets with subpixel-accurate ground truth. In *German Conference on Pattern Recognition*, pages 31–42. Springer, 2014. 107, 132

[30] Dengxin Dai, Hayko Riemenschneider, and Luc Van Gool. The synthesizability of texture examples. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3027–3034, 2014. 107, 139, 140, 141

[31] Robert H Wurtz and Eric R Kandel. Central visual pathways. *Principles of neural science*, 4: 523–545, 2000. 107

[32] Stefan Winkler. Vision models and quality metrics for image processing applications. *These de doctorat, École Polytechnique Fédérale de Lausanne*, 2000. 107

[33] Andrew B Watson, HB Barlow, John G Robson, et al. What does the eye see best? *Nature*, 302(5907):419–422, 1983. 107

[34] Gregory Ward Larson, Holly Rushmeier, and Christine Piatko. A visibility matching tone reproduction operator for high dynamic range scenes. *IEEE Transactions on Visualization and Computer Graphics*, 3(4):291–306, 1997. 107

[35] Lauren Barghout-Stein. *On differences between peripheral and foveal pattern masking.* PhD thesis, University of California, Berkeley, 1999. 107

[36] Otto H. Schade. Optical and photoelectric analog of the eye. *Journal of the Optical Society of America*, 46(9):721–739, Sep 1956. 108

[37] James Mannos and David Sakrison. The effects of a visual fidelity criterion of the encoding of images. *IEEE Transactions on information theory*, 20(4):525–536, 1974. 108

[38] Andrew B Watson. Visual detection of spatial contrast patterns: Evaluation of five simple models. *Optics Express*, 6(1):12–33, 2000. 108

[39] Andrew B Watson and Albert J Ahumada. A standard model for foveal detection of spatial contrast. *Journal of vision*, 5(9):6–6, 2005. 108

[40] Kathy T Mullen. The contrast sensitivity of human colour vision to red-green and blue-yellow chromatic gratings. *The Journal of Physiology*, 359(1):381–400, 1985. 108

[41] Keiko Momose and M Saito. Determination of the chromatic contrast sensitivity using sweep vep technique. In *Engineering in Medicine and Biology, 2002. 24th Annual Conference and the Annual Fall Meeting of the Biomedical Engineering Society EMBS/BMES Conference, 2002. Proceedings of the Second Joint*, volume 3, pages 2145–2146. IEEE, 2002. 108

[42] J. G. Robson. Spatial and temporal contrast-sensitivity functions of the visual system. *Journal of the Optical Society of America*, 56(8):1141–1142, 1966. 108

[43] Eli Peli, Lawrence E Arend, George M Young, and Robert B Goldstein. Contrast sensitivity to patch stimuli: Effects of spatial bandwidth and temporal presentation. *Spatial Vision*, 7(1): 1–14, 1993. 108

[44] Johanna Rousson, Jérémy Haar, Ljiljana Platiša, Bastian Piepers, Tom R Kimpe, and Wilfried Philips. Subjective contrast sensitivity function assessment in stereoscopic viewing of gabor patches. In *SPIE/IS&T Electronic Imaging*, pages 9391 – 9391 – 11. International Society for Optics and Photonics, 2015. 108

[45] Johanna Rousson, Jérémy Haar, Sarah Santal, Asli Kumcu, Ljiljana Platiša, Bastian Piepers, Tom Kimpe, and Wilfried Philips. Contrast sensitivity function in stereoscopic viewing of gabor patches on a medical polarized three-dimensional stereoscopic display. *Journal of Electronic Imaging*, 25(2):25 – 25 – 15, 2016. 108

[46] Marcus J Nadenau, Julien Reichel, and Murat Kunt. Performance comparison of masking models based on a new psychovisual test method with natural scenery stimuli. *Signal processing: Image communication*, 17(10):807–823, 2002. 108

[47] John M Foley. Human luminance pattern-vision mechanisms: masking experiments require a new model. *JOSA A*, 11(6):1710–1719, 1994. 108

[48] FW Campbell and JJ Kulikowski. Orientational selectivity of the human visual system. *The Journal of physiology*, 187(2):437, 1966. 108

[49] Lark Kwon Choi and Alan C Bovik. Flicker sensitive motion tuned video quality assessment. In *IEEE Southwest Symposium on Image Analysis and Interpretation (SSIAI)*, pages 29–32, 2016. 108

[50] Dawid Pajak, Robert Herzog, Radosław Mantiuk, Piotr Didyk, Elmar Eisemann, Karol Myszkowski, and Kari Pulli. Perceptual depth compression for stereo applications. In *Computer Graphics Forum*, volume 33, pages 195–204. Wiley Online Library, 2014. 108

[51] Hugh Davson. *Physiology of the Eye*. Elsevier, 2012. 108

[52] Horace B Barlow. Increment thresholds at low intensities considered as signal/noise discriminations. *The Journal of Physiology*, 136(3):469, 1957. 108

[53] Ralf Engbert, André Longtin, and Reinhold Kliegl. A dynamical model of saccade generation in reading based on spatially distributed lexical processing. *Vision research*, 42(5):621–636, 2002. 109

[54] Jukka Hyönä. Foveal and parafoveal processing during reading. *Oxford handbook of eye movements*, pages 819–838, 2011. 109

[55] Gordon E Legge and John M Foley. Contrast masking in human vision. *Journal of the Optical Society of America*, 70(12):1458–1471, 1980. 111

[56] John Canny. A computational approach to edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, (6):679–698, 1986. 112

[57] Wotao Yin, Donald Goldfarb, and Stanley Osher. Total variation based image cartoon-texture decomposition. Technical report, DTIC Document, 2005. 112

[58] Suzanne P McKee, Mary J Bravo, Harvey S Smallman, and Gordon E Legge. The 'uniqueness constraint'and binocular masking. *Perception*, 24(1):49–65, 1995. 114

[59] Timothy S Meese, Mark A Georgeson, and Daniel H Baker. Interocular masking and summation indicate two stages of divisive contrast gain control. In *Twenty-eighth European Conference on Visual Perception*, 2005. 114

[60] Jeremy M Wolfe. Stereopsis and binocular rivalry. *Psychological review*, 93(3):269, 1986. 114, 123

[61] Guangtao Zhai, Xiaolin Wu, Xiaokang Yang, Weisi Lin, and Wenjun Zhang. A psychovisual quality metric in free-energy principle. *IEEE Transactions on Image Processing*, 21(1):41–52, 2012. 114

[62] Albert J Seyler and Zigmantas L Budrikis. Detail perception after scene changes in television image presentations. *IEEE Transactions on Information Theory*, 11(1):31–43, 1965. 114

[63] Bruno G Breitmeyer and Haluk Ogmen. Recent models and findings in visual backward masking: A comparison, review, and update. *Perception & psychophysics*, 62(8):1572–1595, 2000. 114

[64] Xiaokang Yang, Weisi Lin, Zhongkhang Lu, EePing Ong, and Susu Yao. Motion-compensated residue preprocessing in video coding based on just-noticeable-distortion profile. *IEEE Transactions on Circuits and Systems for Video Technology*, 15(6):742–752, 2005. 114, 127

[65] Chun-Hsien Chou and Chi-Wei Chen. A perceptually optimized 3-D subband codec for video communication over wireless channels. *IEEE Transactions on Circuits and Systems for Video Technology*, 6(2):143–156, 1996. 114

[66] D Varuna SX De Silva, Erhan Ekmekcioglu, Warnakulasuriya Anil Chandana Fernando, and Stewart T Worrall. Display dependent preprocessing of depth maps based on just noticeable depth difference modeling. *IEEE Journal of Selected Topics in Signal Processing*, 5(2):335–351, 2011. 115, 117, 118, 122, 130, 131

[67] Chaminda TER Hewage, Stewart T Worrall, Safak Dogan, Stephane Villette, and Ahmet M Kondoz. Quality evaluation of color plus depth map-based stereoscopic video. *IEEE Journal of Selected Topics in Signal Processing*, 3(2):304–318, 2009. 115

[68] D Varuna SX De Silva, Warnakulasuriya Anil Chandana Fernando, Gokce Nur, Erhan Ekmekcioglu, and Stewart T Worrall. 3d video assessment with just noticeable difference in depth evaluation. In *17th IEEE International Conference on Image Processing (ICIP)*, 2010. 115, 117

[69] Rafik Bensalma and Mohamed-Chaker Larabi. Binocular energy estimation based on properties of the human visual system. *Cognitive Computation*, 5(4):589–609, 2013. 119

[70] Daniel Scharstein and Richard Szeliski. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. *International journal of computer vision*, 47(1-3):7–42, 2002. 120, 138

[71] Jian Sun, Yin Li, Sing Bing Kang, and Heung-Yeung Shum. Symmetric stereo matching for occlusion handling. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2005. 120

[72] Markus Flierl and Bernd Girod. Multiview video compression. *IEEE Signal Processing Magazine*, 24(6):66, 2007. 122

[73] Ying Chen, Miska M Hannuksela, Ling Zhu, Antti Hallapuro, Moncef Gabbouj, and Houqiang Li. Coding techniques in multiview video coding and joint multiview video model. In *Picture Coding Symposium*, pages 1–4. IEEE, 2009. 122

[74] Chao Liu, Ping An, Yifan Zuo, and Zhaoyang Zhang. Applications of just-noticeable depth difference model in joint multiview video plus depth coding. In *SPIE/COS Photonics Asia*, pages 92732R–92732R. International Society for Optics and Photonics, 2014. 122, 132

[75] Yawen Shi, Yongfang Wang, and Yubing Wang. A perceptual preprocess method for 3D-HEVC. In *International Conference on Optical Instruments and Technology 2015*, pages 962219–962219. International Society for Optics and Photonics, 2015. 122, 132

[76] Feng Qi, Debin Zhao, Xiaopeng Fan, and Tingting Jiang. Stereoscopic video quality assessment based on visual attention and just-noticeable difference models. *Signal, Image and Video Processing*, pages 1–8, 2015. 122, 124, 133

[77] Anthony Vetro. Representation and coding formats for stereo and multiview video. In *Intelligent Multimedia Communication: Techniques and Applications*, pages 51–73. Springer, 2010. 122

[78] Payman Aflaki, Miska M Hannuksela, Jukka Häkkinen, Paul Lindroos, and Moncef Gabbouj. Subjective study on compressed asymmetric stereoscopic video. In *17th IEEE International Conference on Image Processing (ICIP)*, 2010. 123

[79] Randolph Blake and Nikos K Logothetis. Visual competition. *Nature Reviews Neuroscience*, 3 (1):13–21, 2002. 123

[80] Rui Zhong, Ruimin Hu, Yi Shi, Zhongyuan Wang, Zhen Han, Lu Liu, and Jinhui Hu. Just noticeable difference for 3D images with depth saliency. In *Pacific-Rim Conference on Multimedia*, pages 414–423. Springer, 2012. 124

[81] Yun Zhang, Gangyi Jiang, Mei Yu, Ken Chen, and Qionghai Dai. Stereoscopic visual attention-based regional bit allocation optimization for multiview video coding. *EURASIP Journal on Advances in Signal Processing*, 2010:60, 2010. 124

[82] Harry Shum and Sing B Kang. Review of image-based rendering techniques. In *Visual Communications and Image Processing 2000*, pages 2–13. International Society for Optics and Photonics, 2000. 125

[83] Christoph Fehn. Depth-image-based rendering (DIBR), compression, and transmission for a new approach on 3D-TV. In *Electronic Imaging 2004*, pages 93–104. International Society for Optics and Photonics, 2004. 125

[84] Nicolas Aspert, Diego Santa Cruz, and Touradj Ebrahimi. Mesh: measuring errors between surfaces using the hausdorff distance. In *ICME (1)*, pages 705–708, 2002. 125

[85] David M Hoffman, Ahna R Girshick, Kurt Akeley, and Martin S Banks. Vergence–accommodation conflicts hinder visual performance and cause visual fatigue. *Journal of vision*, 8(3):33–33, 2008. 125

[86] Yong Ju Jung, Hosik Sohn, Seong-il Lee, Filippo Speranza, and Yong Man Ro. Visual importance-and discomfort region-selective low-pass filtering for reducing visual discomfort in stereoscopic displays. *IEEE Transactions on Circuits and Systems for Video Technology*, 23(8): 1408–1421, 2013. 125

[87] Lei Zhang, Qiang Peng, Qiong-Hua Wang, and Xiao Wu. Stereoscopic perceptual video coding based on just-noticeable-distortion profile. *IEEE Transactions on Broadcasting*, 57(2):572–581, 2011. 127

[88] X Wang, GY Jiang, JM Zhou, Y Zhang, F Shao, ZJ Peng, and M Yu. Visibility threshold of compressed stereoscopic image: effects of asymmetrical coding. *The Imaging Science Journal*, 61(2):172–182, 2013. 127

[89] Baozhen Du, Mei Yu, Gangyi Jiang, Yun Zhang, Feng Shao, Zongju Peng, and Tianzhi Zhu. Novel visibility threshold model for asymmetrically distorted stereoscopic images. In *IEEE International Conference on Visual Communications and Image Processing (VCIP)*, pages 1–4, 2016. 127

[90] Daniel Scharstein and Richard Szeliski. Middlebury stereo vision page. *Online at http://www.middlebury.edu/stereo*, 2002. 127, 132

[91] Daniel Scharstein and Richard Szeliski. High-accuracy stereo depth maps using structured light. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, volume 1, pages I–195, 2003. 127

[92] Varuna De Silva, Anil Fernando, Stewart Worrall, Hemantha Kodikara Arachchi, and Ahmet Kondoz. Sensitivity analysis of the human visual system for depth cues in stereoscopic 3-D displays. *IEEE Transactions on Multimedia*, 13(3):498–506, 2011. 130

[93] Yun Zhang, Sam Kwong, Long Xu, Sudeng Hu, Gangyi Jiang, and C-C Jay Kuo. Regional bit allocation and rate distortion optimization for multiview depth video coding with view synthesis distortion model. *IEEE Transactions on Image Processing*, 22(9):3497–3512, 2013. 130

[94] Zongju Peng, Fen Chen, Gangyi Jiang, Mei Yu, Feng Shao, and Yo-Song Ho. Depth video spatial and temporal correlation enhancement algorithm based on just noticeable rendering distortion model. *Journal of Visual Communication and Image Representation*, 33:309–322, 2015. 130

[95] Hak Gu Kim, Seong-il Lee, and Yong Man Ro. Experimental investigation of the effect of binocular disparity on the visibility threshold of asymmetric noise in stereoscopic viewing. *Optics Express*, 24(17):19607–19615, 2016. 130, 135, 146, 151

[96] Hui Ding, Zhaohui Li, and Dongmei Li. Depth map pre-processing algorithm for compression based on 3D-HEVC scheme. In *16th IEEE International Conference on Communication Technology (ICCT)*, pages 290–294, 2015. 131

[97] Yuejin Bai, Yana Zhang, and Zhaohui Li. 3D Video Coding Using Just Noticeable Depth Difference Based on H. 265/HEVC. In *Computational Intelligence and Security (CIS), 2015 11th International Conference on*, pages 142–145. IEEE, 2015. 131

[98] Seung-Won Jung and Sung-Jea Ko. Depth sensation enhancement using the just noticeable depth difference. *IEEE Transactions on Image Processing*, 21(8):3624–3637, 2012. 131

[99] Seung-Won Jung. A modified model of the just noticeable depth difference and its application to depth sensation enhancement. *IEEE Transactions on Image Processing*, 22(10):3892–3903, 2013. 131

[100] Jianjun Lei, Cuicui Zhang, Yuming Fang, Zhouye Gu, Nam Ling, and Chunping Hou. Depth sensation enhancement for multiple virtual view rendering. *IEEE Transactions on Multimedia*, 17(4):457–469, 2015. 131

[101] Ji-Won Lee, Hee-Dong Kim, Hak-Yeol Choi, Sung-Hee Choi, and Heung-Kyu Lee. Stereoscopic watermarking by horizontal noise mean shifting. In *IS&T/SPIE Electronic Imaging*, pages 830307–830307. International Society for Optics and Photonics, 2012. 131

[102] Heeseok Oh and Sanghoon Lee. Visual Presence: Viewing Geometry Visual Information of UHD S3D Entertainment. *IEEE Transactions on Image Processing*, 25(7):3358–3371, 2016. 131

[103] Feng Shao, Weisi Lin, Zhutuan Li, Gangyi Jiang, and Qionghai Dai. Toward simultaneous visual comfort and depth sensation optimization for stereoscopic 3-D experience. *IEEE transactions on cybernetics*, 47(12):4521–4533, 2017. 131

[104] Bilel Sdiri, Azeddine Beghdadi, Faouzi Alaya Cheikh, Marius Pedersen, and Ole Jakob Elle. An adaptive contrast enhancement method for stereo endoscopic images combining binocular just noticeable difference model and depth information. *Electronic Imaging*, 2016(13):1–7, 2016. 131

[105] Sid Ahmed Fezza, Mohamed-Chaker Larabi, and Kamel Mohamed Faraoun. Asymmetric coding of stereoscopic 3D based on perceptual significance. In *IEEE International Conference on Image Processing (ICIP)*, pages 5656–5660, 2014. 131

[106] Yapei Zhu, Mei Yu, Xin Jin, Zongju Peng, Feng Shao, and Gangyi Jiang. Fast mode decision algorithm for multiview video coding based on binocular just noticeable difference. *Journal of Computers*, 9(10):2428–2434, 2014. 131

[107] Gaofeng Zhu, Mei Yu, Gangyi Jiang, Zongju Peng, Feng Shao, Fen Chen, and Yo-Sung Ho. A novel macroblock level rate control method for stereo video coding. *The Scientific World Journal*, 2014, 2014. 131

[108] Sid Ahmed Fezza, Mohamed-Chaker Larabi, and Kamel Mohamed Faraoun. Stereoscopic image quality metric based on local entropy and binocular just noticeable difference. In *IEEE International Conference on Image Processing (ICIP)*, 2014. 132

[109] Walid Hachicha, Azeddine Beghdadi, and Faouzi Alaya Cheikh. Stereo image quality assessment using a binocular just noticeable difference model. In *IEEE International Conference on Image Processing (ICIP)*, pages 113–117, 2013. 132

[110] Feng Shao, Gang-yi Jiang, Mei Yu, Fucui Li, Zongju Peng, and Randi Fu. Binocular energy response based quality assessment of stereoscopic images. *Digital Signal Processing*, 29:45–53, 2014. 132

[111] Wujie Zhou, Gangyi Jiang, Mei Yu, Feng Shao, and Zongju Peng. PMFS: a perceptual modulated feature similarity metric for stereoscopic image quality assessment. *IEEE Signal Processing Letters*, 21(8):1003–1006, 2014. 132

[112] Yu Cao, Wenhao Hong, and Lu Yu. Full-reference perceptual quality assessment for stereoscopic images based on primary visual processing mechanism. In *IEEE International Conference on Multimedia and Expo (ICME)*, pages 1–6, 2016. 132

[113] Yancong Lin, Jiachen Yang, Lu Wen, Qinggang Meng, Zhihan Lv, and Houbing Song. Quality index for stereoscopic images by jointly evaluating cyclopean amplitude and cyclopean phase. *IEEE Journal of Selected Topics in Signal Processing*, 2016. 132

[114] Xu Wang, Sam Kwong, and Yun Zhang. Considering binocular spatial sensitivity in stereoscopic image quality assessment. In *IEEE International Conference on Visual Communications and Image Processing (VCIP)*, pages 1–4, 2011. 132

[115] Wujie Zhou, Lu Yu, Zhongpeng Wang, Mingwei Wu, Ting Luo, and Lihui Sun. Binocular visual characteristics based fragile watermarking scheme for tamper detection in stereoscopic images. *AEU-International Journal of Electronics and Communications*, 70(1):77–84, 2016. 132

[116] Feng Shao, Wenchong Lin, Weisi Lin, Gangyi Jiang, Mei Yu, and Randi Fu. Stereoscopic visual attention guided seam carving for stereoscopic image retargeting. *Journal of Display Technology*, 12(1):22–30, 2016. 132

[117] Jing Wang, Mei Yu, Ting Luo, Feng Shao, Zongju Peng, and Gangyi Jiang. Joint just noticeable distortion based stereo image watermarking method with self-recovery. *Future Information Engineering (2 Volume Set)*, 49:51, 2014. 132

[118] Ankit K Jain, Lam C Tran, Ramsin Khoshabeh, and Truong Q Nguyen. Efficient stereo-to-multiview synthesis. In *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 889–892, 2011. 138

[119] Rafael C.. Gonzalez, Richard E.. Woods, and Steven L.. Eddins. *Digital Image Processing Using MATLAB®*. McGraw Hill Education, 2010. 145

[120] Ben E Coutant and Gerald Westheimer. Population distribution of stereoscopic ability. *Ophthalmic and Physiological Optics*, 13(1):3–7, 1993. 145

Chapter 8

Paper III: Just Noticeable Difference Model for Asymmetrically Distorted Stereoscopic Images

Y. Fan[1,2], M.-C. Larabi[1], F. A. Cheikh[2], C. Fernandez-Maloigne[1]

[1]XLIM UMR CNRS 7252, University of Poitiers, France
[2]Norwegian Colour and Visual Computing Lab, Norwegian University of Science and Technology, Gjøvik, Norway

Accepted in *IEEE International Conference on Acoustics, Speech and Signal Processing(ICASSP)*, Brighton,UK., 2019.

Abstract

In this paper, we propose a saliency-weighted stereoscopic just noticeable different (SSJND) model constructed based on psychophysical experiments, accounting for binocular disparity and spatial masking effects of the human visual system (HVS). Specifically, a disparity-aware binocular JND model is first developed using psychophysical data, and then is employed to estimate the JND threshold for non-occluded pixel (NOP). In addition, to derive a reliable 3D-JND prediction, we determine the visibility threshold for occluded pixel (OP) by including a robust 2D-JND model. Finally, SSJND thresholds of one view are obtained by weighting the resulting JND for NOP and OP with their visual saliency. Based on subjective experiments, we demonstrate that the proposed model outperforms the other 3D-JND models in terms of perceptual quality at the same noise level.

Index terms– Just noticeable difference, 3D image/video coding, quality assessment, spatial

masking, visual saliency.

8.1 Introduction

The just noticeable difference (JND) is one of the most important perceptual properties, referring to the minimum visibility threshold below which the pixel intensity variations cannot be perceived by the human visual system (HVS). For decades, the 2D-JND models have been successfully developed and exploited in many applications [1]. However, their use for S3D applications is questionable. They mostly rely on monocular vision, which does not fit with the complexity of our 3D perception requiring specific models accounting for both monocular and binocular depth cues.

Accordingly, it becomes crucial to develop effective 3D-JND models for perceptual improvement of 3D applications. So far, a handful of 3D-JND models can be found in the literature [2–12]. Based on the S3D content format, the existing 3D-JND models are classified into two categories: (1) texture-plus-depth-based models, and (2) stereopair-based models.

The first category estimates the visibility thresholds using either texture-plus-depth content [2, 3, 11, 12], or multi-view video plus depth (MVD) one [6–8]. For instance, De Silva *et al.* [2] propose a JND in depth (JNDD) model which measures the threshold for depth variation that a human can perceive on a 3D display. Similarly, to avoid the impact of the monocular depth cues, Yang *et al.* [11] conduct psychophysical experiments (PEs) based on the dynamic Random Dot Stereogram technique to measure the JNDD thresholds. In a different vein, Lian *et al.* design a JND in multi-view (MJND) model, specially for MVD, by combining spatial and temporal JND with JNDD [6]. Likewise, Zhong *et al.* [8] propose a hybrid JND (HJND) model integrating a 2D-JND model [13] together with depth saliency.

In this paper, we propose a saliency-weighted stereoscopic JND (SSJND) model that belongs to the second category, based on our findings obtained from PEs. Our model is two-fold: 1) a disparity-aware binocular JND (DBJND) dedicated to non-occluded pixels (NOPs) obtained from LA and contrast masking (CM) experiments accounting for binocular perception, and 2) a 2D-JND model devoted to occluded pixels (OPs) in the stereo pair. A final step of the proposed SSJND model consists of weighting the JND thresholds by the pixel visual saliency to account for its modulator effect. The obtained model is validated thanks to subjective experiments and compared in terms of perceptual 3D image quality to a number of 3D-JND models from the literature.

8.2 Psychophysical experiments

According to [14], the HVS is able to quickly adjust to the level of the background light in order to distinguish objects. This ability is known as luminance adaptation (LA). Furthermore, contrast masking (CM) describes the masking effect of the HVS in presence of two or more stimuli, if they are of similar contrast/spatial non-uniformity [1]. With the aim to model LA and CM in the S3D context

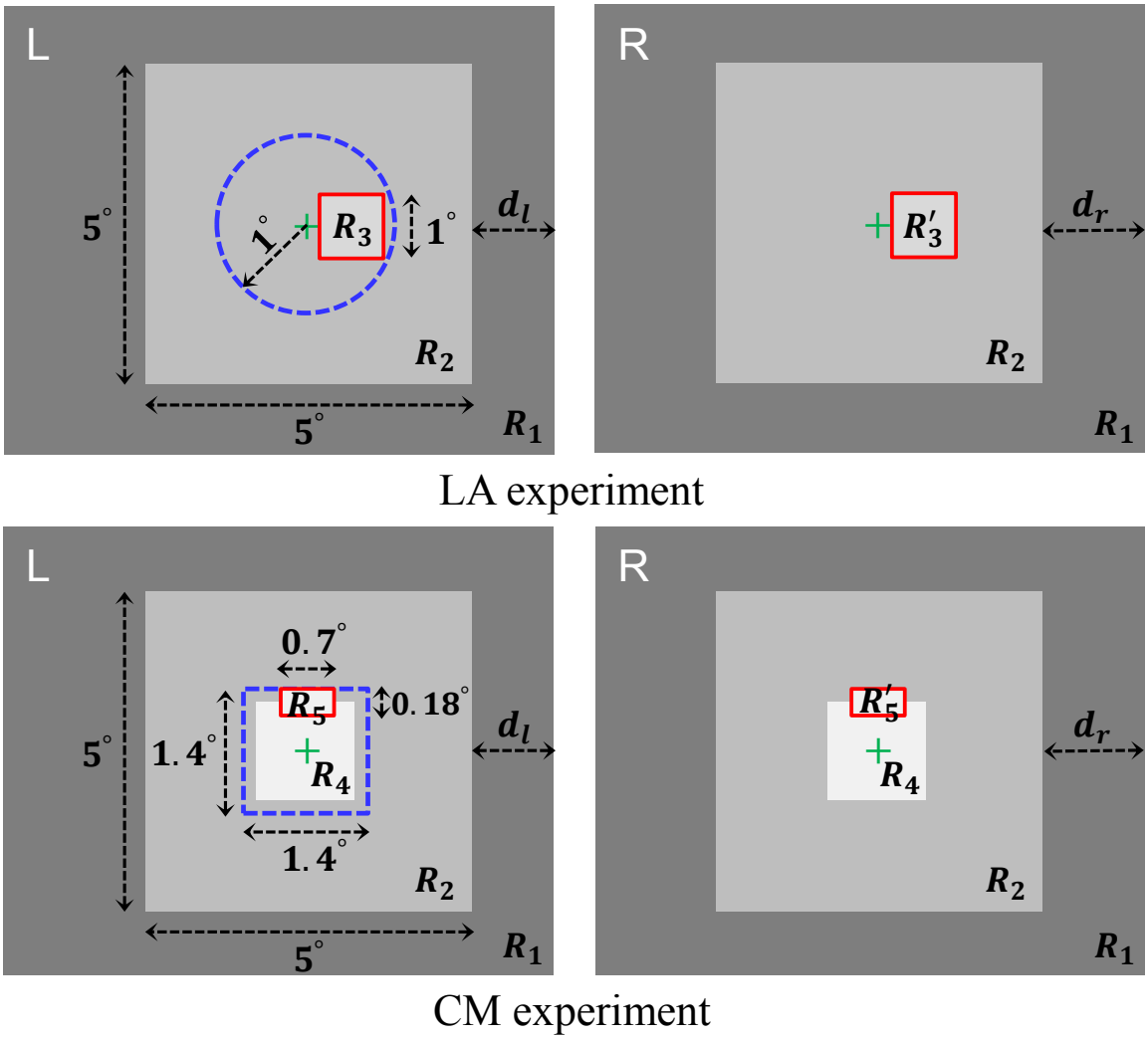


Figure 8.1 – Stereo pair patterns used in psychophysical experiments.

by considering the binocular disparity, we designed two comprehensive PEs.

8.2.1 Stimuli

Fig. 8.1 illustrates the visual stimuli used in LA and CM experiments, respectively. The difference between d_l and d_r denotes the binocular disparity d. The peri-fovea is modeled by a region R_1 with a fixed luminance level 72 pixels (px). The human retinal para-fovea and fovea can cover the information within 5° and 2° of visual angles, respectively, around the fixation point [15]. Consequently, our stimuli in LA/CM experiments contain a fixation cross and a square R_2 with a visual field of $5° \times 5°$ with a luminance level equal to L_b.

8.2.1.1 *LA experiment*

The fovea-covered region is represented by a dashed circle of 2°. In contrast to [4] and [16], the noise area $R_3/R_3^{'}$ is randomly displayed within the dashed circle so as to avoid the memorization of noise location, which may underestimate the JND thresholds. Furthermore, the luminance levels are set to $L_b \pm N_l$ (R_3) and $L_b \pm N_r$ $(R_3^{'})$ with $N_{l|r}$ the noise amplitude injected in the left/right view.

8.2.1.2 *CM experiment*

The fovea-covered region is shown here by a $1.4° \times 1.4°$ dashed square (diagonal of 2°). The noise area $R_5/R_5^{'}$ is located on a randomly chosen side of $R_4/R_4^{'}$ perimeter with an intensity of $N_{l|r}$. Besides, the luminance level of R_4 is set to $L_b - \Delta L$, where ΔL denotes the luminance contrast between R_2 and R_4.

Considering the Percival's zone of comfort [17] and the experiments' duration, we choose five disparity values (i.e., $0°, \pm 0.5°, \pm 1°$) after several trials. Table 8.1 describes the attributes values of the stimuli used in LA and CM experiments. We set $N_l = 0$ for LA experiment to obtain the maximum visibility thresholds of the right image. In total, there are 30 stimuli (6 luminance levels × 5 disparities) in LA experiment, and 60 stimuli (3 luminance levels × 5 disparities × 2 contrast values × 2 noise amplitude levels) in CM experiment.

Table 8.1 – Stimulus attributes for LA and CM experiments.

Attribute	**LA**	**CM**
Noise amplitude N_l (px)	0	0, 2
Luminance contrast ΔL (px)	–	16, 48
Background luminance L_b (px)	22, 32, 48 96, 144, 192	96, 144, 192
Disparity d (degree)	-1, -0.5, 0, 0.5, 1	

8.2.2 Subjects

Twenty-two subjects (ages ranging from 20 to 33) are invited for both LA and CM experiments. Before the experiments, each subject undergoes a visual acuity check based on the Freiburg Vision Test, in addition to the stereoscopic acuity check using the Randot stereo test.

8.2.3 Apparatus

The experiments are conducted in the XLIM psychophysical test room that is isolated from the outside diffuse light and noise. The ambient illumination is adjusted to 65 lux measured by an illuminance-meter. To display the 3D test images, we use a calibrated 46" Hyundai TriDef S465D monitor having HD (1920×1080) resolution with a brightness set to 250 cd/m^2. Polarized 3D glasses are used to

perceive the 3D effect. According to the ITU-R BT.2021-1 recommendations [18], the viewing distance between the subject and the monitor is set to 1.7 m (approx. 3× the height of the display).

8.2.4 Procedure

The experiments are designed using the Psychtoolbox of Matlab [19]. Each subject is informed about the purpose of the experiments, and instructed on how to report the results by using the keyboard thanks to a training sequence before the actual experiments. The JND threshold of the right view is obtained in two steps according to [20]. Step 1 determines the just noticeable noise of the right view A_{JNN}, whereas step 2 measures the just unnoticeable noise A_{JUN}. The noise amplitude of the right view is varying, while the left view remains constant in order to generate an asymmetric noise.

In step 1, for a stimulus, the noise amplitude of the right image N_r is initially set to 0 to make it invisible to subjects. Then, N_r is increased with a step of A_s until it becomes just noticeable, and the final value is saved as the subject's A_{JNN}. A_s was set to 0.0083 and 0.1 for LA and CM experiments, respectively. Subsequently, N_r is increased to $A_{JNN} + A$ immediately to ensure that subjects can easily detect the noise. A is set to 1.7 and 2.0 for LA and CM, respectively.

In step 2, the subjects follow a reversed procedure. Initially, the noise area is visible to subjects. Then, N_r is gradually decreased from $A_{JNN} + A$ by a level of A_s until the noise becomes just unnoticeable. The corresponding value is saved as the subject's A_{JUN}. The JND threshold of the right view is finally obtained as the average of A_{JNN} and A_{JUN}. The procedure is repeated for the whole set of stimuli and subjects are asked to take a rest every 15 minutes.

8.3 Psychophysical data analysis and modeling

8.3.1 Data analysis

To derive a reliable 3D-JND model, we perform an outlier detection [21]. To do so, subject's responses screening is performed following the ITU-R BT 1788 recommendations [22]. The decision criterion is based on the correlation level between subject's values and the mean observations. Consequently, four subjects for LA and three for CM are identified as outliers, and discarded for the further analysis.

With the aim to obtain consistent data for each subject, we proceed to the rejection of outlier observations for each subject [23]. The median-absolute-deviation method is used for LA experimental data, because the distribution for each subject is approximately symmetric. At the opposite, the samples distribution for CM experimental data is mostly asymmetric for which the Tukey's-fences method is preferred. In addition, to confirm the reliability of the JND data after outliers' rejection, we adopt the Jarque-Bera test [24] to verify that all JND values of each stimulus follow a normal distribution ($p - value > 0.05$). Finally, the mean JND threshold is obtained for each stimulus using the post-processed JND data.

To further investigate the effects of background luminance L_b and disparity d on the JND values, we conduct a two-way analysis of variance (ANOVA) with the null hypothesis of no statistical significant difference between JND thresholds for different L_b and d. It is worth noting that the effects of ΔL and N_l are not exploited, because both of them have only two values (see Table 8.1). Before ANOVA, we first validate the normality of the distributions with the Shapiro-Wilk test [25] and the homogeneity of variances with the Levene's test [26].

The resulting $F(1,6) = 290.26$, $p < 0.001$ for LA, and $F(1,3) = 90.01$, $p < 0.001$ for CM demonstrate that there is a significant difference between the luminance levels in terms of JND thresholds. Furthermore, for the binocular disparity, the analysis indicates a significant effect for LA ($F(1,4) = 2.95$, $p = 0.04$) and no effect for CM ($F(1,4) = 0.56$, $p = 0.69$). This is probably caused by the influence of the luminance contrast and the left view noise on JND threshold than by disparity in the complicated CM experiment patterns.

8.3.2 3D-JND modeling

In this section, the post-processed JND data from the conducted experiments are used to derive a 3D-JND model by considering both LA and CM effects, as well as the disparity. Based on the study in [4], the BJND model serves as a framework for our proposed model. Therefore, using L_b, ΔL, N_l and d (*cf.* Table 8.1), we define a disparity-aware binocular JND threshold of the right image DBJND_r as:

$$DBJND_r = T_{r_{max}}(L_b, \Delta L, d)\left[1 - \left(\frac{N_l}{T_{r_{max}}(L_b, \Delta L, d)}\right)^{\lambda}\right]^{\frac{1}{\lambda}}, \tag{8.1}$$

with λ a parameter that controls the influence of N_l, and its estimation will be discussed later. In additions, $T_{r_{max}}$ denotes the maximum JND threshold of the right image by considering both LA and CM effects, and is calculated as follows:

$$T_{r_{max}} = S(L_b)\Delta L + T^{'}_{r_{max}}(L_b, d), \tag{8.2}$$

where $T^{'}_{r_{max}}$ is the LA JND threshold for $N_l = 0$. Fitting the data of Fig. 8.2a requires a curve having two distinct intervals: one for $L_b \leq 48$ and the other for $L_b \geq 48$. L_c represents the intersection point between these two curves, and is equal to 33. As presented in the top of Fig. 8.2a, the values of R-square and the root mean square error (RMSE) indicate a good fitting. Hence, for different L_b and d, $T^{'}_{r_{max}}$ can be expressed as:

$$T^{'}_{r_{max}} = \begin{cases} c_1(L_b^2 + c_2 L_b + c_3 d) + c_4, & L_b \in [0, L_c[\\ c_5(L_b^2 + c_6 L_b + c_7 d) + c_8, & L_b \in]L_c, 255] \end{cases} \tag{8.3}$$

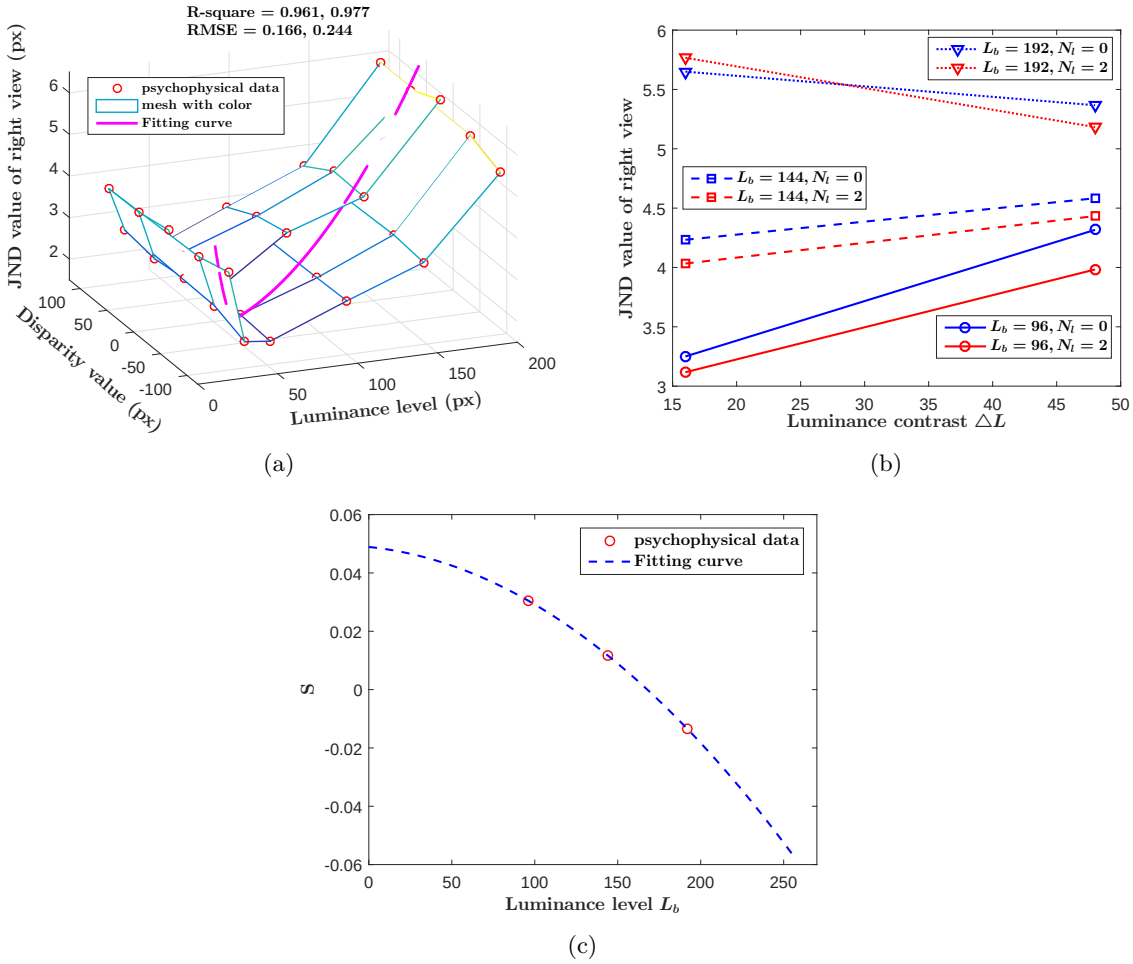


Figure 8.2 – (a) JND thresholds for difference background luminance levels L_b and disparities d from LA experiment, (b) JND thresholds for difference L_b and noise amplitudes of the left view N_l from CM experiment. (c) Average slopes of the two curves in (b) for each L_b.

where the damped least-square fitting method [27] used on LA experimental data allows to identify the different constants as $c_1 = 0.0043$, $c_2 = 83.939$, $c_3 = 0.344$, $c_4 = 9.611$, $c_5 = 0.0001$, $c_6 = 57.884$, $c_7 = 2.333$, and $c_8 = 2.536$.

Moreover, to determine $S(L_b)$ in (8.2), we first depict the average JND values (for five disparity values) according to ΔL under different L_b and N_l in Fig. 8.2b. It illustrates that the JND threshold of the right image increases as the luminance level increases. Furthermore, the JND threshold is inversely proportional to the amplitude of the noise injected in the left image under the same L_b, except for the case where $L_b = 192$. This is because high luminance intensity in CM experiment may result in subjects' misjudgment on the visibility thresholds. The slopes of the two curves for each L_b are determined, and are averaged as S in Eq.8.2. Fig. 8.2c shows the relation between S and L_b based

on the obtained CM data, and its corresponding fitting function is modeled by:

$$S = c_9(L_b^2 + c_{10}L_b) + c_{11}, \tag{8.4}$$

where the fitting parameters c_9, c_{10} and c_{11} are equal to -1.389×10^{-6}, 30.238 and 0.049, respectively. The disparity d in Eq. 8.4 is not considered because of the lack of effect on CM JND values (see Section 8.3.1). As a result, we estimate λ described in Eq. 8.1 by fitting the JND values for $N_l = 0$ and $N_l = 2$, and obtain $\lambda = 3.76$ with RMSE $= 0.421$.

In addition to the above effects, we consider the occlusions for 3D-JND modeling. To this end, image pixels are classified into non-occluded (NOP) and occluded (OP) pixels based on [28]. Then, DBJND (Eq.8.1) is applied to NOP and a robust 2D-JND model [29] is applied to OP. Besides, the studies in [30, 31] demonstrate that JND thresholds are affected by the visual importance of objects in the image, *i.e.*, visual saliency (VS). Specifically, the salient regions, which attract more visual attention, have lower visibility thresholds than the non-salient ones. Thereby, we propose to employ a VS map to weight different JND estimates for NOPs and OPs. The VS of the S3D image is estimated using a promising 3D saliency detection algorithm [32].

Finally, the proposed saliency-weighted stereo JND (SSJND) model is defined as:

$$SSJND_{l|r}(k) = \begin{cases} T_{l|r}(k)(1 + \alpha(T_s - \bar{S}_{l|r}(k))), \bar{S}_{l|r}(k) \in [0, T_s] \\ T_{l|r}(k)(1 - \alpha(\bar{S}_{l|r}(k) - T_s)), \bar{S}_{l|r}(k) \in]T_s, 1] \end{cases} \tag{8.5}$$

where $l \mid r$ refer to the left or right image, k is the kth pixel of the image. $T_{l|r}$ respectively corresponds to $\text{DBJND}_{l|r}$ for NOPs and $JND_{l|r}$ for OPs. $\bar{S}$ represents the visual saliency normalized in the range of $[0, 1]$. In addition, the parameters Ts and α, bounded in [0, 1], control the impact of VS on $SSJND$. For the next section, we set $Ts = 0.5$, and $\alpha = 0.6$.

8.4 Experimental validation

In this section, we validate the performance of the proposed SSJND model by comparing with three very recent 3D-JND models, *i.e.*, BJND [4], JJND [5] and DJND [10], as well as the SSJND model without considering saliency (DBJND).

To achieve this, we use twelve stereo pairs from the Middlebury stereo datasets [33]. Similar to [34] and [35], we compare the perceptual quality between the noise-injected S3D images relying on different 3D-JND models under the same noise level. Note that the noise is injected only in the right image of the stereo pair in order to simulate an asymmetric distortion. The S3D image I^* contaminated by the JND-based noise is calculated as: $I^*(k) = I(k) + C_n \cdot N_{rand}(k) \cdot JND(k)$, where I denotes the original image. C_n is a control parameter that makes the same noise level for different 3D-JND models leading to the same peak signal-to-noise ratio (PSNR), *i.e.* PSNR $\in$ [28dB, 29dB].

To subjectively compare our model to the state-of-the-art, we use the same experimental setup as

Table 8.2 – Quality comparison between our SSJND and state-of-the-art models using 12 images from the Middlebury stereo datasets.

S3D	vs.DBJND		vs. BJND [4]		vs. JJND [5]		vs. DJND [10]	
image	$\bar{M}$	p-value	$\bar{M}$	p-value	$\bar{M}$	p-value	$\bar{M}$	p-value
Art	0.39	0.0001	0.06	0.0001	1.44	0.0058	1.61	0.0015
Reindeer	0.72	0.0001	0.33	0.0001	2.56	0.0001	2.89	0.0001
Moebius	0.39	0.0001	-0.17	0.0001	2.06	0.0001	2.22	0.0001
Dolls	0.72	0.0002	0.50	0.0001	1.72	0.0001	0.94	0.0001
Aloe	0.39	0.0001	0.83	0.0006	0.78	0.0016	1.22	0.0027
Baby2	0.17	0.0034	0.11	0.0004	-0.50	0.0131	0.78	0.0001
Midd2	0.56	0.0001	0.22	0.0007	-0.94	0.0001	0.83	0.0002
Plastic	0.56	0.0001	0.28	0.0001	0.89	0.0045	0.44	0.0013
Motorcycle	-0.11	0.0001	-0.17	0.0001	1.06	0.0001	1.94	0.0001
Piano	-0.22	0.0001	0.28	0.0002	2.56	0.0001	2.33	0.0001
Playroom	0.44	0.0001	0.22	0.0006	1.89	0.0001	1.44	0.0001
Playtable	0.22	0.0001	0.56	0.0001	1.00	0.0052	1.44	0.0007
Average	**0.32**	**0.0003**	**0.24**	**0.0002**	**1.12**	**0.0023**	**1.39**	**0.0005**

for previous PEs. The room ambient illumination and the viewing distance are set to 100 lux and 1.8 m, respectively. Furthermore, eighteen subjects are invited to participate the test. Note that two subjects (side-by-side) participate to the test simultaneously while the influence of viewing direction on the quality judgment will be investigated later. We opted for the stimulus-comparison method described in the ITU-R BT.2021-1 [18]. Firstly, a mid-grey image with zero disparity, containing the image sequence number, is presented to the subjects for 2s. Then, a couple of JND-based distorted 3D images (SSJND and other 3D-JND model) are shown with random position on a mid-gray background for 10s. Subsequently, subjects are asked to provide a score depending on the preference: 0 (the same), 1 (slightly better), 2 (better), 3 (much better). These scores are then used to compute the mean opinion score over all subjects for each S3D image. In addition, we use the Pearson's chi-squared test [36] to verify the statistical significance of the comparative scores. The adopted null hypothesis of this test is: "there is no preference between the proposed SSJND model and the other 3D-JND models".

Table 8.2 shows the quality comparison results in terms of mean opinion scores and p-values for each image. $p - value < 0.05$ for all pair comparison cases rejects the null hypothesis, and thus validates the statistical significant preference between the proposed model and the other 3D-JND models.

Overall, SSJND outperforms all the other models on almost all the used images. Complex scenes may lead to difficulties in VS estimation where SSJND may overestimate the JND thresholds for smooth regions with high luminance intensity when the latter regions are considered as non-salient.

Compared to the BJND, the proposed SSJND model considers occlusion effect, and thus globally provides better estimation for S3D image containing large number of occluded pixels. In the same vein, our model performs quite better than the JJND and DJND models in terms of average scores, because they are both developed based on 2D-JND, which makes them less reliable than the 3D-JND model based on PEs. As a conclusion, our SSJND model performs better for almost the whole dataset except for some rare cases, where it should be noticed that the difference is close to 0.

The results of ANOVA with the null hypothesis of no significant difference of the subject position in terms of subjective scores, give $p - value = 0.28, 0.89, 0.78$, and 0.99 respectively for the DBJND, BJND, JJND and DJND models, and indicate that the viewing direction has not significant influence on subjective scores.

8.5 Conclusion

In this paper, we propose a saliency-weighted stereoscopic JND (SSJND) model. To this end, we first conduct psychophysical experiments in which we measure the visibility thresholds of the asymmetric noise. The psychophysical data is used to develop a disparity-aware binocular JND (DBJND) model allowing to estimate the JND thresholds for non-occluded pixels. The SSJND profile is build on top of DBJND by including a 2D-JND model for occluded-pixels and accounting for visual saliency. The experimental validation shows that the proposed model outperforms the other 3D-JND models in terms of perceptual quality at the same noise level. A more reliable VS detection approach and an effective VS-map-based weighting function will be investigated in the future to improve the effectiveness of the proposed 3D-JND model.

Bibliography

[1] Ee-Leng Tan and Woon-Seng Gan. Computational models for just-noticeable differences. In *Perceptual Image Coding with Discrete Cosine Transform, SpringerBriefs in Electrical and Computer Engineering*, pages 3–19. Springer, Singapore, 2015. 168

[2] D Varuna S X De Silva, Erhan Ekmekcioglu, Warnakulasuriya Anil Chandana Fernando, and Stewart T Worrall. Display dependent preprocessing of depth maps based on just noticeable depth difference modeling. *IEEE J. Sel. Top. Signal Process.*, 5(2):335–351, 2011. 168

[3] Seung-Won Jung. A modified model of the just noticeable depth difference and its application to depth sensation enhancement. *IEEE Trans. Image Process.*, 22(10):3892–3903, 2013. 168

[4] Yin Zhao, Zhenzhong Chen, Ce Zhu, Yap-Peng Tan, and Lu Yu. Binocular just-noticeable-difference model for stereoscopic images. *IEEE Signal Process. Lett.*, 18(1):19–22, 2011. 168, 170, 172, 174, 175

[5] Xiaoming Li, Yue Wang, Debin Zhao, Tingting Jiang, and Nan Zhang. Joint just noticeable difference model based on depth perception for stereoscopic images. In *2011 IEEE Int. Conf. Vis. Commun. Image Process.*, pages 1–4, Tainan, Taiwan, Nov. 2011. 168, 174, 175

[6] Fengzong Lian, Shaohui Liu, Xiaopeng Fan, Debin Zhao, and Wen Gao. A new just-noticeable-distortion model combined with the depth information and its application in multi-view video coding. In *The Era of Interactive Media*, pages 229–240. Springer New York, 2013. 168

[7] Chao Liu, Ping An, Yifan Zuo, and Zhaoyang Zhang. Applications of just-noticeable depth difference model in joint multiview video plus depth coding. In *Proc. SPIE*, volume 9273, pages 9273–9273–11, 2014. 168

[8] Rui Zhong, Ruimin Hu, Zhongyuan Wang, and Shizheng Wang. 3D hybrid just noticeable distortion modeling for depth image-based rendering. *Multimed. Tools Appl.*, 74(23):10457–10478, 2015. 168

[9] Feng Qi, Debin Zhao, Xiaopeng Fan, and Tingting Jiang. Stereoscopic video quality assessment based on visual attention and just-noticeable difference models. *Signal, Image Video Process.*, 10 (4):737–744, 2016. 168

[10] Fei Xue, Cheolkon Jung, and Joongkyu Kim. Disparity-based just-noticeable-difference model for perceptual stereoscopic video coding using depth of focus blur effect. *Displays*, 42:43–50, 2016. 168, 174, 175

[11] Yu Yang, Chaohui Lu, Jianzeng Li, and Huanqi Yao. Just noticeable depth difference of human during viewing of dynamic random dot stereograms. In *9th Inter. Symposium on Computational Intelligence and Design (ISCID)*, volume 1, pages 422–424, Hangzhou, China, Dec. 2016. IEEE. 168

[12] Chunhua Li, Ping An, Liquan Shen, Kai Li, and Jian Ma. A modified just noticeable depth difference model for 3d displays. In *Int. Forum of Digital TV and Wireless Multimedia Commun.*, pages 63–71. Springer Singapore, 2017. 168

[13] X K Yang, W S Ling, Z K Lu, Ee Ping Ong, and S S Yao. Just noticeable distortion model and its applications in video coding. *Signal Process. Image Commun.*, 20(7):662–680, 2005. 168

[14] KN Gelatt and DW Esson. Physiology of the eye. *Veterinary Ophthalmology. 4th ed. Iowa: Blackwell Publishing Ltd*, pages 149–182, 2007. 168

[15] Elizabeth R Schotter, Bernhard Angele, and Keith Rayner. Parafoveal processing in reading. *Attention, Perception, Psychophys.*, 74(1):5–35, 2012. 169

[16] Hak Gu Kim, Seong-il Lee, and Yong Man Ro. Experimental investigation of the effect of binocular disparity on the visibility threshold of asymmetric noise in stereoscopic viewing. *Opt. Express*, 24(17):19607–19615, 2016. 170

[17] Frédéric Devernay and Paul A Beardsley. Stereoscopic Cinema. *Image Geom. Process. 3-D Cinematogr.*, 5:11–51, 2010. 170

[18] Subjective methods for the assessment of stereoscopic 3DTV systems. Tech. Rep. ITU-R BT.2021-1, Int. Telecommun. Union, Geneva Switzerland, 2015. 171, 175

[19] David H Brainard and Spatial Vision. The psychophysics toolbox. *Spat. Vis.*, 10:433–436, 1997. 171

[20] Yiming Li, Hongyi Liu, and Zhenzhong Chen. Perceptually-lossless image coding based on foveated-JND and H. 265/HEVC. *J. Vis. Commun. Image Represent.*, 40:600–610, 2016. 171

[21] Haiqiang Wang, Ioannis Katsavounidis, Jiantong Zhou, Jeonghoon Park, Shawmin Lei, Xin Zhou, Man-On Pun, Xin Jin, Ronggang Wang, Xu Wang, and Others. VideoSet: A large-scale compressed video quality dataset based on JND measurement. *J. Vis. Commun. Image Represent.*, 46:292–302, 2017. 171

[22] Methodology for the subjective assessment of video quality in multimedia applications. Tech. Rep. ITU-R BT.1708-0, Int. Telecommun. Union, Geneva Switzerland, 2007. 171

[23] Pete R Jones. A note on detecting statistical outliers in psychophysical data. *bioRxiv*, 2016. 171

[24] Carlos M Jarque and Anil K Bera. A test for normality of observations and regression residuals. *Int. Stat. Rev.*, pages 163–172, 1987. 171

[25] Samuel Sanford Shapiro and Martin B Wilk. An analysis of variance test for normality (complete samples). *Biometrika*, 52(3/4):591–611, 1965. 172

[26] Howard Levene and Others. Robust tests for equality of variances. *Contrib. to Probab. Stat.*, 1: 278–292, 1960. 172

[27] Philip E Gill and Walter Murray. Algorithms for the solution of the nonlinear least-squares problem. *SIAM J. Numer. Anal.*, 15(5):977–992, 1978. 173

[28] Daniel Scharstein and Richard Szeliski. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. *Int. J. Comput. Vis.*, 47(1-3):7–42, 2002. 174

[29] Anmin Liu, Weisi Lin, Manoranjan Paul, Chenwei Deng, and Fan Zhang. Just noticeable difference for images with decomposition model for separating edge and textured regions. *IEEE Trans. Circuits Syst. Video Technol.*, 20(11):1648–1652, 2010. 174

[30] Hadi Hadizadeh. A saliency-modulated just-noticeable-distortion model with non-linear saliency modulation functions. *Pattern Recognit. Lett.*, 84:49–55, 2016. 174

[31] Hadi Hadizadeh, Atiyeh Rajati, and Ivan V Bajić. Saliency-guided just noticeable distortion estimation using the normalized laplacian pyramid. *IEEE Signal Process. Lett.*, 24(8):1218–1222, 2017. 174

[32] Yuming Fang, Junle Wang, Manish Narwaria, Patrick Le Callet, and Weisi Lin. Saliency detection for stereoscopic images. *IEEE Trans. Image Process.*, 23(6):2625–2636, 2014. 174

[33] Daniel Scharstein and Richard Szeliski. Middlebury stereo datasets. *Online at http://vision.middlebury.edu/stereo/data/*, 2002. 174

[34] Shiqi Wang, Lin Ma, Yuming Fang, Weisi Lin, Siwei Ma, and Wen Gao. Just noticeable difference estimation for screen content images. *IEEE Trans. Image Process.*, 25(8):3838–3851, 2016. 174

[35] Jinjian Wu, Leida Li, Weisheng Dong, Guangming Shi, Weisi Lin, and C-C Jay Kuo. Enhanced just noticeable difference model for images with pattern complexity. *IEEE Trans. Image Process.*, 26(6):2682–2693, 2017. 174

[36] N Balakrishnan, Vassilly Voinov, and Mikhail Stepanovich Nikulin. *Chi-squared goodness of fit tests with applications*. Academic Press, 2013. 175

Chapter 9

Paper IV: Stereoscopic Image Quality Assessment based on the Binocular Properties of the Human Visual System

Y. Fan[1,2], M.-C. Larabi[1], F. A. Cheikh[2], C. Fernandez-Maloigne[1]

[1]XLIM UMR CNRS 7252, University of Poitiers, France
[2]Norwegian Colour and Visual Computing Lab, Norwegian University of Science and Technology, Gjøvik, Norway

In *IEEE International Conference on Acoustics, Speech and Signal Processing(ICASSP)*, pages 2037–2041, March 2017.

Abstract

One of the most challenging issues in stereoscopic image quality assessment (IQA) is how to effectively model the binocular behaviors of the human visual system (HVS). The latter has a great impact on the perceptual stereoscopic 3D (S3D) quality. This paper presents a stereoscopic IQA metric based on the properties of the HVS. Instead of measuring the quality of the left and the right views separately, the proposed method predicts the quality of a cyclopean image to ensure that the overall S3D quality is as close as possible to the binocular vision. The cyclopean image is synthesized based on the local entropy of each view with the aim to simulate the phenomena of the binocular rivalry/suppression. A 2D IQA metric is employed to assess the quality of both the cyclopean image and the disparity map.

Additionally, the quality of the cyclopean image is modulated according to the visual importance of each pixel defined by the just noticeable difference (JND). Finally, the 3D quality score is derived by combining the quality estimates of the cyclopean image and disparity map. Experimental results show that the proposed method outperforms many other state-of-the-art SIQA methods in terms of prediction accuracy and computational efficiency.

Index terms– stereoscopic image quality assessment, cyclopean image, binocular rivalry/suppression, just noticeable difference (JND).

9.1 Introduction

In the past few years, great efforts in Stereoscopic 3D (S3D) technologies have been made to bring a realistic 3D visual experience to consumers. However, S3D technology development brings some challenges especially to 3D-TVs makers. One of the major challenges is linked to the user's quality of experience (QoE) including comfort and fatigue aspects. In order to achieve this, it's important to develop accurate and reliable IQA metrics for 3D stereoscopic content. While 2D IQA has greatly advanced in the recent years, stereoscopic IQA (SIQA) is only in its infancy. Mainly because 3D perceptual quality can be affected by the characteristics of both monocular and binocular vision. Even though 3D quality can be measured using subjective experiments [1], these are tedious and expensive. Therefore, objective metrics are needed to automatically assess the perceived 3D visual quality.

A stereo pair contains two slightly different views (i.e., left and right views), each of which is projected separately onto the retina. When a S3D image is observed, the human visual system (HVS) merges the two views to yield a single mental view (i.e., cyclopean image) based on the properties of the binocular vision [2]. Thereby, the 3D perceptual quality depends not only on the quality of each individual view [3], but also on the depth information [4] and the binocular characteristics [5]. The idea is to explore how these attributes contribute to the overall 3D quality. Therefore, to design reliable and accurate S3D metrics, it is important to understand and account for the different perceptual processes of the HVS.

In this paper, we propose a new SIQA metric based on the HVS properties, combining the quality scores of the cyclopean image [5] and the disparity map. The major contribution of this work lies in the development of a novel 3D quality metric by modeling the phenomena of binocular rivalry/suppression, and accounting for disparity map quality as well as the monocular spatial sensitivity of the HVS. Besides, we provide a comprehensive experimental evaluation for our proposed method, and an extensive comparison with other SIQA methods. The remainder of the paper is organized as follows. In Sect. 9.2, we provide a brief review of recent SIQA metrics. Sect. 9.3 describes the proposed SIQA method. We evaluate and discuss the performance of the proposed metric in Sect. 9.4. This paper ends with some conclusions and future work.

9.2 Related work

In this section, we briefly review the recent SIQA methods. Based on the type and the amount of the information used from stereoscopic views, the SIQA methods can be divided into three categories [6]: (1) stereo-pair-based methods, (2) methods based on stereo-pair and depth information, (3) methods considering the HVS properties.

9.2.1 Stereo-pair-based methods

The SIQA methods of the first category try to extend the 2D IQA algorithms directly to measure the distortions of S3D images. Most early approaches [7, 8] assess the quality of left and right views separately using state-of-the-art 2D quality metrics, and then combine both scores into an overall 3D quality score. For instance, Campisi *et al.* [7] evaluated the S3D quality by four 2D quality metrics including structural similarity metric (SSIM) [9], universal image quality index (UQI) [10], C4 [11] and reduced-reference QA [12]. However, considering the combination of the qualities for each view as an overall 3D quality does not correlate well with the human quality judgments [13]. This is mainly due to the fact that these 2D metrics do not take into account depth information, which plays an important role on 3D perception.

9.2.2 Methods based on stereo-pair and depth information

Consequently, the second category employs both views of a stereo pair in addition to depth/disparity information to estimate 3D quality. In this category, 2D quality metrics are used to measure the quality of both the stereo-pair and the disparity map. Then, these two quality values are combined to yield a 3D quality score. In an early research, Benoit *et al.* [3] proposed a full reference 3D metric that applies SSIM and C4 metrics on left and right images independently, and then combined these 2D scores with the estimate of disparity map distortion. Later, You *et al.* [14] explored the performance of 2D quality metrics used in the context of 3D quality assessment with different ways of combining between the disparity map quality and views' quality. Hwang and Wu [15] developed a 3D quality prediction model that integrates the stereo-pair quality with depth quality and S3D visual saliency. Recently, Wang *et al.* [16] designed a reduced reference SIQA model, considering the quality of both luminance images and disparity map, based on image statistics in the contourlet domain. Since the ground truth depth/disparity maps are not always available, this category of methods estimate the disparity maps by using stereo matching algorithms. Thereby the accuracy of the stereo matching algorithms may affect the performance of 3D quality prediction.

9.2.3 Methods considering the HVS properties

In fact, the views of a stereo-pair may suffer from an equal amount of distortion (namely symmetric distortion) or different amounts and/or types of distortions (namely asymmetric distortion). Symmet-

ric distortion results in binocular fusion [17], whereas asymmetric distortions lead to either binocular rivalry [18] or binocular suppression[19] depending on the strength of the difference. These latter have a great impact on perceived 3D quality. The SIQA methods of the two above-mentioned categories are quite useful in the case of symmetric distortion, but perform much less effectively for asymmetrically distorted stereo-pairs that are very common in real application such as coding. Thus, to improve the performance of the 3D metric, the third category of SIQA methods consider the monocular and/or binocular visual properties in addition to stereo-pair quality and depth information.

It is known that the human eyes are incapable of perceiving pixel changes below a specific visual threshold namely the just noticeable difference (JND) due to their underlying temporal/spatial sensitivity and masking effects [20]. Some JND models for S3D content (3D-JND) accounting for both monocular and binocular depth cues have been proposed [21]. For instance, a binocular just noticeable difference (BJND) model [22], which investigates the properties of the binocular vision in response to asymmetric noise in a stereo-pair based on HVS characteristics, has been applied in 3D quality estimation[23, 24].

Other SIQA approaches combine left and right views into one cyclopean image, and the final 3D quality is measured by analyzing this merged image. For example, Chen *et al.* [5] developed a SIQA metric by computing the quality of the cyclopean images constructed by a linear model. The weights of this model are derived from the Gabor filter magnitude responses, which simulate the binocular rivalry. Similarly, Fezza and Larabi [25] proposed a full reference SIQA method based on the quality of the test cyclopean image generated by using local entropy and depth information. Besides, Lin and Wu [26] predicted the 3D quality based on both binocular combination and binocular frequency integration. In the following section, we propose a SIQA method that estimates the degradations of cyclopean image and disparity map.

9.3 The Proposed SIQA Method

As mentioned above, the HVS is not sensitive to the quality in the left or right image separately. Instead, it perceives distortions of the cyclopean image as 2D impairments, and depth/disparity distortion as 3D impairment. Thereby our proposed SIQA method is based on the assumption that the overall 3D quality is a combination of the qualities of binocular-based cyclopean image and the disparity map. Figure 1 shows the framework of the proposed SIQA method. This 3D quality prediction model consists of 5 steps:

1. Disparity estimation for both reference and distorted stereo pairs;
2. Formation of the cyclopean image for each stereo pair based on local entropy;
3. Quality assessment of the cyclopean image and disparity map separately using the UQI metric;
4. Weighting the cyclopean image quality with the JND map of the reference cyclopean image;

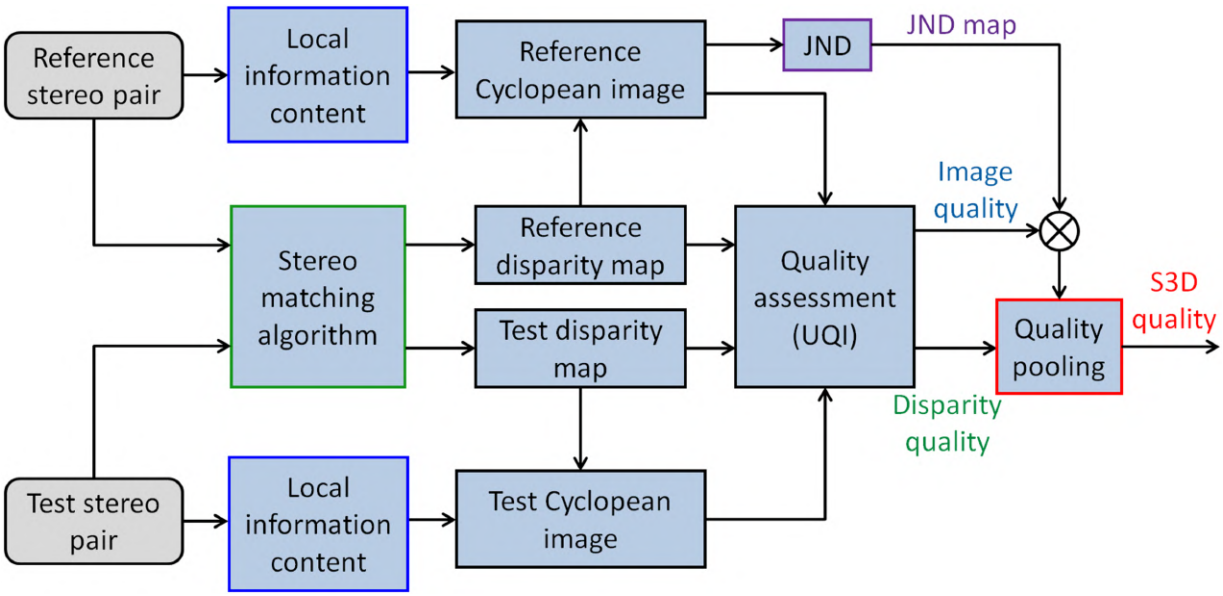


Figure 9.1 – Framework of the proposed SIQA method.

5. S3D quality estimation by combining the quality of the JND-based cyclopean image with the qualiy of disparity map.

The first step is to form the cyclopean images. According to a linear model proposed in [5, 27], by modeling the rivalry/suppression when a stereo stimulus is presented, we synthesize the cyclopean image as follows:

$$I_c(i,j) = W_l(i,j) \times I_l(i,j) + W_r(i,j-d_l) \times I_r(i,j-d_l), \tag{9.1}$$

where I_l and I_r represent the left and right images respectively, and I_c is the cyclopean image. W_l and W_r are the weighting coefficients for their corresponding images, and used to describe the rivalry process, thus $W_l + W_r = 1$. Moreover, i and j are the pixel coordinates, and $d_l(i,j)$ represents the disparity value of the pixel (i,j) of left image that corresponds to the horizontal shift of one pixel from the left to the right image. To determine the disparity map, we propose to use a stereo matching algorithm recently proposed by Lee *et al.* [28]. This algorithm efficiently achieves high performance in disparity estimation and deals with the issues of occlusion and depth discontinuities.

As described in [27], the experience of binocular rivalry is correlated to the relative stimulus strength of each view instead of absolute stimulus strength. Moreover, the studies in [5] [13] found that the 3D human perception is dominated by the view of high contrast or rich contours. In other words, the perceptual 3D quality follows the quality of the view containing a higher amount of information. Therefore, the local information content is used to determine the relative stimulus strengths W_l and W_r of two views, where $W_l(i,j)$ and $W_r(i,j)$ are defined by:

$$W_l(i,j) = \frac{EN_l(i,j)}{EN_T(i,j)}, W_r(i,j-d_l) = \frac{EN_r(i,j-d_l)}{EN_T(i,j)}, \tag{9.2}$$

$$EN_T(i,j) = EN_l(i,j) + EN_r(i,j-d_l), \tag{9.3}$$

where $EN_l(i,j)$ and $EN_r(i,j)$ are the left and right local entropy of the pixel (i,j) in the left and right views respectively. The image entropy is related to the amount of information that can be coded in the compression process. For instance, a low entropy image contains very little contrast. The entropy of a pixel computed based on 11-by-11 neighborhood with specific shape around this pixel[29] is described as follows:

$$EN(i,j) = -\sum_{s=g_{min}}^{g_{max}} p(x_s) \times \log_2(p(x_s)), \tag{9.4}$$

where g_{min} and g_{max} are minimum and maximum values respectively in the corresponding neighborhood pixels. $p(x_s)$ denotes the probability that the difference between two adjacent pixels is equal to s. Based on equations 2, 3 and 4, our SIQA method simulates the binocular rivalry/suppression. For example, different local entropies in two views lead to binocular rivalry/suppression, and the 3D quality is more affected by the view containing higher local entropies.

Given the cyclopean images (I_{rc}, I_{dc}) and the disparity maps (Dp_r, Dp_d) of the reference and distorted stereo pairs, we independently measure the quality of the cyclopean image and the disparity map by using 2D IQA metric. In [14], You *et al.* found that UQI performs the best for 3D quality prediction among all the tested 2D IQA metrics. On the other hand, UQI metric has the best performance for IQA on the disparity map. Actually, UQI used in disparity quality estimation is based on comparing the structural information, and the disparity can express such information of the original images.

Thereby we propose to employ UQI to predict the quality of the stereo pair and disparity map independently:

$$Q_c(i,j) = UQI(I_{rc}, I_{dc}), Q_d = UQI(Dp_r, Dp_d), \tag{9.5}$$

where Q_c is the UQI index map of the test cyclopean image, and Q_d denotes the quality score of the disparity map. In order to improve the SIQA performance, we used the visual importance of the pixel to weight the cyclopean quality score [23]. The visual importance, which corresponds to monocular spatial sensitivity of HVS, is described by JND thresholds [30] of the reference cyclopean image. Accordingly, the JND-based cyclopean quality Q_c^{JND} is calculated by:

$$Q_c^{JND} = \frac{\sum_{i,j}^{N} \left[\frac{1}{JND(i,j)} \times Q_c(i,j)\right]}{\sum_{i,j}^{N} \frac{1}{JND(i,j)}}, \tag{9.6}$$

where N is the number of pixels in the cyclopean image. High value of the JND in a pixel means that this pixel can tolerate a large degradation, and thus has a low visual importance in the perceptual

Table 9.1 – Performance of SIQA methods on LIVE 3D IQA database (phase II). The symbols AS and S are respectively the asymmetric and symmetric distortions. CT denotes the computational runtime (in second) for all images. Italicized entries are 2D quality metrics, while the best performance are bolded.

Method	LCC			SROCC			RMSE			CT
	S	As	Total	S	As	Total	S	As	Total	Total
SSIM [9]	*0.852*	*0.767*	*0.802*	*0.826*	*0.736*	*0.793*	*6.543*	*6.510*	*6.736*	***30***
MS-SSIM [31]	*0.927*	*0.719*	*0.795*	*0.912*	*0.684*	*0.777*	*4.694*	*7.047*	*6.851*	*49*
FSIM [32]	*0.929*	*0.731*	*0.808*	*0.912*	*0.684*	*0.786*	*4.623*	*6.913*	*6.654*	*919*
VIF [33]	*0.928*	*0.777*	*0.837*	*0.916*	*0.732*	*0.819*	*4.653*	*6.383*	*6.184*	*684*
UQI [10]	***0.940***	*0.794*	*0.863*	***0.938***	*0.755*	*0.841*	***4.223***	*6.159*	*5.685*	*38*
Wang [23]	0.862	0.743	0.771	0.826	0.696	0.771	6.334	6.787	7.188	82
Fezza [24]	0.788	0.713	0.751	0.778	0.676	0.734	7.685	7.104	7.453	163
Fezza [25]	0.930	0.820	0.871	0.921	0.796	0.862	4.576	5.801	5.553	1410
Chen [5]	0.939	**0.878**	**0.909**	0.927	**0.858**	**0.904**	4.277	**4.846**	**4.700**	14089
Proposed	**0.940**	0.875	0.906	**0.938**	0.839	0.893	4.272	4.903	4.795	2392

quality. Finally, the S3D quality score Q_{3D} is calculated by a linear model:

$$Q_{3D} = \alpha \times Q_c^{JND} + \beta \times Q_d \tag{9.7}$$

where α and β are the weights of the 2D JND-based cyclopean quality and the disparity quality respectively, with $\alpha + \beta = 1$. In our implementation, we assume that the 2D quality has more importance than disparity quality, thus we fixed $\alpha = 0.6$ and $\beta = 0.4$.

9.4 Experiemntal results

In this section, we evaluate the performance of the proposed SIQA method on the publicly available LIVE 3D IQA database (phase II) [34]. LIVE 3D IQA database is composed of 8 original images and 360 distorted stereo pairs with symmetric and asymmetric distortions, including additive white gaussian noise (WN), gaussian blur (Gblur), JPEG, JPEG 2000 compression (JP2K) and fast fading (FF). We compare the proposed method with four other SIQA methods[5, 23–25]. For the SIQA methods, we used the same stereo matching algorithm [28] to estimate the disparity maps to ensure a fair comparison. In addition, we evaluate the performance of SIQA methods using only 2D metrics including SSIM, MS-SSIM [31], FSIM [32], VIF [33] and UQI. For these 2D-based SIQA methods, we estimated the 3D perceptual quality by averaging the quality predicted from the left and right views. The performance of the 3D quality metrics has been evaluated using three well-known measures: the Linear Correlation Coefficient (LCC), the Spearman Rank Order Correlation Coefficient (SROCC) and RMSE. Three measures were computed between DMOS and the predicted scores after a non-linear regression with a five-parameter logistic function described in [35]. All tests were performed by

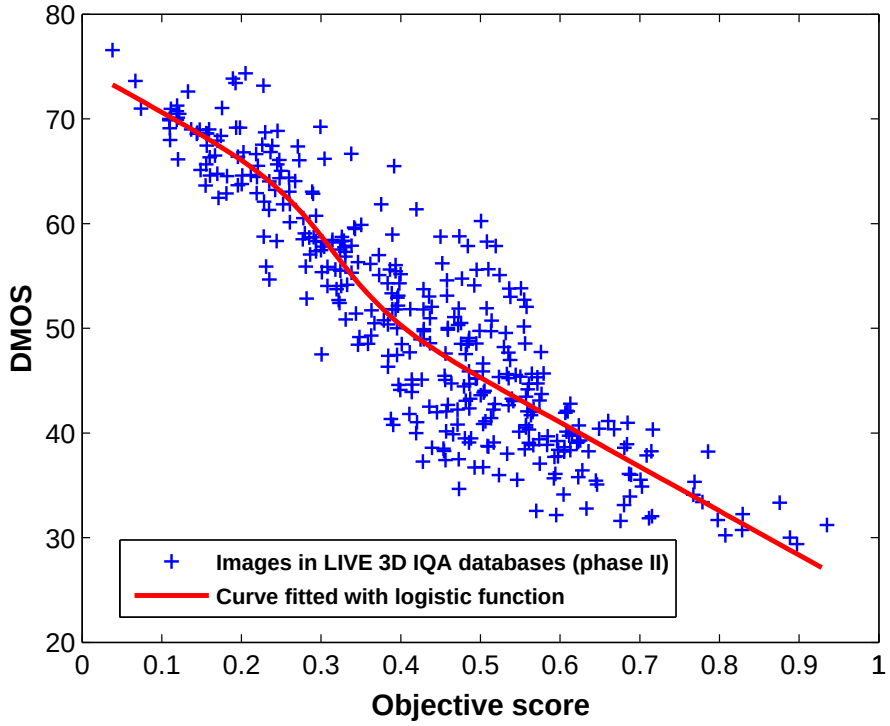


Figure 9.2 – Scatter plots of DMOS versus predicted scores obtained by proposed SIQA method.

running MATLAB code on a portable computer (Inter Core i7-2630 QM Processor at 2.00 GHz, 4 GB RAM, Windows 7).

9.4.1 Overall performance

Table 9.1 shows the performance of SIQA methods on LIVE 3D IQA database. These results demonstrate that the proposed method outperforms the others methods except Chen's method for the cases of symmetric and asymmetric distortions. Actually, the proposed method is quite similar to Chen's method [5] in terms of overall performance, but the proposed method is obviously much faster than Chen's method. To summarize, our proposed method achieves high performance with low computational costs. On the other hand, most of 2D-based SIQA methods are as efficient as the 3D IQA methods for the symmetrically distorted stereo pairs, but they generally give bad performance than 3D IQA methods for asymmetric distortions. This is mainly due to the fact that 2D-based SIQA methods evaluate the S3D quality without considering neither the depth/disparity information nor the characteristics of the binocular vision. It should be noted that the method using UQI metric performs best within all 2D-based SIQA methods.

The performance of Wang's method [23] and Fezza's method [24] are lower than the proposed approach despite their use of binocular properties. This may be explained by the fact of predicting

Table 9.2 – Performance of the proposed SIQA method on LIVE 3D IQA database (phase II).

Strategies	LCC	SROCC	RMSE
without JND	0.902	0.883	4.870
without DQA	0.889	0.864	5.163
without JND and DQA	0.887	0.866	5.178
with JND and DQA	**0.906**	**0.893**	**4.795**

3D quality separately of left and right views, and failing in accounting for the binocular properties. Thereby the methods based on cyclopean image (i.e., Chen's [5], Fezza 's [25] and our proposed method) achieved better performance than other 3D IQA methods. In addition to the performance comparison mentioned above, we provide in Figure 9.2 the scatter distributions of DMOS versus predicted scores obtained with the proposed method, as well as the non-linear fitting curve.

9.4.2 Discussion about the proposed strategy

In this section, we show the advantages of considering both JND and quality assessment for disparity map in our SIQA method. We compare the performance and the influence of each component of the proposed metric (see Figure 9.1). The performance of the four SIQA methods on one database are shown in Table 9.2. SIQA method without JND does not use the JND map to weight the quality of reference cyclopean image, whereas the SIQA method without quality assessment for disparity map (DQA) does not consider the quality of disparity map. From the results, we can notice that the proposed SIQA method (i.e, with JND and DQA) gives the best performance among all strategies. However, the proposed method slightly outperforms method without JND in terms of LCC. In addition, SIQA method without JND performs better than SIQA method without DQA. This can be explained by the fact that the depth information is more important than the sensitivity of HVS for 3D quality prediction. In summary, the results of Table 9.2 mean that 3D quality prediction performance can be improved by accounting for both JND and disparity quality estimation. We also explored the performance of proposed method for different types of distortions. Our method performs quite well for both GBlur and FF distortion. We cannot show here due to page limitation.

9.5 Conclusion

In this paper, we proposed a quality assessment method for stereoscopic images based on HVS properties. Our method models the human stereo vision by fusing the left and right views to generate a cyclopean image, and taking into account the disparity information as well as the monocular spatial sensitivity of HVS. The experimental results showed that the proposed method outperforms well-known 2D-based SIQA methods and 3D IQA methods in terms of prediction accuracy and computational costs. In future works, the performance of the proposed method will be evaluated on other databases.

In addition, other reliable approaches modeling the process of binocular rivalry will be considered to improve the performance of our method.

Acknowledgment

This research has been funded by both the Research Council of Norway through project no. 221073 "HyPerCept – Colour and quality in higher dimensions" and the Region "Nouvelle Aquitaine".

Bibliography

[1] ITU. Subjective methods for the assessment of stereoscopic 3dtv systems. *Recommendation ITU-R BT*, 2021. 182

[2] Bela Julesz. Foundations of cyclopean perception. 1971. 182

[3] Alexandre Benoit, Patrick Le Callet, Patrizio Campisi, and Romain Cousseau. Quality assessment of stereoscopic images. *EURASIP journal on image and video processing*, 2008. 182, 183

[4] ZM Sazzad, Roushain Akhter, J Baltes, and Yuukou Horita. Objective no-reference stereoscopic image quality prediction based on 2d image features and relative disparity. *Advances in Multimedia*, 2012:8, 2012. 182

[5] Ming-Jun Chen, Che-Chun Su, Do-Kyoung Kwon, Lawrence K Cormack, and Alan C Bovik. Full-reference quality assessment of stereopairs accounting for rivalry. *Signal Processing: Image Communication*, 28(9):1143–1155, 2013. 182, 184, 185, 187, 188, 189

[6] Yi Zhang and Damon M Chandler. 3d-mad: A full reference stereoscopic image quality estimator based on binocular lightness and contrast perception. *IEEE Trans. Image Process.*, 24(11):3810–3825, 2015. 183

[7] Patrizio Campisi, Patrick Le Callet, and Enrico Marini. Stereoscopic images quality assessment. In *Signal Processing Conference, 2007 15th European*, pages 2110–2114. IEEE, 2007. 183

[8] Jiachen Yang, Chunping Hou, Yuan Zhou, Zhuoyun Zhang, and Jichang Guo. Objective quality assessment method of stereo images. In *3DTV Conference: The True Vision-Capture, Transmission and Display of 3D Video*, pages 1–4. IEEE, 2009. 183

[9] Zhou Wang, Alan Conrad Bovik, Hamid Rahim Sheikh, and Eero P Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE Trans. Image Process.*, 13(4): 600–612, 2004. 183, 187

[10] Zhou Wang and Alan C Bovik. A universal image quality index. *IEEE Signal Processing Letters*, 9(3):81–84, 2002. 183, 187

[11] Mathieu Carnec, Patrick Le Callet, and Dominique Barba. An image quality assessment method based on perception of structural information. In *ICIP*, volume 3, pages III–185. IEEE, 2003. 183

[12] Zhou Wang and Eero P Simoncelli. Reduced-reference image quality assessment using a wavelet-domain natural image statistic model. In *Electronic Imaging 2005*, pages 149–159. International Society for Optics and Photonics, 2005. 183

[13] Daniel V Meegan, Lew B Stelmach, and W James Tam. Unequal weighting of monocular inputs in binocular combination: implications for the compression of stereoscopic imagery. *Journal of Experimental Psychology: Applied*, 7(2):143, 2001. 183, 185

[14] Junyong You, Liyuan Xing, Andrew Perkis, and Xu Wang. Perceptual quality assessment for stereoscopic images based on 2d image quality metrics and disparity analysis. In *Proc. of Int. Workshop on Video Process. and Quality Metrics for Consumer Electronics, Scottsdale, AZ, USA*, 2010. 183, 186

[15] Jae-Jeong Hwang and Hong Ren Wu. Stereo image quality assessment using visual attention and distortion predictors. *TIIS*, 5(9):1613–1631, 2011. 183

[16] Xu Wang, Qiong Liu, Ran Wang, and Zhuo Chen. Natural image statistics based 3d reduced reference image quality assessment in contourlet domain. *Neurocomputing*, 151:683–691, 2015. 183

[17] Jeremy M Wolfe. Stereopsis and binocular rivalry. *Psychological review*, 93(3):269, 1986. 184

[18] Randolph Blake and Nikos K Logothetis. Visual competition. *Nature Reviews Neuroscience*, 3 (1):13–21, 2002. 184

[19] Jan Brascamp, Hansem Sohn, Sang-Hun Lee, and Randolph Blake. A monocular contribution to stimulus rivalry. *PNAS*, 110(21):8337–8344, 2013. 184

[20] Nikil Jayant, James Johnston, and Robert Safranek. Signal compression based on models of human perception. *Proceedings of the IEEE*, 81(10):1385–1422, 1993. 184

[21] Yu Fan, Mohamed-Chaker Larabi, Faouzi Alaya Cheikh, and Christine Fernandez-Maloigne. On the performance of 3d just noticeable difference models. In *2016 IEEE International Conference on Image Processing (ICIP)*, pages 1017–1021. IEEE, 2016. 184

[22] Yin Zhao, Zhenzhong Chen, Ce Zhu, Yap-Peng Tan, and Lu Yu. Binocular just-noticeable-difference model for stereoscopic images. *IEEE Signal Processing Letters*, 18(1):19–22, 2011. 184

[23] Xu Wang, Sam Kwong, and Yun Zhang. Considering binocular spatial sensitivity in stereoscopic image quality assessment. In *VCIP*, pages 1–4. IEEE, 2011. 184, 186, 187, 188

[24] Sid Ahmed Fezza, Mohamed-Chaker Larabi, and Kamel Mohamed Faraoun. Stereoscopic image quality metric based on local entropy and binocular just noticeable difference. In *ICIP*, pages 2002–2006. IEEE, 2014. 184, 187, 188

[25] Sid Ahmed Fezza and Mohamed-Chaker Larabi. Stereoscopic 3d image quality assessment based on cyclopean view and depth map. In *ICCE*, pages 335–339. IEEE, 2014. 184, 187, 189

[26] Yu-Hsun Lin and Ja-Ling Wu. Quality assessment of stereoscopic 3d image compression by binocular integration behaviors. *IEEE Trans. Image Process.*, 23(4):1527–1542, 2014. 184

[27] Willem JM Levelt. *On binocular rivalry*. PhD thesis, Van Gorcum Assen, 1965. 185

[28] Sehyung Lee, Jin Han Lee, Jongwoo Lim, and Il Hong Suh. Robust stereo matching using adaptive random walk with restart algorithm. *Image and Vision Computing*, 37:1–11, 2015. 185, 187

[29] Jan Beirlant, Edward J Dudewicz, László Györfi, and Edward C Van der Meulen. Nonparametric entropy estimation: An overview. *IJMSS*, 6(1):17–39, 1997. 186

[30] Anmin Liu, Weisi Lin, Manoranjan Paul, Chenwei Deng, and Fan Zhang. Just noticeable difference for images with decomposition model for separating edge and textured regions. *IEEE Trans. Circuits Syst. Video Technol.*, 20(11):1648–1652, 2010. 186

[31] Zhou Wang, Eero P Simoncelli, and Alan C Bovik. Multiscale structural similarity for image quality assessment. In *Signals, Systems and Computers, 2004. Conference Record of the Thirty-Seventh Asilomar Conference on*, volume 2, pages 1398–1402. IEEE, 2003. 187

[32] Lin Zhang, Lei Zhang, Xuanqin Mou, and David Zhang. Fsim: a feature similarity index for image quality assessment. *IEEE Trans. Image Process.*, 20(8):2378–2386, 2011. 187

[33] Hamid Rahim Sheikh and Alan C Bovik. Image information and visual quality. *IEEE Trans. Image Process.*, 15(2):430–444, 2006. 187

[34] Anush Krishna Moorthy, Che-Chun Su, Anish Mittal, and Alan Conrad Bovik. Subjective evaluation of stereoscopic image quality. *Signal Processing: Image Communication*, 28(8):870–883, 2013. 187

[35] Hamid Rahim Sheikh, Muhammad Farooq Sabir, and Alan Conrad Bovik. A statistical evaluation of recent full reference image quality assessment algorithms. *IEEE Trans. Image Process.*, 15(11): 3440–3451, 2006. 187

Chapter 10

Paper V: Full-Reference Stereoscopic Image Quality Assessment account for Binocular Combination and Disparity Information

Y. Fan[1,2], M.-C. Larabi[1], F. A. Cheikh[2], C. Fernandez-Maloigne[1]

[1]XLIM UMR CNRS 7252, University of Poitiers, France
[2]Norwegian Colour and Visual Computing Lab, Norwegian University of Science and Technology, Gjøvik, Norway

In *IEEE International Conference on Image Processing (ICIP)*, 760–764, September 2017.

Abstract

One of the most challenging issues in stereoscopic image quality assessment (SIQA) is how to effectively model the binocular behavior of the human visual system (HVS). The latter has a great impact on the perceptual 3D quality. In this paper, we propose a SIQA metric accounting for binocular combination properties and disparity information. Instead of computing the quality of the left and the right views separately, the proposed metric predicts the quality of a cyclopean image so as to have a good consistency with 3D human perception. The cyclopean image is synthesized based on the local entropy and the visual saliency of each view with the aim to simulate the phenomena of binocular fusion/rivalry. A 2D IQA metric is employed to assess the quality of both the cyclopean image and the disparity map. The obtained scores are use to derive the 3D quality score thanks to a pooling stage. Experimental

results on three public 3D IQA databases show that the proposed method outperforms many other state-of-the-art SIQA methods, and achieves high prediction accuracy on these databases.

Index terms– Stereoscopic image quality assessment, cyclopean image, binocular fusion/rivalry, visual saliency.

10.1 Introduction

In recent years, three-dimensional (3D) multimedia has become popular thanks to new sensations of immersion. With the rapid development of stereoscopic 3D (S3D) technologies, sources of 3D content and 3D display are more common nowadays. As a result, the perceptual quality assessment of 3D images and videos is quite important in order to guarantee the visual quality of experiences (QoE) at every processing stage ranging from 3D acquisition, compression, transmission and display. While 2D image quality assessment (IQA) has greatly advanced over the last decade, SIQA is still in its early stage and hence challenging [1], especially for asymmetrically distorted S3D images. This is mainly because 3D perceptual quality is affected by both monocular and binocular factors including 2D quality, disparity/depth quality and visual comfort. Although 3D quality can be measured using subjective experiments, these are costly, time-consuming, and thus impractical for real-time applications. Consequently, objective SIQA metrics are needed to automatically predict the perceptual quality of S3D images.

According to the availability of the reference stereo pair, SIQA metrics can be generally classified into three groups: full-reference (FR) [2, 3], reduced-reference (RR) [4, 5] and no-reference (NR) [6, 7] methods. While FR-SIQA metrics use the whole reference S3D images to measure the 3D quality, RR-SIQA metrics make use of a set of features extracted form the reference images. NR-SIQA metrics measure the image quality without using any specific information of the reference images. Our SIQA metric presented in this paper belongs to the FR group.

Meanwhile, FR-SIQA methods can also be categorized into three classes [2] based on the type and amount of information extracted from stereo pairs. The SIQA methods of the first class [8, 9] employ 2D IQA metrics to measure the quality of left and right views separately, and then combine both scores into an overall 3D quality score. This class of methods does not correlate well with human quality judgments, since 2D metrics do not take into account binocular depth cues playing a critical role in 3D perception. Methods of the second class [10, 11] assess the 3D quality using depth/disparity information in addition to both views of a stereo pair. It is worth noting that the performance of the methods in this class depends on the accuracy of the depth/disparity maps estimated by stereo matching algorithms.

In fact, the left and right views of a stereo pair may suffer from the same distortion type and level (namely symmetric distortion) or different distortion levels and/or types (namely asymmetric distortion). Symmetric distortions lead to binocular fusion (BF), whereas asymmetric distortions result in either binocular rivalry (BR) [12] or binocular suppression (BS) [13] according to the difference

of distortion strength. These latter have a great impact on the perceptual 3D quality. The FR-SIQA methods of the two above-described classes can perform quite well in the case of symmetric distortion, but are much less effective for asymmetrically distorted S3D images that are very common in real application such as 3D coding. Thus, to improve the SIQA performance for asymmetric distortions, a 3D metric should accurately model the stimulus strength and account for the binocular combination. The third class of SIQA methods takes into account the monocular and/or binocular visual properties in addition to 2D image quality and disparity/depth information. Several SIQA methods that simulate the binocular visual phenomena have been proposed. These methods assess the quality of a single-view separately, and then combine both quality scores into a 3D quality with the help of weights modeling stimulus strength. For instance, Wang *et al.* [3] proposed an information content and distortion weighted SSIM metric for left and right views, and employed a BR inspired multi-scale model to predict the perceived 3D quality from the 2D images based on image local variance. Recently, Cao *et al.* [14] developed a FR-SIQA method based on several visual characteristics of the human visual system (HVS). The patch-based image gradient entropy was used for modeling the stimulus strength. When a stereo pair is observed by a human subject, the HVS merges both views of the stereo pair to yield a single mental view (namely, cyclopean perceptual image) according to the binocular combination behavior [15]. The cyclopean perceptual image can be used to model BF and BR properties. Therefore, based on different binocular combination strategies, many other SIQA approaches [2, 16–19] in the literature combine left and right views into one cyclopean image, and the final 3D quality is evaluated by analyzing this merged image. For example, Chen *et al.* [17] developed a metric by assessing the quality of the cyclopean images constructed by a linear model. The weights of this model are derived from Gabor filter magnitude responses, which simulate the BR. Recently, Zhang and Chandler [2] presented a FR-SIQA metric based on monocular image quality estimated from left and right views, and cyclopean image quality measured using lightness distance and pixel-based contrast. Although these methods achieve much progress, various characteristics of the HVS have not been deeply explored, which limit the prediction accuracy. Therefore, to design more reliable and accurate SIQA metrics, it is important to understand and account for different perceptual processes of the HVS.

In this paper, we propose a new SIQA method based on binocular combination properties and disparity information, combining quality scores of the cyclopean image and the disparity map. Specifically, the major contribution of this work lies in the development of a 3D perceptual quality prediction framework by modeling the BF/BR phenomena, and accounting for disparity distortion as well as monocular visual saliency in the binocular combination. Besides, we provide a comprehensive experimental evaluation for our proposed method and an extensive comparison with other SIQA methods on three databases. The rest of this paper is organized as follows. In Section 10.2, we detail the proposed SIQA method. Section 10.3 gives experimental results and comparative analysis. Finally, we conclude this paper in Section 10.4.

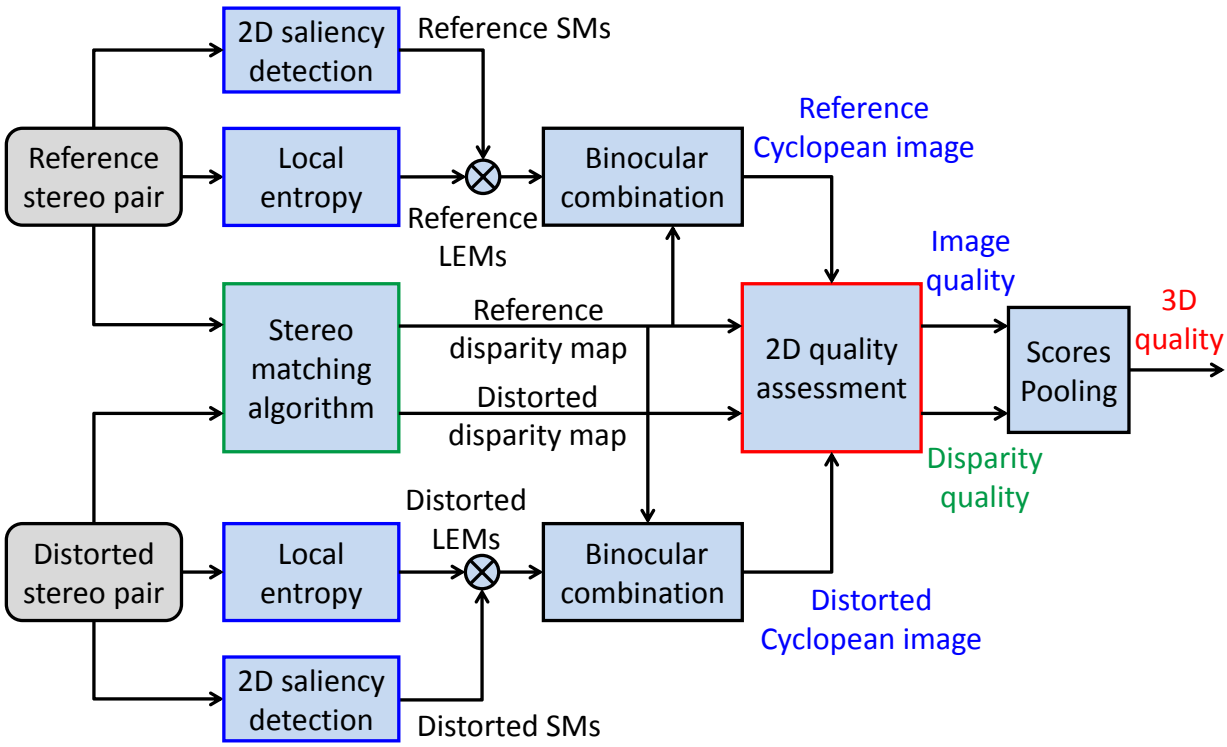


Figure 10.1 – Flowchart of the proposed SIQA method.

10.2 Proposed SIQA method

As mentioned previously, the HVS does not account for left and right stimuli separately. Instead, it perceives distortions of the cyclopean image as 2D impairments, and depth/disparity distortion as 3D ones. Inspired by this, our proposed SIQA method predicts the overall 3D quality by combining the cyclopean image quality with disparity quality. Fig.10.1 illustrates the flowchart of the proposed method.

As shown in Fig.10.1 , the first step is to determine the disparity images for reference and distorted stereo-pairs. To achieve this, we use the stereo matching algorithm proposed in [20] for low-resolution S3D images and SSIM-based stereo algorithm [17] for high-resolution S3D images. These two algorithms can efficiently achieve good performance in disparity estimation and deal with the issue of occlusion and depth discontinuities. Next, inspired by the linear combination model proposed in [17, 21], by modeling the BF/BR phenomena when a stereo stimulus is presented, we generate the synthesized cyclopean image I_c as follows:

$$I_c(i,j) = \frac{LE_l(i,j+d_r)}{LE_T(i,j)} \times I_l(i,j+d_r) + \frac{LE_r(i,j)}{LE_T(i,j)} \times I_r(i,j), \tag{10.1}$$

$$LE_T(i,j) = LE_l(i,j+d_r) + LE_r(i,j), \tag{10.2}$$

where I_l and I_r represent the left and right views respectively. LE_l and LE_r denote the local energy maps for their corresponding images, and used to describe the stimuli to left and right eyes. In addition, (i, j) is the pixel coordinate. The left image I_l and its local energy map LE_l are warped to their corresponding locations in the right view using the disparity of right image d_r that corresponds to the horizontal shift of the pixel from the right to the left view. As shown in Eq. 10.1, the BR phenomenon is correlated to the relative stimulus strength of each view instead of the absolute stimulus strength [17].

The next step is to compute the local energy maps of two views to model the stimulus strength. The study in [22] found that the 3D human perception is dominated by the view of high contrast or rich contours. In other words, the perceptual 3D quality follows the quality of the view containing a higher amount of information. Therefore, the local entropy is used to determine the stimulus strength of each view. Moreover, we assume that the local energy of one view depends on the visual importance of the stimulus corresponding to the 2D visual saliency. The local energy $LE(i, j)$ of one view is defined by:

$$LE(i, j) = (EN(i, j) \times VS(i, j))^2, \tag{10.3}$$

where $EN(i, j)$ denotes the local entropy of a pixel (i, j) in one view of a stereo pair, and VS is the visual saliency map of this view. On the one hand, we use the method proposed in [23] to estimate the saliency map, because it performs well in terms of saliency prediction accuracy and computational efficiency. On the other hand, the image entropy is related to the amount of information that can be coded in the compression process. For example, a low entropy image contains very little contrast. The local entropy of a pixel computed based on 11-by-11 neighborhood with specific shape around this pixel is described as follows:

$$EN(i, j) = -\sum_{s=g_{min}}^{g_{max}} p(x_s) \times \log_2(p(x_s)), \tag{10.4}$$

where g_{min} and g_{max} are the minimum and maximum values respectively in the corresponding neighborhood pixels. $p(x_s)$ denotes the probability that the difference between two adjacent pixels is equal to s. Based on Eqs. 10.1, 10.2, 10.3, and 10.4,, the proposed SIQA metric tries to simulates the BF/BR phenomena. Specifically, different local energies in both views lead to BR, and the 3D quality of a region is more affected by the view containing higher contrast energies.

Given the cyclopean images (I_{rc}, I_{dc}) and the disparity maps (Dp_r, Dp_d) of the reference and distorted stereo pairs, we separately measure the cyclopean quality and disparity quality by using 2D IQA metric. In [11], You *et al.* found that universal image quality index (UQI) [24] performs the best for 3D quality prediction among all tested 2D IQA metrics. Furthermore, the study in [25] revealed that the visual information fidelity (VIF) [26] metric can achieve an accurate quality prediction for 2D IQA database consisting of 2D high-resolution images such as CSIQ database [27]. On the other hand, UQI metric provided the best performance for IQA on the disparity map. In fact, UQI used in

Table 10.1 – Performance of SIQA methods on LIVE 3D IQA database (phase I). Italicized entries denote 2D-based IQA, and the results of the best-performing SIQA method are highlighted in boldface.

Distortion type	Criteria	*UQI [24]*	*VIF [26]*	*GMSD [25]*	Benoit [10]	You [11]	Fezza [18]	Chen [17]	Shao [28]	Lin [19]	Proposed
WN	LCC	*0.927*	*0.930*	*0.950*	0.925	0.941	0.947	**0.955**	0.945	0.927	0.932
	SROCC	*0.926*	*0.931*	*0.943*	0.929	0.940	0.944	**0.948**	0.941	0.929	0.927
	RMSE	*6.240*	*6.103*	*5.197*	6.308	5.622	5.351	**4.963**	-	6.257	6.038
JPEG	LCC	*0.769*	*0.603*	*0.664*	0.641	0.487	0.706	0.527	0.520	0.755	**0.781**
	SROCC	*0.737*	*0.580*	*0.620*	0.603	0.439	0.657	0.521	0.495	0.716	**0.748**
	RMSE	*4.178*	*5.216*	*4.888*	5.022	5.710	4.632	5.557	-	4.291	**4.086**
JP2K	LCC	*0.944*	*0.888*	*0.933*	0.940	0.878	0.937	0.920	0.921	0.952	**0.954**
	SROCC	*0.910*	*0.902*	*0.906*	0.910	0.860	0.896	0.887	0.895	0.913	**0.915**
	RMSE	*4.270*	*5.959*	*4.676*	4.427	6.207	4.532	5.070	-	3.963	**3.868**
GB	LCC	*0.952*	***0.962***	*0.960*	0.949	0.920	0.934	0.943	0.959	0.958	0.958
	SROCC	*0.925*	*0.934*	*0.939*	0.931	0.882	0.909	0.924	**0.940**	0.933	0.926
	RMSE	*4.451*	*3.955*	***4.051***	4.571	5.680	5.173	4.813	-	4.137	4.182
FF	LCC	*0.879*	*0.862*	*0.839*	0.747	0.730	0.783	0.776	0.859	0.862	**0.891**
	SROCC	*0.833*	*0.804*	*0.791*	**0.889**	0.583	0.693	0.700	0.796	0.829	0.844
	RMSE	*5.925*	*6.306*	*6.755*	8.258	8.492	7.730	7.832	-	6.299	**5.644**
ALL	LCC	*0.943*	*0.925*	*0.944*	0.903	0.881	0.821	0.922	0.935	0.937	**0.944**
	SROCC	*0.937*	*0.920*	*0.936*	0.889	0.879	0.922	0.914	0.925	0.931	**0.940**
	RMSE	*5.478*	*6.230*	*5.404*	7.062	7.746	9.358	6.351	5.816	5.744	**5.404**

disparity quality estimation is based on comparing the structural information, and the disparity can express such information of the original images. Thereby the qualities of the cyclopean image and the disparity map are calculated as follows:

$$Q_c = UQI/VIF(I_{rc}, I_{dc}), Q_d = UQI(Dp_r, Dp_d), \tag{10.5}$$

where Q_c is the quality score of the test cyclopean image, and Q_d denotes the quality score of the disparity map. To estimate the cyclopean image quality, we use the UQI metric for LIVE 3D IQA databases (phase I [29] and phase II [17]), and the VIF metric for Waterloo-IVC 3D database (phase I) [3]. Finally, the S3D quality score Q_{3D} is calculated by a linear model:

$$Q_{3D} = \alpha \times Q_c + (1 - \alpha) \times Q_d \tag{10.6}$$

where α is the weight for adjusting the relative importance of Q_c and Q_d. In the implementation,

Table 10.2 – Performance of SIQA methods on LIVE 3D IQA database (phase II). The symbols As and S are respectively the asymmetric and symmetric distortions.

Method	LCC			SROCC			RMSE		
	S	As	All	S	As	All	S	As	All
UQI [24]	*0.941*	*0.795*	*0.864*	*0.939*	*0.755*	*0.842*	*4.213*	*6.154*	*5.677*
VIF [26]	*0.928*	*0.777*	*0.837*	*0.916*	*0.732*	*0.819*	*4.652*	*6.382*	*6.182*
GMSD [25]	*0.920*	*0.738*	*0.803*	*0.910*	*0.716*	*0.783*	*4.897*	*6.842*	*6.723*
Benoit [10]	0.921	0.746	0.764	0.910	0.732	0.748	5.712	6.976	7.281
You [11]	0.911	0.659	0.721	0.898	0.604	0.721	7.128	8.009	7.141
Fezza [18]	0.930	0.820	0.871	0.921	0.796	0.862	4.576	5.801	5.553
Chen [17]	0.935	0.870	0.902	0.923	**0.851**	0.895	4.438	4.991	4.866
Shao [28]	-	-	0.863	-	-	0.849	-	-	5.706
Proposed	**0.945**	**0.883**	**0.912**	**0.941**	0.848	**0.898**	**4.076**	**4.754**	**4.643**

we set $\alpha = 0.65$ for LIVE 3D phase I and II databases, and $\alpha = 0.6$ for Waterloo-IVC 3D phase I database. This is because the overall disparity of the stereo pairs in Waterloo-IVC 3D database is generally larger than that in LIVE 3D databases, and the disparity quality plays more important role in Waterloo-IVC 3D phase I database.

10.3 Experimental results and analysis

In this section, we evaluate the performance of the proposed and other SIQA methods on three publicly available 3D IQA databases providing subjective scores (DMOS values): LIVE 3D IQA databases (phase I [29] and phase II [17]), and the recently created Waterloo-IVC 3D database (phase I) [3]. LIVE 3D phase I database contains 20 reference stereo pairs and 365 symmetrically distorted stereo pairs, including five distortion types: additive white Gaussian noise (WN), JPEG, JPEG 2000 compression (JP2K), Gaussian blur (GB), and fast fading (FF). The LIVE 3D phase II database is composed of 8 reference stereo pairs and 360 symmetrically and asymmetrically distorted stereo pairs corresponding to the same distortion types. Waterloo-IVC 3D phase I database consists of 6 reference stereo pairs and 330 distorted stereo pairs with symmetric and asymmetric distortion levels and types including WN, JPEG and GB. The image resolution per view is 640×360 in LIVE 3D databases , and 1920×1080 in Waterloo-IVC 3D phase I database.

We compare the proposed method with other representative FR-3D-IQA methods [10, 11, 17–19, 28]. Besides, we further explored the performance of SIQA methods using only 2D IQA metrics including UQI [24], VIF [26], GMSD [25] and FSIM [30]. We choose these metrics in this paper since they yield promising results on 3D databases compared to other 2D IQA metrics. The performance of the SIQA metrics has been evaluated using three well-known measures: the Linear Correlation Coefficient (LCC), the Spearman Rank Order Correlation Coefficient (SROCC) and Root-Mean-Square Error (RMSE). The larger LCC and SROCC values, and the smaller RMSE value indicate better

Table 10.3 – Performance of SIQA methods on Waterloo-IVC 3D database (phase I).

Method	LCC			SROCC		
	S	As	All	S	As	All
UQI [24]	*0.814*	*0.724*	*0.753*	*0.635*	*0.631*	*0.640*
VIF [26]	*0.918*	***0.788***	*0.839*	***0.914***	***0.755***	***0.801***
FSIM [30]	*0.839*	*0.668*	*0.767*	*0.918*	*0.625*	*0.704*
Benoit [10]	0.850	0.697	0.680	0.728	0.577	0.585
You [11]	0.868	0.709	0.713	0.752	0.571	0.600
Fezza [18]	0.881	0.611	0.692	0.782	0.484	0.553
Chen [17]	0.837	0.536	0.657	0.649	0.496	0.382
Wang [32]	0.833	0.609	0.677	0.683	0.500	0.552
Proposed	**0.949**	0.774	**0.841**	**0.914**	0.729	0.790

performance in terms of correlation with human opinion. The three performance measures were computed between DMOS and the predicted scores after a non-linear regression with a five-parameter logistic function described in [31].

Table 10.1 shows the performance of SIQA methods on LIVE 3D IQA phase I database. Overall, the proposed method outperforms all the other 2D/3D IQA methods. Lin's [19] and Chen's [17] methods achieve better performance than Benoit's [10] and You's [11] methods thanks to consideration of the binocular vision properties. However, Lin's and Chen's methods are slower than the proposed method due to using 2D Gabor filter in their methods. Interestingly, all 2D-based IQA metrics perform quite well on the symmetrically distorted databases, and GMSD and UQI metrics perform even better than certain 3D IQA methods. Specifically, we also examine the performance of the SIQA metrics on each individual distortion type. As shown in Table 10.1, the proposed SIQA method provides better predictions on most distortion types in comparison with other methods except for white noise and gaussian blur. However, the obtained performance for the latter distortions remain competitive and in a very acceptable level. For the WN distortion, Chen's method [17] performs the best since the MS-SSIM metric used in this method can yield a high prediction for WN distorted images. This observation indicates that the performances of some SIQA methods highly depend on the performance of the used 2D metric. Generally, all 2D-based or 3D IQA methods achieve reasonably accurate prediction results on LIVE 3D phase I database.

The quality prediction on LIVE 3D phase II database, which partially contains asymmetrically distorted stereo pairs, is more challenging than on LIVE 3D phase I database. For each SIQA method, Table 10.2 shows the overall performance and the performance on separate subsets of symmetrically and asymmetrically S3D images in LIVE 3D phase II database. These results demonstrate that the proposed SIQA method delivers the best performance compared to the others methods. Moreover, the proposed method is particularly effective for asymmetric distortions. As expected, 2D-based

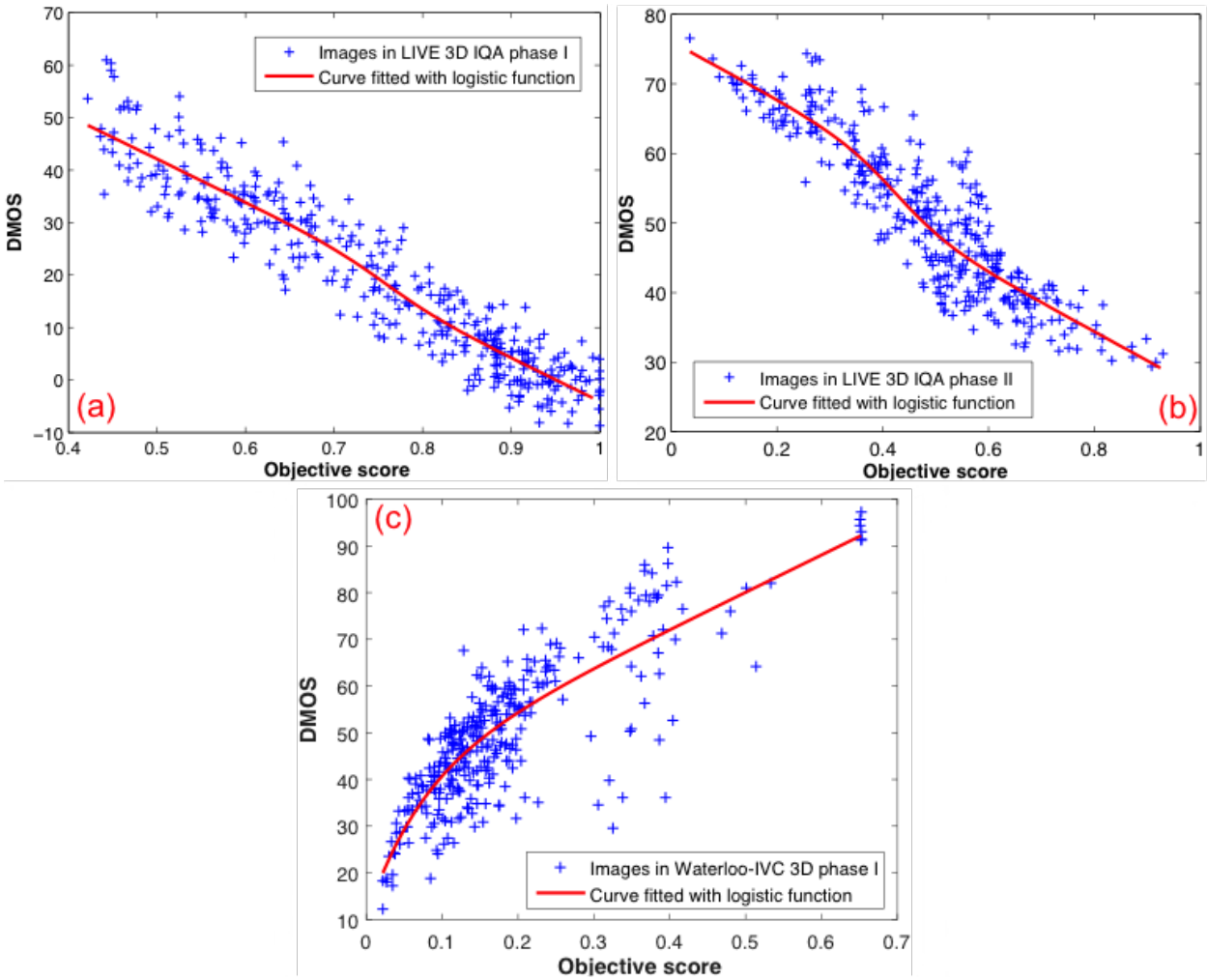


Figure 10.2 – Scatter distribution of predicted scores obtained by proposed SIQA metric versus DMOS for three databases, (a) LIVE 3D phase I, (b) LIVE 3D phase II, (c) Waterloo-IVC 3D phase I.

IQA methods achieve high performance for the symmetric distortions, but they generally perform worse than most 3D IQA methods for asymmetric distortions. This is mainly due to the fact that 2D-based SIQA methods assess the perceptual 3D quality considering neither the depth/disparity information nor the binocular vision characteristics. It is worth noting that UQI-based SIQA method performs best within all 2D-based SIQA methods. Despite the consideration of the disparity distortion, the performance of Benoit's [10] and You's [11] metrics are much lower than the proposed metric, and particularly for asymmetric distortions. This is because these methods have not accounted for binocular vision properties such as BF. The methods based on cyclopean image (*i.e.*, Chen's [17], Fezza 's [18] and our proposed methods) achieve better performance than the other 3D IQA methods.

In addition to performance evaluation on LIVE 3D databases, the performance comparison between the proposed and other SIQA methods on Waterloo-IVC 3D database (phase I) is shown in Table 10.3. The proposed method performs better than other SIQA methods and ranks second (far from

the remaining methods) for asymmetrically distorted S3D image. The decrease of performance in our method is mostly due to the use of asymmetrically mixed distortion types in this database making it more challenging to be assessed by most SIQA metrics. The scatter plots of DMOS vs. our quality scores are shown on Fig. 10.2.

10.4 Conclusion

In this paper, we proposed a full-reference quality assessment method for stereoscopic images accounting for binocular combination and disparity distortion. The proposed method models the human stereo vision by fusing the left and right views to generate a cyclopean image based on local entropy and monocular visual saliency. Then, a 2D quality metric is employed to separately evaluate the quality of both the cyclopean image and disparity map derived from a stereo matching algorithm. Finally, two quality scores are combined to yield an overall 3D quality score. A extensive performance comparison of the proposed method with some 2D-based IQA and 3D QA methods is conducted on three databases. The experimental results demonstrate that the proposed method achieves better performance than other SIQA methods for most of the databases and distortions. This is less true for Waterloo 3D database (phase I) because of the use of mixed asymmetric distortions types. This latter case, will be explored in the future in order to improve the quality prediction of our metric.

Acknowledgment

This research has been funded by both the Research Council of Norway through project no. 221073 "HyPerCept – Colour and quality in higher dimensions" and the Region "Nouvelle Aquitaine".

Bibliography

[1] Che-Chun Su, Anush Krishna Moorthy, and Alan Conrad Bovik. Visual quality assessment of stereoscopic image and video: challenges, advances, and future trends. In *Visual Signal Quality Assessment*, pages 185–212. Springer, 2015. 196

[2] Yi Zhang and Damon M Chandler. 3d-mad: A full reference stereoscopic image quality estimator based on binocular lightness and contrast perception. *IEEE Transactions on Image Processing*, 24(11):3810–3825, 2015. 196, 197

[3] Jiheng Wang, Abdul Rehman, Kai Zeng, Shiqi Wang, and Zhou Wang. Quality prediction of asymmetrically distorted stereoscopic 3d images. *IEEE Trans. Image Process.*, 24(11):3400–3414, 2015. 196, 197, 200, 201

[4] Feng Qi, Debin Zhao, and Wen Gao. Reduced reference stereoscopic image quality assessment based on binocular perceptual information. *IEEE Transactions on Multimedia*, 17(12):2338–2344, 2015. 196

[5] Lin Ma, Xu Wang, Qiong Liu, and King Ngi Ngan. Reorganized dct-based image representation for reduced reference stereoscopic image quality assessment. *Neurocomputing*, 215:21–31, 2016. 196

[6] Wei Zhang, Chenfei Qu, Lin Ma, Jingwei Guan, and Rui Huang. Learning structure of stereoscopic image for no-reference quality assessment with convolutional neural network. *Pattern Recognition*, 59:176–187, 2016. 196

[7] Feng Shao, Weijun Tian, Weisi Lin, Gangyi Jiang, and Qionghai Dai. Toward a blind deep quality evaluator for stereoscopic images based on monocular and binocular interactions. *IEEE Transactions on Image Processing*, 25(5):2059–2074, 2016. 196

[8] Patrizio Campisi, Patrick Le Callet, and Enrico Marini. Stereoscopic images quality assessment. In *Signal Processing Conference, 2007 15th European*, pages 2110–2114. IEEE, 2007. 196

[9] Jiachen Yang, Chunping Hou, Yuan Zhou, Zhuoyun Zhang, and Jichang Guo. Objective quality assessment method of stereo images. In *3DTV Conference: The True Vision-Capture, Transmission and Display of 3D Video*, pages 1–4. IEEE, 2009. 196

[10] Alexandre Benoit, Patrick Le Callet, Patrizio Campisi, and Romain Cousseau. Quality assessment of stereoscopic images. *EURASIP journal on image and video processing*, 2008. 196, 200, 201, 202, 203

[11] Junyong You, Liyuan Xing, Andrew Perkis, and Xu Wang. Perceptual quality assessment for stereoscopic images based on 2d image quality metrics and disparity analysis. In *Proc. of Int. Workshop on Video Process. and Quality Metrics for Consumer Electronics, Scottsdale, AZ, USA*, 2010. 196, 199, 200, 201, 202, 203

[12] Randolph Blake and Nikos K Logothetis. Visual competition. *Nature Reviews Neuroscience*, 3 (1):13–21, 2002. 196

[13] Jan Brascamp, Hansem Sohn, Sang-Hun Lee, and Randolph Blake. A monocular contribution to stimulus rivalry. *PNAS*, 110(21):8337–8344, 2013. 196

[14] Yu Cao, Wenhao Hong, and Lu Yu. Full-reference perceptual quality assessment for stereoscopic images based on primary visual processing mechanism. In *Multimedia and Expo (ICME), 2016 IEEE International Conference on*, pages 1–6. IEEE, 2016. 197

[15] Bela Julesz. Foundations of cyclopean perception. 1971. 197

[16] Aldo Maalouf and Mohamed-Chaker Larabi. Cyclop: A stereo color image quality assessment metric. In *ICASSP*, pages 1161–1164. IEEE, 2011. 197

[17] Ming-Jun Chen, Che-Chun Su, Do-Kyoung Kwon, Lawrence K Cormack, and Alan C Bovik. Full-reference quality assessment of stereopairs accounting for rivalry. *Signal Processing: Image Communication*, 28(9):1143–1155, 2013. 197, 198, 199, 200, 201, 202, 203

[18] Sid Ahmed Fezza and Mohamed-Chaker Larabi. Stereoscopic 3d image quality assessment based on cyclopean view and depth map. In *ICCE*, pages 335–339. IEEE, 2014. 197, 200, 201, 202, 203

[19] Yancong Lin, Jiachen Yang, Lu Wen, Qinggang Meng, Zhihan Lv, and Houbing Song. Quality index for stereoscopic images by jointly evaluating cyclopean amplitude and cyclopean phase. *IEEE Journal of Selected Topics in Signal Processing*, 2016. 197, 200, 201, 202

[20] Sehyung Lee, Jin Han Lee, Jongwoo Lim, and Il Hong Suh. Robust stereo matching using adaptive random walk with restart algorithm. *Image and Vision Computing*, 37:1–11, 2015. 198

[21] Jian Ding and George Sperling. A gain-control theory of binocular combination. *Proceedings of the National Academy of Sciences of the United States of America*, 103(4):1141–1146, 2006. 198

[22] Daniel V Meegan, Lew B Stelmach, and W James Tam. Unequal weighting of monocular inputs in binocular combination: implications for the compression of stereoscopic imagery. *Journal of Experimental Psychology: Applied*, 7(2):143, 2001. 199

[23] Hamed Rezazadegan Tavakoli, Esa Rahtu, and Janne Heikkilä. Fast and efficient saliency detection using sparse sampling and kernel density estimation. In *Scandinavian Conference on Image Analysis*, pages 666–675. Springer, 2011. 199

[24] Zhou Wang and Alan C Bovik. A universal image quality index. *IEEE Signal Processing Letters*, 9(3):81–84, 2002. 199, 200, 201, 202

[25] Wufeng Xue, Lei Zhang, Xuanqin Mou, and Alan C Bovik. Gradient magnitude similarity deviation: A highly efficient perceptual image quality index. *IEEE Transactions on Image Processing*, 23(2):684–695, 2014. 199, 200, 201

[26] Hamid Rahim Sheikh and Alan C Bovik. Image information and visual quality. *IEEE Trans. Image Process.*, 15(2):430–444, 2006. 199, 200, 201, 202

[27] Eric C Larson and Damon M Chandler. Most apparent distortion: full-reference image quality assessment and the role of strategy. *Journal of Electronic Imaging*, 19(1):011006–011006, 2010. 199

[28] Feng Shao, Kemeng Li, Weisi Lin, Gangyi Jiang, Mei Yu, and Qionghai Dai. Full-reference quality assessment of stereoscopic images by learning binocular receptive field properties. *IEEE Transactions on Image Processing*, 24(10):2971–2983, 2015. 200, 201

[29] Anush Krishna Moorthy, Che-Chun Su, Anish Mittal, and Alan Conrad Bovik. Subjective evaluation of stereoscopic image quality. *Signal Processing: Image Communication*, 28(8):870–883, 2013. 200, 201

[30] Lin Zhang, Lei Zhang, Xuanqin Mou, and David Zhang. Fsim: A feature similarity index for image quality assessment. *IEEE transactions on Image Processing*, 20(8):2378–2386, 2011. 201, 202

[31] Hamid Rahim Sheikh, Muhammad Farooq Sabir, and Alan Conrad Bovik. A statistical evaluation of recent full reference image quality assessment algorithms. *IEEE Trans. Image Process.*, 15(11): 3440–3451, 2006. 202

[32] Xu Wang, Sam Kwong, and Yun Zhang. Considering binocular spatial sensitivity in stereoscopic image quality assessment. In *VCIP*, pages 1–4. IEEE, 2011. 202

Chapter 11

Paper VI: No-Reference Quality Assessment of Stereoscopic Images based on Binocular Combination of Local Features Statistics

Y. Fan[1,2], M.-C. Larabi[1], F. A. Cheikh[2], C. Fernandez-Maloigne[1]

[1]XLIM UMR CNRS 7252, University of Poitiers, France
[2]Norwegian Colour and Visual Computing Lab, Norwegian University of Science and Technology, Gjøvik, Norway

In *IEEE International Conference on Image Processing (ICIP)*, pages 3538–3542, October 2018.

Abstract

No-reference (NR) stereoscopic 3D (S3D) image quality assessment (SIQA) is still challenging due to the poor understanding of how the human visual system (HVS) judges image quality based on binocular vision. In this paper, we propose an efficient opinion-aware NR Stereoscopic Quality predictor based on local contrast statistics combination (SQSC). Specifically, for left and right views, we first extract statistical features of the gradient magnitude (GM) and Laplacian of Gaussian (LoG) responses, describing the image local structures from different perspectives. The HVS is insensitive to low-order statistical redundancies that can be removed by LoG filtering. Hence, the monocular statistical features are then fused to derive the binocular features based on a linear combination model using LoG responses-based weightings. These weightings can efficiently simulate the binocular rivalry (BR) phe-

nomenon. Finally, the binocular features and the subjective scores were jointly employed to construct a learned regression model obtained by the support vector regression (SVR) algorithm. Experimental results on three widely used 3D IQA databases demonstrate the high prediction performance of the proposed method when compared to recent well performing SIQA methods.

Index terms– Stereoscopic 3D image quality assessment, Laplacian of Gaussian, binocular rivalry, support vector regression.

11.1 Introduction

The digital era has allowed to simplify the spread of stereoscopic three-dimensional (S3D) technologies in our daily life, providing new 3D viewing experience to consumers. In order to guarantee and improve the visual quality of S3D content, reliable and efficient methods of image quality assessment (IQA) are needed for 3D content to evaluate the performance of S3D processing algorithms/systems. Although IQA can be performed by either subjective experiments or objective metrics, the latter are more convenient to deal with real-world problems. In the recent years, 2D IQA has made a remarkable progress compared to 3D IQA. Thus some challenges still exist due to the complexity of the binocular vision [1].

Similar to 2D IQA methods, 3D IQA methods can be categorized into full-reference (FR) [2–4], reduced-reference (RR) [5–7] and no-reference (NR) or blind [8–17] methods according to the availability of the reference S3D images. This paper focuses on NR SIQA to be close to real application where the reference image is unavailable. Previous studies [8–12] for stereoscopic image quality assessment (SIQA) showed significant success based on cyclopean image, which is generated by fusing the left and right views with a binocular combination model and different weights. For instance, Chen *et al.* [8] developed a NR 3D IQA model by extracting features from the cyclopean images, the estimated disparity maps and the uncertainty maps. Recently, Zhou *et al.* [10] proposed an extreme learning machine-based Blind method based on features obtained from the cyclopean image in addition to those from the left and right views. The cyclopean-based approaches require the disparity map, which is not always available, and its estimation is often inaccurate and time-consuming. To avoid these constraints, other NR SIQA methods predicted the quality based on the binocular difference [13] or binocular similarity [14, 15]. For instance, Zhang *et al.* [13] developed a blind SIQA method by learning the primitive structures of both stereopair and difference map between left view and right view using a convolutional neural network. Zhou *et al.* [14] built their approach based on the inter- and intra-pixel binocular quality-predictive features of the local similarity maps.

Recently, some Blind SIQA models have been developed using combination of monocular 2D images features [16, 17]. For instance, Zhou and Yu [17] employed the complementary local patterns of binocular energy response and the binocular rivalry (BR) response. In this paper, we propose a new NR SIQA method based on binocular combination of monocular primitive structures, which are described by statistics of the image local contrasts. Existing studies highlighted the significance of

the Laplacian of Gaussian (LoG) response for IQA by detecting the edges where zero-crossings occur [10, 18, 19].Image gradient features play a crucial role in many 2D- [20, 21] and 3D-IQA approaches [11, 15] which show remarkable performance. Following the strategy described in [22], we first use the responses of gradient magnitude (GM) and LoG to extract the statistical features of monocular contrast. Then the monocular features are combined with different weights depending on the views so as to derive the binocular contrast features. To model the BR phenomenon [23], we propose to use the LoG maps of both views to calculate the weights of the linear combination model. This is mainly because the LoG function can model the early human visual process [24], and thus accurately model the relative stimulus strength of each view. Finally, the binocular features and the human subjective scores are jointly used to construct the learned regression model based on the support vector regression (SVR) technique [25].The performance of the proposed metric are studied in addition to a set of competitive SIQA methods on three widely used databases.

The remainder of this paper is organized as follows. Section 11.2 details the proposed approach. In Section 11.3, we analyze and discuss the experimental results on three databases. Finally, Section 11.4 concludes this paper and provides some insights about future works.

11.2 Proposed approach

To develop a reasonable and reliable NR SIQA model, one needs to consider not only the efficient perception- or distortion-relevant features, but also the binocular behavior of the human visual system (HVS) related to these features. Previous studies showed that the image structural information (e.g., edges, textures) of the image scenes are crucial for perceptual quality assessment tasks [13, 17, 26]. In other words, the HVS can detect image distortions by measuring the information in terms of image structures. The details of image structures can be captured by image derivatives corresponding to the local spatial contrast. Relying on the success of [22], we employ the image GM and LoG responses to describe the monocular structural features from different perspectives. The GM map shows the strength of local luminance variation, whereas LoG measures the local luminance contrast (e.g., image edges) after smoothing the noise.

Fig. 11.1 shows the framework of the proposed NR SIQA metric. First, the primitive structures of left and right views are independently characterized by their corresponding GM and LoG maps. In fact, the Gaussian (first and/or second order) derivative functions can model the receptive field responses of neurons along the visual pathway [24]. Here, we thus compute the GM and LoG maps using the first and second order derivatives of a circularly symmetric 2D Gaussian function G defined as follows:

$$G(x, y, \sigma) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}} , \tag{11.1}$$

where x and y denote the horizontal and vertical directions, respectively. The parameter σ is the standard deviation. Then we calculate the first order partial derivative of $G(x, y, \sigma)$ with respect to x

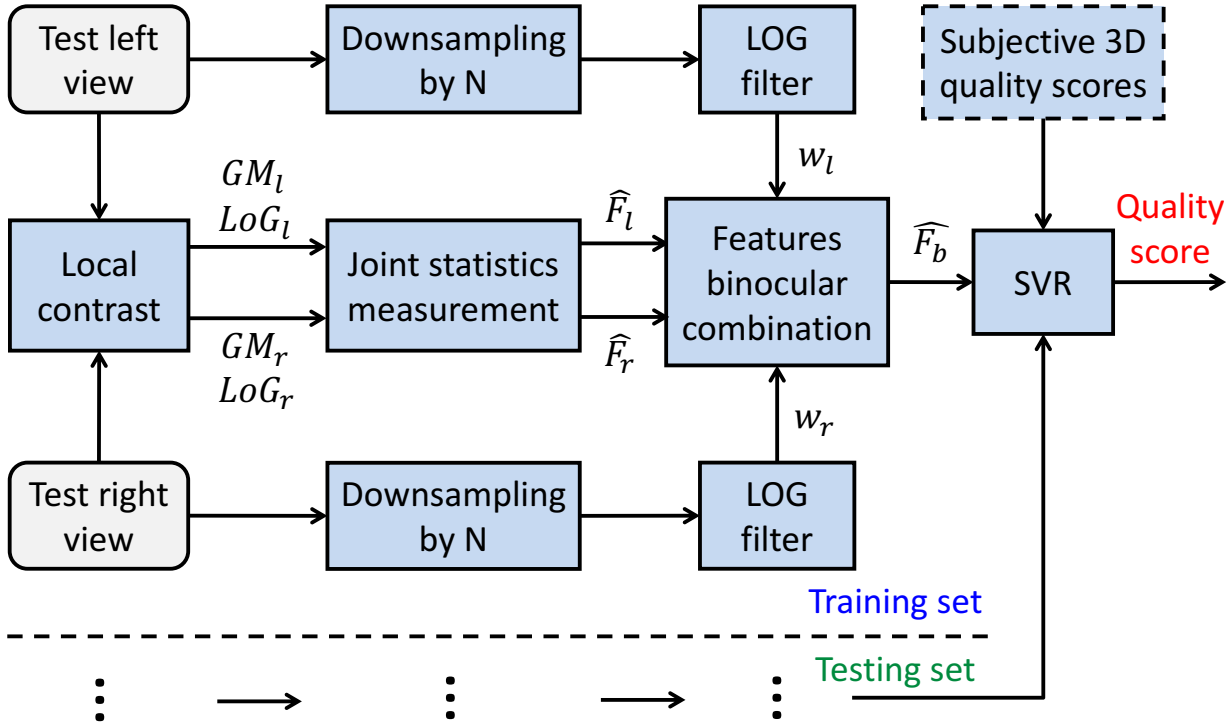


Figure 11.1 – Block Diagram of the proposed NR SIQA model

or y by

$$\frac{\partial G(x,y,\sigma)}{\partial d} = -\frac{1}{2\pi\sigma^2}\frac{d}{\sigma^2}e^{-\frac{x^2+y^2}{2\sigma^2}} \ , \ d \in \{x,y\} \ , \tag{11.2}$$

and the GM map of the image is estimated by

$$GM_v = \sqrt{(I_v \otimes \frac{\partial G}{\partial x})^2 + (I_v \otimes \frac{\partial G}{\partial y})^2} \ , \tag{11.3}$$

where the symbol $\otimes$ denotes the convolution operation. $v \in \{l,r\}$, l and r refer to left and right views, respectively. Besides, the LoG filter, corresponding to second order Gaussian partial derivative, is defined as follows:

$$\begin{aligned} h_{LoG}(x,y,\sigma) &= \frac{\partial^2 G(x,y,\sigma)}{\partial^2 x} + \frac{\partial^2 G(x,y,\sigma)}{\partial^2 y} \\ &= -\frac{1}{\pi\sigma^4}(1 - \frac{x^2+y^2}{2\sigma^2})e^{-\frac{x^2+y^2}{2\sigma^2}} \ . \end{aligned} \tag{11.4}$$

Accordingly, we estimate the LoG map of the left and right views (i.e., LoG_l and LoG_r in Fig. 11.1) by:

$$LoG_k = I_k \otimes h_{LoG} \ . \tag{11.5}$$

Subsequently, a joint adaptive normalization procedure is applied to normalize both GM and LoG

maps of each view so as to reduce the dependency of image statistics on the image content. To further extract the efficient statistical structural features, we measure the joint statistics of the normalized GM and LoG maps for each view by using both marginal and independent distributions. In particular, we derive the histogram-based feature vector (i.e., $\widehat{F_l}$ or $\widehat{F_r}$) for each view by concatenating their marginal and independent vectors.

Based on the findings of [16], we assume in our work that 3D image quality is highly related to combination of monocular features distributions. Therefore, we propose to use binocular contrast features to characterize the 3D quality distortion. The next step is to generate the binocular features for further data training. To simulate the BR phenomenon, we generate the binocular features $\widehat{F_b}$ by a linear combination model as follows:

$$\widehat{F_b} = w_l \cdot \widehat{F_l} + w_r \cdot \widehat{F_r} \ , \tag{11.6}$$

where

$$w_v = \frac{LoG_v^2}{LoG_l^2 + LoG_r^2}, \ v \in \{l, r\} \ , \tag{11.7}$$

is the weighting coefficient of the left or right views measuring the relative stimulus strength of each view. It is usually characterized by the image local energy feature such as Gabor filter magnitude response [8, 27], local entropy [28], and local variance [3]. Here, we utilize the LoG response to describe the stimulus strength, because the LoG filter can avoid the statistical redundancy in natural scene to accurately model the binocular combination. Indeed, the LoG response is similar to the receptive fields of the ganglion cells in the retina and neurons in the lateral geniculate nucleus (LGN) [24]. Besides, Wang *et al.* [29] concluded that the image scale influences the performance of the IQA model, and the first- and second- ranked performances were given respectively by scale 2 and 3. Consequently, we apply the LoG filter (see Eq. 11.4) in a single-scale image with an optimal scale size N according to the image resolution. The values of N for the experimental implementation and the performance evaluation with different scales will be discussed in Section 11.3.

Given a test stereopair, we obtain 40 binocular features, among which 20 features for two marginal distribution for GM and LoG responses and the others for independent distributions between GM and LoG responses. For the training stage, we adopt the SVR technique [25] to map the abovementioned 40 features to the subjective 3D quality scores. In particular, we employ in this paper ε-SVR to generate the learned regression model with the kernel radial basis function. Finally, we use the learned regression model to predict 3D quality scores for the testing stage. Compared to [16], we do not use subjective scores of monocular images in the training stage because most of the 3D databases only provide 3D scores. Furthermore, the regression model (i.e., SVR) in our method is only used once to predict the 3D scores at testing stage. Therefore, the proposed approach is more efficient in terms of computational complexity.

11.3 Experimental results

11.3.1 Database and experiment description

We evaluate the performance of the proposed Stereoscopic Quality based on Statistics Combination (SQSC) predictor and the state-of-the-art methods on three publicly available S3D image databases associating with differential mean opinion scores (DMOS): LIVE 3D IQA phase II database (LIVE 3D II) [27], and Waterloo IVC 3D phase I and II databases (Waterloo IVC I and Waterloo IVC II) [3]. LIVE 3D II consists of 8 reference stereopairs and 360 symmetrically and asymmetrically distorted S3D images, including five distortion types: additive white Gaussian noise (WN), JPEG, JPEG 2000 compression (JP2K), Gaussian blur (GBlur), and fast fading (FF). Waterloo IVC I contains 6 reference stereopairs and 324 distorted symmetric and asymmetric stereopairs with different distortion levels and types including WN, JPEG and GBlur. Waterloo IVC II is composed of 10 reference stereopairs and 450 distorted stereopairs corresponding to the same distortion types as Waterloo IVC I. Note that Waterloo IVC I and II provide DMOS of both 2D and 3D qualities, whereas LIVE 3D II only contains 3D subjective scores.

In the experimental implementation, we set the parameter σ (in Eq. 11.1) of the Gaussian function depending on the database to 0.7, 0.7 and 1.0 for LIVE 3D II, Waterloo IVC I and Waterloo IVC II, respectively. Furthermore, the downsampling size N equals to 2 for LIVE 3D II database containing small-resolution images [27], and to 4 for Waterloo IVC I and II databases consisting of high-resolution images [3]. Note that $N = 2$ for cross-database evaluation so as to fairly test across-database prediction capabilities. In addition, the parameters for ε-SVR learning (C, γ) were set as $(128, 8)$ and $(128, 16)$ for respectively LIVE 3D II and Waterloo IVC databases. Besides, to highlight the importance of modeling the BR phenomenon, we implemented our proposed method using two strategies for monocular features combination. The first strategy derives the binocular features using the LoG-based adaptive weights (AW) estimated in Eq. 11.7, whereas the second one using the fixed weights (FW) i.e., $w_l = w_r = 0.5$ of Eq. 11.6. Our first and second strategies-based NR SIQA methods are denoted as SQSC-AW and SQSC-FW respectively for further performance evaluation.

We compared the proposed SQSC-AW and SQSC-FW with eight competing SIQA methods, including two FR 3D IQA methods (Chen-FR [27] and Wang [3]), three NR 3D IQA methods (Chen-NR [8], SINQ [11] and SSQA [12]), two 2D-extended FR IQA methods (UQI [30] and VIF [31]) and one 2D-extended NR IQA method (BRISQUE [20]). For 2D-extended SIQA methods, we predicted the 3D quality by averaging the quality scores of 2D left and right views. And BRISQUE was trained using 2D subjective quality scores from the LIVE image quality assessment Release 2 [32].

The performance of each SIQA method was evaluated by three criteria: Pearson Linear Correlation Coefficient (PLCC), Spearman Rank Order Correlation Coefficient (SROCC) and Root-Mean-Square Error (RMSE). A Higher values for PLCC and SROCC, or a lower RMSE value indicate better performance in terms of correlation with human opinion. To reduce the non linearity of subjective

Table 11.1 – Overall performance of the SIQA methods on three databases. The top two performing methods for each criterion are highlighted in boldface. FW and AW refer to fixed and adaptive weighting, respectively.

Database	Criterion	UQI	VIF	BRISQUE	Chen-FR	Wang	Chen-NR	SINQ	SSQA	**SQSC (FW)**	**SQSC (AW)**
LIVE 3D II	PLCC	0.864	0.837	0.785	0.910	0.916	0.895	**0.936**	**0.932**	0.901	0.928
	SROCC	0.842	0.819	0.770	0.905	0.919	0.880	**0.931**	**0.930**	0.891	0.920
	RMSE	5.677	6.182	6.989	4.693	-	5.102	**3.959**	4.120	4.992	**4.181**
Waterloo IVC I	PLCC	0.753	0.839	0.828	0.654	0.930	0.938	**0.956**	0.948	**0.956**	**0.957**
	SROCC	0.631	0.801	0.814	0.510	0.918	0.922	0.937	0.924	**0.960**	**0.954**
	RMSE	10.360	8.571	8.832	11.910	-	5.237	4.960	5.564	**4.143**	**3.926**
Waterloo IVC II	PLCC	0.681	0.773	0.814	0.613	0.892	0.929	0.922	0.941	**0.945**	**0.956**
	SROCC	0.590	0.737	0.814	0.578	0.869	0.922	0.911	0.936	**0.937**	**0.945**
	RMSE	14.023	12.150	11.114	15.740	-	6.439	7.377	6.373	**6.194**	**5.408**

quality rating, PLCC, SROCC and RMSE were computed between DMOS and the predicted scores after a non-linear regression with a five-parameter logistic function as described in [32]. For NR 3D IQA methods, in each train-test procedure, we randomly picked 80% of all images from the database for training, and the remaining 20% for testing without overlap. To remove the performance bias, we repeated 1000 times train-test procedure, and the median values across all trials were reported as the final validation results.

11.3.2 Performance evaluation

We first evaluated the overall performance of the SIQA methods in terms of PLCC, SROCC and RMSE values on three databases. As shown in Table 11.1, both proposed models SQSC-FW and SQSC-AW outperform the other models on Waterloo IVC I and II. In addition, SQSC-AW model delivers highly competitive performance and even better than FR 3D IQA models (i.e., Chen-FR and Wang) on LIVE 3D II. Although the recently proposed SINQ and SSQA models perform better than our SQSC-AW model on LIVE 3D II, their models are relatively less efficient in terms of computational complexity due to many inputs for features extraction. Clearly, SQSC-AW model achieves better performance compared to SQSC-FW model, because the LoG-based monocular features combination in SQSC-AW effectively simulate the binocular visual phenomena.

Since the asymmetrically distorted stereopairs are more challenging for IQA task than the symmetrically distorted stereopairs, we thus tested the performance of the SIQA methods on separated symmetric and asymmetric distortions of each database, and showed the results in Table 11.2. We observe that SQSC-FW and SQSC-AW metrics significantly outperform most other SIQA metrics on Waterloo IVC I and II. For LIVE 3D II, the proposed SQSC-AW model performs quite well for asymmetrically distorted S3D images, and yields the promising results for symmetrically distorted S3D images. Overall, the SQSC-AW model achieves outstanding performance for asymmetric distortion

Table 11.2 – Performance of the SIQA methods on symmetric and asymmetric distortions of the three databases. The symbols S and As denote the symmetric and asymmetric distortions, respectively.

Method	LIVE 3D II				Waterloo IVC I				Waterloo IVC II			
	PLCC		SROCC		PLCC		SROCC		PLCC		SROCC	
	S	As	S	As	S	As	S	As	S	As	S	As
UQI	**0.941**	0.795	**0.939**	0.755	0.814	0.724	0.635	0.631	0.763	0.628	0.572	0.563
VIF	0.928	0.777	0.916	0.732	0.918	0.788	0.913	0.755	0.861	0.719	0.830	0.687
BRISQUE	0.857	0.700	0.849	0.663	0.925	0.786	0.891	0.787	0.929	0.748	0.900	0.772
Chen-FR	**0.940**	0.878	0.927	0.859	0.869	0.592	0.774	0.442	0.848	0.634	0.768	0.567
Wang	0.937	**0.898**	0.923	0.902	0.964	0.929	0.948	0.910	0.938	0.880	0.905	0.848
Chen-NR	-	-	0.918	0.834	-	-	0.934	0.907	-	-	0.944	0.899
SINQ	-	-	0.933	**0.905**	-	-	**0.967**	0.934	-	-	0.917	0.904
SSQA	-	-	**0.940**	**0.904**	-	-	0.962	0.921	-	-	**0.958**	**0.929**
SQSC-FW	0.922	0.870	0.903	0.865	**0.965**	**0.958**	**0.978**	**0.956**	**0.960**	**0.924**	0.956	0.914
SQSC-AW	0.934	**0.920**	0.921	0.902	**0.972**	**0.959**	**0.967**	**0.938**	**0.969**	**0.939**	**0.967**	**0.932**

Table 11.3 – PLCC values of the NR-SIQA methods on cross-database.

Training database	**Testing database**	Chen-NR	SINQ	SSQA	**SQSC (FW)**	**SQSC (AW)**
LIVE 3D II	Waterloo IVC I	0.461	0.521	0.659	**0.770**	**0.729**
	Waterloo IVC II	0.515	0.450	0.675	**0.763**	**0.676**
Waterloo IVC II	LIVE 3D II	0.484	0.597	**0.691**	0.592	**0.759**
	Waterloo IVC I	0.831	0.916	**0.924**	0.918	**0.920**

across the three databases.

To further demonstrate the robustness and the generalization of the proposed SIQA metric, we tested the performance by training them on one database, and testing on other databases. Table 11.3 indicates the cross-database evaluation PLCC results. Obviously, the proposed SQSC-AW and SQSC-FW models achieve the first- and second-ranked performances among all 3D NR IQA models when using LIVE 3D II database for training. Moreover, SQSC-AW model significantly outperforms most other models when using Waterloo IVC II for training. In fact, almost all models deliver relatively poor performance when using LIVE 3D II for training/testing, and Waterloo IVC I or II for testing/training. The reason is that the individual distortions types and asymmetric distortions in these two databases are totally different. In conclusion, the results of Table 11.3 validates the database independence of the proposed SQSC-AW model.

Table 11.4 – Performance of the proposed SQSC-AW method using the downsampling procedure by different factors N.

Database	N	PLCC	SROCC	RMSE
LIVE 3D II	1	0.923	0.916	4.375
	2	0.928	**0.920**	**4.181**
	4	**0.932**	0.919	4.213
Waterloo IVC I	1	0.934	0.929	4.798
	2	0.952	0.940	4.107
	4	**0.957**	**0.954**	**3.426**
Waterloo IVC II	1	0.933	0.919	6.587
	2	0.950	0.941	5.749
	4	**0.956**	**0.945**	**5.408**

Besides, to investigate the influence of the image downsampling factor N of the proposed framework (see Fig. 11.1) on the prediction accuracy, we evaluated the prediction performance with different downsampling factors N and listed the results in Table 11.4. Note that $N = 1$ represents original resolution of the image. We observe that SQSC-AW model yields the best performance in the case of $N = 2$ for LIVE 3D II database, and $N = 4$ for Waterloo IVC I and II databases. The results confirms our selection of different downsampling factors for these databases. The proposed model with $N = 2$ and $N = 4$ perform significantly better than the model with $N = 1$. This concludes that our SIQA model tends to supply higher quality scores with the increase of the downsampling factors.

11.4 Conclusion

In this paper, we present an opinion-aware NR SIQA method using the binocular histogram-based features from joint statistics of GM and LoG responses. The SVR algorithm with radial basis function kernel is then used to learn the prediction model, and finally to predict the perceived 3D quality. This work includes three main contributions. First, unlike previous SIQA methods using Gabor filter magnitude [8] or local entropy [28] or local variance [3], we employ the LoG responses-based local contrast on left and right views to estimate the weights of the binocular combination model so as to model the binocular rivalry phenomenon. Second, to reduce the computational complexity, we only use the binocular features derived from monocular statistical features combination for the training and testing processes. Third, we provide an extensive experimental evaluation showing high correlation with human opinion scores of our method on three widely used databases. Future work will be focused on two aspects. First, multi-scale LoG- and GM-based features will be considered to improve the performance of our method because of the usefulness of multiple scales for IQA [29]. Besides, we plan to develop an opinion-unaware NR SIQA metric based on LoG responses for real industrial application.

Acknowledgement

This research is funded by Research Council of Norway, project no. 221073 "HyPerCept", Region "Nouvelle Aquitaine" and CPER NUMERIC - e-Creation - Cinema.

Bibliography

[1] Che-Chun Su, Anush Krishna Moorthy, and Alan Conrad Bovik. Visual quality assessment of stereoscopic image and video: challenges, advances, and future trends. In *Visual Signal Quality Assessment*, pages 185–212. Springer, 2015. 210

[2] Yi Zhang and Damon M Chandler. 3d-mad: A full reference stereoscopic image quality estimator based on binocular lightness and contrast perception. *IEEE Trans. Image Process.*, 24(11):3810–3825, 2015. 210

[3] Jiheng Wang, Abdul Rehman, Kai Zeng, Shiqi Wang, and Zhou Wang. Quality prediction of asymmetrically distorted stereoscopic 3d images. *IEEE Trans. Image Process.*, 24(11):3400–3414, 2015. 210, 213, 214, 217

[4] Jian Ma, Ping An, Liquan Shen, and Kai Li. Full-reference quality assessment of stereoscopic images by learning binocular visual properties. *Appl. Opt.*, 56(29):8291–8302, 2017. 210

[5] Feng Qi, Debin Zhao, and Wen Gao. Reduced reference stereoscopic image quality assessment based on binocular perceptual information. *IEEE Trans. Multimedia*, 17(12):2338–2344, 2015. 210

[6] Lin Ma, Xu Wang, Qiong Liu, and King Ngi Ngan. Reorganized dct-based image representation for reduced reference stereoscopic image quality assessment. *Neurocomputing*, 215:21–31, 2016. 210

[7] Jiachen Yang, Bin Jiang, Yafang Wang, Wen Lu, and Qinggang Meng. Sparse representation based stereoscopic image quality assessment accounting for perceptual cognitive process. *Inform. Sci.*, 430:1–16, 2018. 210

[8] Ming-Jun Chen, Lawrence K Cormack, and Alan C Bovik. No-reference quality assessment of natural stereopairs. *IEEE Trans. Image Process.*, 22(9):3379–3391, 2013. 210, 213, 214, 217

[9] Feng Shao, Weijun Tian, Weisi Lin, Gangyi Jiang, and Qionghai Dai. Toward a blind deep quality evaluator for stereoscopic images based on monocular and binocular interactions. *IEEE Trans. Image Process.*, 25(5):2059–2074, 2016. 210

[10] Wujie Zhou, Lu Yu, Yang Zhou, Weiwei Qiu, Ming-Wei Wu, and Ting Luo. Blind quality estimator for 3d images based on binocular combination and extreme learning machine. *Pattern Recognit.*, 71:207–217, 2017. 210, 211

[11] Lixiong Liu, Bao Liu, Che-Chun Su, Hua Huang, and Alan Conrad Bovik. Binocular spatial activity and reverse saliency driven no-reference stereopair quality assessment. *Signal Process., Image Commun.*, 58:287–299, 2017. 210, 211, 214

[12] Lixiong Liu, Bing Yang, and Hua Huang. No-reference stereopair quality assessment based on singular value decomposition. *Neurocomputing*, 275:1823–1835, 2018. 210, 214

[13] Wei Zhang, Chenfei Qu, Lin Ma, Jingwei Guan, and Rui Huang. Learning structure of stereoscopic image for no-reference quality assessment with convolutional neural network. *Pattern Recognit.*, 59:176–187, 2016. 210, 211

[14] Wujie Zhou, Shuangshuang Zhang, Ting Pan, Lu Yu, Weiwei Qiu, Yang Zhou, and Ting Luo. Blind 3d image quality assessment based on self-similarity of binocular features. *Neurocomputing*, 224:128–134, 2017. 210

[15] Yaqi Lv, Mei Yu, Gangyi Jiang, Feng Shao, Zongju Peng, and Fen Chen. No-reference stereoscopic image quality assessment using binocular self-similarity and deep neural network. *Signal Process., Image Commun.*, 47:346–357, 2016. 210, 211

[16] Feng Shao, Kemeng Li, Weisi Lin, Gangyi Jiang, and Mei Yu. Using binocular feature combination for blind quality assessment of stereoscopic images. *IEEE Signal Process. Lett.*, 22(10):1548–1551, 2015. 210, 213

[17] Wujie Zhou and Lu Yu. Binocular responses for no-reference 3d image quality assessment. *IEEE Trans. Multimedia*, 18(6):1077–1084, 2016. 210, 211

[18] Min Zhang, Xuanqin Mou, and Lei Zhang. Non-shift edge based ratio (nser): An image quality assessment metric based on early vision features. *IEEE Signal Process. Lett.*, 18(5):315–318, 2011. 211

[19] Congmin Chen and Xuanqin Mou. A reduced-reference image quality assessment model based on joint-distribution of neighboring log signals. *Electronic Imaging*, 2016(18):1–8, 2016. 211

[20] Anish Mittal, Anush Krishna Moorthy, and Alan Conrad Bovik. No-reference image quality assessment in the spatial domain. *IEEE Trans. Image Process.*, 21(12):4695–4708, 2012. 211, 214

[21] Lixiong Liu, Yi Hua, Qingjie Zhao, Hua Huang, and Alan Conrad Bovik. Blind image quality assessment by relative gradient statistics and adaboosting neural network. *Signal Process., Image Commun.*, 40:1–15, 2016. 211

[22] Wufeng Xue, Xuanqin Mou, Lei Zhang, Alan C Bovik, and Xiangchu Feng. Blind image quality assessment using joint statistics of gradient magnitude and laplacian features. *IEEE Trans. Image Process.*, 23(11):4850–4862, 2014. 211

[23] Randolph Blake and Nikos K Logothetis. Visual competition. *Nat. Rev. Neurosci.*, 3(1):13–21, 2002. 211

[24] Bruno A Olshausen and David J Field. Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature*, 381(6583):607, 1996. 211, 213

[25] Alex J Smola and Bernhard Schölkopf. A tutorial on support vector regression. *Statist. Comput.*, 14(3):199–222, 2004. 211, 213

[26] Qiaohong Li, Weisi Lin, Jingtao Xu, and Yuming Fang. Blind image quality assessment using statistical structural and luminance features. *IEEE Trans. Multimedia*, 18(12):2457–2469, 2016. 211

[27] Ming-Jun Chen, Che-Chun Su, Do-Kyoung Kwon, Lawrence K Cormack, and Alan C Bovik. Full-reference quality assessment of stereopairs accounting for rivalry. *Signal Process., Image Commun.*, 28(9):1143–1155, 2013. 213, 214

[28] Sid Ahmed Fezza, Aladine Chetouani, and Mohamed-Chaker Larabi. Using distortion and asymmetry determination for blind stereoscopic image quality assessment strategy. *J. Vis. Commun. Image Represent.*, 49:115–128, 2017. 213, 217

[29] Zhou Wang, Eero P Simoncelli, and Alan C Bovik. Multiscale structural similarity for image quality assessment. In *Proc. Conf. Rec. 37th Asilomar Conf. Signals, Syst. Comput.*, volume 2, pages 1398–1402. IEEE, 2003. 213, 217

[30] Zhou Wang and Alan C Bovik. A universal image quality index. *IEEE Signal Process. Lett.*, 9 (3):81–84, 2002. 214

[31] Hamid Rahim Sheikh and Alan C Bovik. Image information and visual quality. *IEEE Trans. Image Process.*, 15(2):430–444, 2006. 214

[32] Hamid Rahim Sheikh, Muhammad Farooq Sabir, and Alan Conrad Bovik. A statistical evaluation of recent full reference image quality assessment algorithms. *IEEE Trans. Image Process.*, 15(11): 3440–3451, 2006. 214, 215

Chapter 12

Paper VII: Stereoscopic Image Quality Assessment based on Monocular and Binocular Visual Properties

Y. Fan[1,2], M.-C. Larabi[1], F. A. Cheikh[2], C. Fernandez-Maloigne[1]

[1]XLIM UMR CNRS 7252, University of Poitiers, France
[2]Norwegian Colour and Visual Computing Lab, Norwegian University of Science and Technology, Gjøvik, Norway

Submitted in *Journal of Visual Communication and Image Representation*, 2019.

Abstract

Stereoscopic/3D image quality assessment (SIQA) is still challenging to efficiently predict the 3D quality consisting with human judgments due to poor understanding of how to effectively model the binocular behavior of the human visual system (HVS). In this paper, we propose a full-reference SIQA model accounting for stereopair-based monocular quality and binocular-based cyclopean quality. In particular, we assess the qualities of the left- and right- view images separately based on gradient-based 2D IQA metric, and then linearly combine the qualities of both views into a monocular quality with the weights of simulating the strength of the views dominance in binocular rivalry (BR) behavior of the HVS. These weights are determined based on the local energy and visual saliency of each view. Subsequently, we apply the gradient-based 2D metric to estimate the quality of the cyclopean

image, which is synthesized based on a binocular combination model with the BR-related weightings. Next, the cyclopean image quality of the test stereopair is modulated with a just noticeable difference (JND) map of the reference stereopair, which reveals different visual sensitivities on the image's region degradation. Finally, the overall 3D quality is computed by integrating BR-inspired monocular quality with binocular-based JND-weighted cyclopean image quality based on a pooling stage. We achieve an extensive and comprehensive performance evaluation of our model and many other state-of-the-art SIQA models on seven publicly available 3D image quality databases. The experimental results demonstrate that the proposed method outperforms many recent well-performing methods, and achieves highly competitive prediction accuracy on these databases.

Index terms– Stereoscopic/3D image quality assessment, human visual system, binocular rivalry, visual saliency, binocular combination, just noticeable difference.

12.1 Introduction

With the recent advances in hardware technologies, the immersive multimedia technologies (e.g., stereoscopic/3D, virtual and augmented reality) have made a great progress and thus been increasingly commercialized in entertainment and medical industries in order to improve the quality of experience (QoE) and simplify the human daily life. Specifically, the amount of S3D content delivered by television, cinema, games, and remote education has been significantly growing over the recent years. 3D multimedia has thus become progressively popular because of providing a realistic and immersive viewing experience to end users. However, S3D technologies development has brought some challenges and issues to 3D displays manufactures and content producers. Therefore, because of the caused visual discomfort, visual fatigue in addition to many other symptoms [1] [2], 3D-TV has not met the expected success. In order to guarantee good QoE at every stage such as acquisition, compression, storage, transmission, and display, perceptual quality assessment (QA) of stereoscopic content is crucial so as to evaluate/optimize the performance of 3D processing algorithms (e.g., compression [3–5]) or systems (e.g., 3D display [6]).

Perceived quality of stereoscopic images can be assessed through either subjective experiments or objective measures. During subjective experiments, the subjects are asked to observe 3D images and then provide their opinion scores. Although such experiments can deliver reliable and referenced results, they are unfortunately costly, time-consuming, and thus impractical for real-time applications [7]. Accordingly, reliable and efficient objective algorithms are needed to automatically assess the perceived quality of 3D images. While methods/algorithms of 2D image QA (IQA) have greatly advanced over the last decade [8–12], stereoscopic IQA (SIQA) remains in an early stage and is thus challenging [13], especially for asymmetrically distorted stereo pairs [14]. This is mainly because perceived 3D quality can be affected by both monocular and binocular factors including 2D quality, depth cues, BR and binocular suppression (BS) [15] effects, in addition to visual discomfort.

According to the availability of the reference stereo pair, SIQA methods can be usually categorized

into three classes: full-reference (FR) [16–25], reduced-reference (RR) [26–30] and no-reference (NR) [14, 31–39] methods. FR-SIQA metrics require the whole reference stereoscopic images to predict 3D quality, while RR-SIQA metrics use only a set of features extracted from reference images. Oppositely, NR-SIQA metrics evaluate 3D quality without any information about the reference images. Compared to RR and NR SIQA metrics, FR SIQA metrics are frequently used for real applications thanks to their effectiveness. Our work mainly belongs to FR SIQA category.

As known, a stereo pair (i.e., a 3D image) consists of two slightly different views (i.e., left and right images), projected separately onto the corresponding retina. When a 3D image is observed by a human subject, the human visual system (HVS) merges both views of the stereo pair to yield a single mental view, called the cyclopean perceptual image [40, 41]. The construction of this mental image takes into account different binocular phenomena such as 1) the binocular fusion (BF) when both images are of similar quality, 2) the binocular rivalry (BR) when there is an important gap between both views and 3) the binocular suppression (BS) when the difference of quality between the views is limited and below a given threshold. Because of its compactness and the nature of its content, the cyclopean image is often taken into account when designing SIQA metrics.. Meanwhile, previous studies on FR-SIQA show significant success considering the combination of 2D monocular features/qualities [17, 18, 21, 23]. As a result, our proposed SIQA system takes into account the combined qualities of 2D monocular images too.

In addition to the above-mentioned two aspects, we further consider the visual spatial sensitivity of the HVS which probably impacts the overall 3D quality [42, 43]. According to the HVS properties [44], visual sensitivities of the human eyes are likely different to diverse distortion types and levels of the image's pixels/regions [45]. Hence, it is desirable to account for the visual sensitivity of monocular and binocular visions in order to correlate highly with human quality judgments. The visual sensitivity can be determined by measuring the just noticeable difference (JND) thresholds, which reflects the maximum tolerable distortion undetectable by the HVS. With 2D- [46, 47] and/or 3D-JND models [48–50], several FR-SIQA methods have been successfully developed [19, 51–56].

Based on the aforementioned points, we propose an approach to automatically assess the perceptual 3D quality efficiently and robustly. In other words, it is reasonable to demonstrate the effectiveness of the proposed approach base on comprehensive SIQA evaluation databases. Based on our previous work [57, 58], in this paper, we propose a FR-SIQA system considering the qualities of (1) left- and right-views images (related to monocular vision) and (2) a cyclopean-view image (related to BV). Specifically, we first compute the quality of left and right image separately based on gradient magnitude similarity metric (GMSM), and then linearly combine the qualities of both views into a 2D quality with the weights modeling the related stimulus strength of each view. Next, using the GMSM metric, we estimate the quality of the cyclopean image, which is derived from a BF-inspired combination model.

In addition, the cyclopean quality of the test stereo pair is weighted with a JND map of the reference stereo pair that reveals different visual sensitivities on the image's region degradation. Finally, the

overall 3D quality score is computed by integrating 2D monocular image quality with 3D binocular-based JND-weighted cyclopean image quality. The major contributions of this work include:

1. A new SIQA framework accounting for degradation of the stereo-pair-based monocular scene and of the cyclopean-based binocular scene using different visual stimulus strength modeling methods.

2. An overview of the existing 3D IQA databases. Comprehensive experimental evaluations of the proposed SIQA system, and extensive performance comparisons between our model and state-of-the-art SIQA methods on seven publicly available 3D IQA databases.

3. Investigation of the importance of BR-inspired monocular 2D quality and cyclopean quality on overall 3D quality.

4. Study of the impacts of different JND models and strategies of simulating the strength of view dominance on 3D quality prediction accuracy.

The remainder of this paper is organized as follows. Section 12.1 provides a literature review of stereoscopic quality metrics in addition to 2D- and 3D-JND models. In Section 12.2, we describe the proposed SIQA model by giving the details of its components. Section 12.3 presents the experimental results including performance comparison between our SIQA model and state-of-the-art SIQA models on seven publicly available 3D IQA databases. We further discuss the contribution of the main components of the proposed model on the 3D quality prediction in Section 12.4. Finally, this paper ends by some conclusions in addition to future works in Section 12.5.

12.2 Related work

In this section, we briefly review the recent FR-SIQA methods, simulating BS/BR properties. Based on the type and amount of the information used from 3D images, the FR-SIQA methods are mainly divided into three categories [17]: (1) methods without disparity information, (2) methods based on 2D monocular quality and disparity information, (3) methods considering the binocular combination mechanism and other HVS characteristics.

12.2.1 Methods without disparity information

The SIQA methods of the first category employ off-the-shelf 2D IQA algorithms directly to measure distortions without using disparity information. Specifically, this category measures the quality of left and right views separately using 2D-IQA metrics and then combine both scores into an overall 3D quality score. Based on different pooling strategies of the 2D quality scores, we further divide this category of methods into two classes: (1) Stereo-pair-based methods [59–61] and (2) BR-inspired methods.

12.2.1.1 Stereo-pair-based methods

Most early approaches fuse the quality scores of left- and right views without considering BR/BS behavior that models the strength of each view. For example, Campisi *et al.* [59] assess 3D quality by computing the quality scores of stereo pair with four 2D-IQA metrics including structural similarity metric (SSIM) [62], universal image quality index (UQI) [63], C4 [64] and a RR-QA metric [65], and then pooling these scores based on three different combination strategies. Later, Gorley and Holliman [60] estimate 3D quality degradation by computing the difference of average luminance contrast between matched interest points extracted from reference and distorted stereo pairs using SIFT [66] and RANSAC [67] algorithms.

12.2.1.2 BR-inspired methods

Although the above-mentioned methods do not take any binocular depth cue (e.g., binocular disparity), they can effectively deal with symmetric (i.e., same type and amount of) distortions of the stereo pair. However, their QA performance significantly drop for asymmetrically distorted (i.e., different distortion types and/or amounts on) ones [14, 16]. The latter distortions are very common in real application such as 3D coding [3, 6, 68, 69]. In fact, previous studies demonstrated that the quality of asymmetrically blurred stereopairs is dominated by the higher quality view, and the overall 3D quality is more affected by lower quality view for blockiness-based asymmetrically distorted stereopair [68, 70].

For a stereopair, a symmetric distortion leads to binocular fusion (BF) [71], whereas an asymmetric distortion results in either BR [72] or BS [15] depending on strength of the inter-view difference. Accordingly, to improve QA performance for asymmetrically distorted stereopairs, several existing SIQA approaches linearly combine 2D quality estimates into a 3D quality score using the weights based on the strength of view dominance in binocular combination [20, 73]. For instance, Wang *et al.* [73] employ the content information and a distortion weighted SSIM metric to compute the quality score of left and right view respectively, and then derive the 3D quality scores with a BR-inspired pooling strategy based on the local variance and the HVS spatial frequency sensitivity of each view. Recently, Geng *et al.* [20] develop a SIQA metric considering BF and BR behaviors modeled by both-views image features similarity based on independent component analysis and local luminance consistency of the stereo pair.

12.2.2 Methods based on 2D monocular quality and disparity information

Previous studies showed that the receptive field responses of some neurons in the primary visual cortex are related to binocular disparity [74, 75]. Thus, the above-described SIQA methods do not correlate well with the human quality judgments due to the lack of consideration of depth/disparity information, which plays a critical role in 3D perception. Consequently, the second category of methods use both views of a stereo pair in addition to depth/disparity information to assess the 3D quality [51–56, 76].

For instance, Benoit *et al.* [42] propose a FR-SIQA algorithm that applies SSIM and C4 metrics on the left and right images independently, and then combine these 2D scores with the estimate of disparity map distortion. Later, You *et al.* [43] explore the performance of 2D-IQA metrics used in the context of 3D quality assessment using different combination strategies for single-view images and disparity map qualities. In addition, several proposed methods take the visual sensitivity account in their SIQA model design. Hwang and Wu [76] develop a 3D quality prediction model that integrates the stereopair quality with depth quality and saliency map quality. Recently, Khan and Channappayya [25] compute 3D quality scores by combining the views quality based on a 2D-saliency-modulated gradient structural similarity map, and depth quality estimated based on edge-related depth saliency.

As mentioned in Section 12.1, the visual sensitivity of monocular vision and binocular vision can be conveyed by 2D-JND and 3D-JND thresholds, respectively. The binocular JND (BJND) is one of the most used models in this context to estimate 3D-JND thresholds. Thereby, many SIQA methods have used such models to account for the importance of monocular and binocular pixels [51–56]. For example, Shao *et al.* assess 3D quality by integrating the quality of binocular regions with BJND-modulated quality of the BF and BS regions based on responses of the log-Gabor filter amplitude and phase. Similarity, Zhou *et al.* [55] propose a SIQA method by first extracting monocular and BF features modulated by 2D-JND and BJND respectively, and then computing the overall 3D score using support vector regression algorithm (SVR) [77]. It is worth noting that the performance of the methods in this category depends on the accuracy of depth/disparity maps, which are often estimated by stereo matching algorithms.

12.2.3 Methods considering binocular combination mechanism and the HVS characteristics

In addition to stereopair quality and depth/disparity information, the third category of methods account for monocular and/or binocular properties (BF and BR/BS behaviors) and other HVS characteristics (e.g., visual spatial sensitivity). To simulate the BF and BR phenomena, these SIQA metrics employ the cyclopean image (introduced in Section 12.1), which combines disparity-compensation left and right views, based on a binocular combination model with the weights mimicking strength of view dominance of a stereo pair. The final 3D quality is predicted by using the cyclopean image only or together with the stereopair. During the past three decades, researchers have deeply explored the binocular combination (BC) mechanism of the HVS, and thus proposed several physiology-inspired BC models mainly including (1) vector-summation model [78, 79], (2) neural-network model [80, 81], (3) contrast gain-control model [82–85] successfully applied in SIQA [16–18, 21, 23, 31].

12.2.3.1 Methods based on the cyclopean image

Some methods in this category only estimate the cyclopean image quality and use it as the 3D quality score. An early SIQA method is proposed by accounting for the qualities of both synthesized cyclopean

image in addition to the disparity map [41]. In the same vein, Fezza and Larabi [86] develop a SIQA model based on the quality of the test cyclopean image generated by using local entropy and depth information. Chen *et al.* [16] develop a SIQA metric by assessing the quality of the cyclopean images constructed by a linear model. The weights of this model are derived from the Gabor filter magnitude responses, which simulate the BR. Besides, Lin and Wu [87] predict the 3D quality based on both BC and binocular frequency integration.

12.2.3.2 Methods based on the cyclopean image and monocular views

In additions to BF/BR-inspired cyclopean image quality, the degradations on monocular views may affect the overall 3D quality according to [17]. Hence, to design a more efficient and robust SIQA model, it is reasonable to consider the features/qualities of both the stereo pair and the synthesized cyclopean image, in addition to HVS characteristics [17, 18, 21]. For instance, Zhang and Chandler [17] develop a FR-SIQA metric based on a combination of monocular quality estimated using the stereopair and block-based contrast, and cyclopean image quality measured using lightness distance and pixel-based contrast. Recently, Ma *et al.* [21] compute 3D quality scores by jointly considering monocular perception (related to simple cells' response) using a push-combination of receptive fields model, and binocular perception (related to complex cells' response) based on BR and binocular energy modeling.

12.3 Proposed SIQA model

To accurately predict 3D image quality, we propose in this paper a MOnocular and Binocular Impairments based QUality Metric (MOBIQUM). Specifically, the proposed metric is based on the assumption that the overall 3D quality is a combination of stereopair-based monocular quality and binocular-based cyclopean image quality. Therefore, we design MOBIQUM based on two main stages: (1) monocular quality estimations derived from the combination of views quality and (2) gradient magnitude similarity (GMSM) based quality estimations on the synthesized cyclopean image. These two quality estimates are finally fused to obtain a global 3D quality score. The following subsections describe each stage of MOBIQUM as described in Figure.12.1.

12.3.1 Monocular images quality

In the monocular QA stage, we employ a gradient-based metric to compute the quality scores of the left and right views separately. Both scores are used to estimate 3D quality score based on an adapted weighting reproducing the view dominance/importance phenomenon. At this stage, any robust 2D-IQA metric (e.g., UQI [63], VIF [88] or FSIM [89]) can be used to assess the single-view quality. In this paper, we employ a gradient magnitude based metric (GMSM [90]) because of its high prediction accuracy and efficiency on 2D IQA. The main features of this metric are described below.

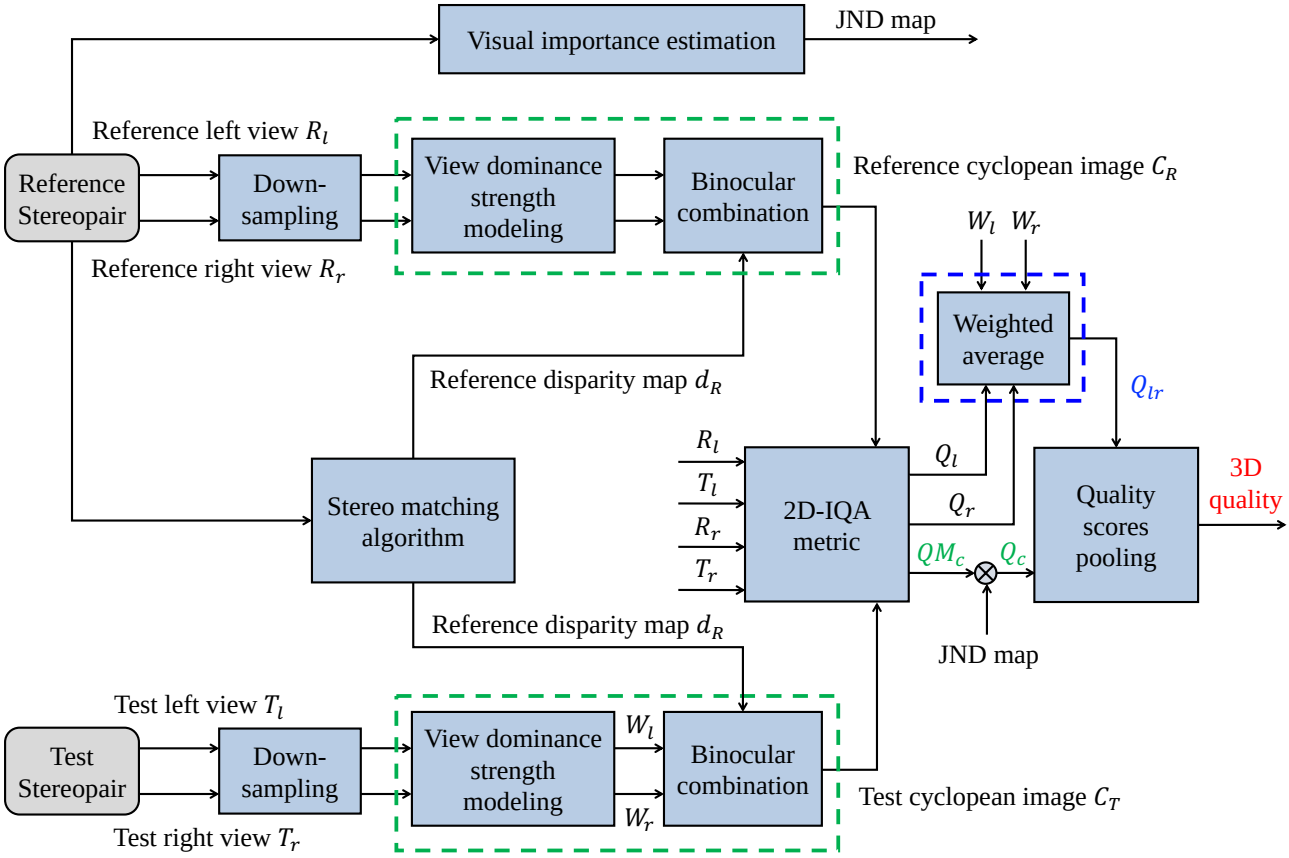


Figure 12.1 – Flowchart of the proposed SIQA model

12.3.1.1 GMSM-based QA

The image gradient information, showing the image luminance variation, can describe the image local structures (e.g., edges, textures). The HVS is highly sensitive to degradations on such structures. Hence, the image gradient feature plays a crucial role on perceptual QA tasks, and has been successfully considered in 2D- [89, 91, 92] and 3D-IQA metrics design [25, 56, 93]. Most gradient-based FR 2D-IQA algorithms use a gradient similarity (GS) map to compute the quality score. For instance, Xue *et al.* [90] developed the GMSM and a gradient magnitude similarity deviation (GMSD) IQA metrics, which compute quality scores with the average and the standard deviation pooling strategies of the GS maps respectively. We calculate the GMSM score by following three main steps. In the first step the luminance component I is extracted from the YIQ color space [94] by using the following formula:

$$I = 0.299\ r + 0.587\ g + 0.114\ b, \tag{12.1}$$

where r, g and b denote the color image's red, green and blue components respectively. Before calculating the GM of I, we iteratively smooth the luminance image I using a mean filter with a 2×2 square kernel and downsample the filtered image by a factor of N in order to choose the appropriate

scaling coefficient depending on the viewing condition [95]. The possible values of N depend on the used the dataset. This will be discussed in Section 12.4.

The second step consists in computing GM maps of the reference single-view image R and the test single-view image T by:

$$GM_s(x,y) = \sqrt{(s \otimes f_x)^2(x,y) + (s \otimes f_y)^2(x,y)},\ s \in \{R,T\}\,, \tag{12.2}$$

where $\otimes$ represents the 2D convolution operation. f_x and f_y, denote respectively the horizontal and vertical kernels of Prewitt filter [96] and are defined as follows:

$$f_x = \frac{1}{3}\begin{bmatrix} 1 & 0 & -1 \\ 1 & 0 & -1 \\ 1 & 0 & -1 \end{bmatrix},\ f_y = \frac{1}{3}\begin{bmatrix} 1 & 1 & 1 \\ 0 & 0 & 0 \\ -1 & -1 & -1 \end{bmatrix} \tag{12.3}$$

In fact, other high-pass filters (e.g., Sobel, Canny, ...) can be used to calculate the image GM. In our proposed model, the Prewitt operator is used because it can efficiently achieve high performance on test databases compared with other reference operators. The third step is to determine the GM maps difference between the reference and test images using the GS map as follows:

$$GS(x,y) = \frac{2\ GM_R(x,y)\ GM_T(x,y) + c}{GM_R^2(x,y) + GM_T^2(x,y) + c}, \tag{12.4}$$

where c is a positive constant to avoid instability and its value will be discussed in Section 12.4. Note that $GS(x,y) \in]0,1]$ and its value close to 1 indicates less distortion on pixel (x,y). Hence, GS of the left and right images (GS_l and GS_r) are determined independently. Finally, we estimate the quality score of the left and right images based on GS_l/GS_r using an average pooling as follows:

$$Q_v = \frac{1}{XY}\sum_{x=1}^{X}\sum_{y=1}^{Y} GS_v(x,y), v \in \{l,r\}\,, \tag{12.5}$$

12.3.1.2 BR-inspired combination of the stereopair qualities

The importance of the quality of the left and right views are considered the same if both views contain similar spatial content/information. In this case, we can simply use the average pooling to fuse Q_l and Q_r into a combined monocular image quality Q_{lr}. In contrast, when two views suffer from different distortion levels and/or types that result in the BR behavior of the HVS, the average pooling is not suitable to accurately model the overall 3D quality judgment of the HVS. Consequently, it is appropriate to use an adaptive pooling using the weights simulating the strength of the views quality dominance.

Next, the main question is on how to accurately estimate the weighting coefficient for each view corresponding to image local energy [97–99], which can effectively describe the response properties of

the binocular neurons in the primary visual cortex. Existing SIQA approaches generally estimate local entropy (LE) of each view based on Gabor/LoG-Gabor filter responses [16, 19, 22, 23], block-based local contrast/variance [17, 73], LoG/Difference of Gaussian (DoG) responses [33, 87, 100], image gradient features [21, 93] or image LE [57, 69, 101].

On the one hand, the previous findings in [17, 102] showed that 3D human perception or 3D quality is largely dominated by the monocular view or the quality of monocular view containing higher contrast or rich contours. Accordingly, the local spatial contrast of each view can be used to reflect their corresponding relative stimulus strength. On the other hand, our previous work [57, 58] demonstrated that the image LE can successfully mimic the stimulus strength of the binocular combination. However, LE-based weighting strategy is less efficient than LoG-based one. Therefore, to accurately model the strength of the view dominance of a stereopair, we propose to jointly use the LoG-based weighting strategy and monocular visual saliency that reflects different spatial sensitivities of the HVS on image regions. Note that LoG-, GM-, and LE-based weighting strategies will be discussed in terms of 3D quality prediction accuracy in Section 12.4.

As shown in Figure 12.1, the combined monocular quality Q_{lr} is obtained based on a linear summation model defined as follows:

$$Q_{lr} = w_l \cdot Q_l + w_r \cdot Q_r, \tag{12.6}$$

$$w_v = \frac{g_v}{g_l + g_r}, \ v \in \{l, r\}, \tag{12.7}$$

$$g_v = \sum_{x=1}^{X} \sum_{y=1}^{Y} \left[E_v(x,y) \cdot VS_v(x,y)\right]^2, \ v \in \{l, r\}, \tag{12.8}$$

where VS_v denotes the visual saliency map of the single-view image. Several saliency detection algorithms for 2D [103–107] or 3D images [108, 109] have been proposed in the recent years. Here, we apply an algorithm developed in [104] to estimate image saliency map thanks to its high detection accuracy and computational efficiency. As described previously, E_v represents LoG, GM or LE maps of the single-view image. GM map is estimated by Eq. 12.2 and 12.3, whereas the LoG map of an image is computed using second-order derivatives of a circularly symmetric 2D Gaussian function G defined as follows:

$$G(x,y,\sigma) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}}, \tag{12.9}$$

where the parameter σ is the standard deviation. Then, the LoG filter h_{LoG} is determined by:

$$\begin{aligned} h_{LoG}(x,y,\sigma) &= \frac{\partial^2 G(x,y,\sigma)}{\partial^2 x} + \frac{\partial^2 G(x,y,\sigma)}{\partial^2 y} \\ &= -\frac{1}{\pi\sigma^4}\left(1 - \frac{x^2+y^2}{2\sigma^2}\right) e^{-\frac{x^2+y^2}{2\sigma^2}}, \end{aligned} \tag{12.10}$$

and the LoG map of each view (i.e., LoG_l and LoG_r) is finally estimated by convolving the view image

with h_{LoG} as follows:

$$LoG_v(x,y) = I_v(x,y) \otimes h_{LoG}(x,y),\ v \in \{l,r\}\,. \tag{12.11}$$

In addition, the image local energy can also be estimated by quantifying the information contained in the image, which corresponds to the entropy introduced by Shannon [110]. Kapur *et al.* [111] in their early work describe how to calculate the global entropy of an image. Following this work, in order to describe the amount of spatial information, we compute the LE map instead of the global entropy value. Therefore, we define a local neighborhood Ω as a window of size $m \times n$ centered at target pixel (x,y). In this paper, a size of 11×11 with a disk shape around a corresponding pixel is used to compute the LE values. Then, the local entropy of the pixel (x,y) of the left/right view (i.e., $LE_v(x,y)$), corresponding to the LE of Ω (i.e., $En_v(\Omega)$) is expressed as:

$$LE_v(x,y) = En(\Omega_{(x,y)}) = -\sum_{k=0}^{K-1} Pr(k) \times \log_2(Pr(k)),\ v \in \{l,r\} \tag{12.12}$$

where $Pr(k)$ denotes the probability that the grey level (i.e., luminance value) k appears in Ω and is calculated by:

$$Pr(k) = \frac{n_k}{m \times n}, \tag{12.13}$$

where n_k is the number of pixels with grey level k in Ω, and K is the maximum grey level. Image regions with a low contrast may result in low LE values for these regions, because LE in the target region reflects the luminance variance in its corresponding neighborhood. Figure 12.2 illustrates a reference image, images with different distortion types and their associated LoG, LE and GM maps. For WN distortion, the artifacts are much more introduced influenced in flat areas than in texture areas when comparing the corresponding LoG, GM and LE maps between the reference and distorted images. For GB distortion, the blur artifacts reduce the fine details of flat areas and enhance the edge structures for LE and GM maps, while most of the details in Figure 12.2h are removed because of the smoothing effect in LoG response. For JPEG distortion, Figures 12.2f, 12.2j and 12.2n describe the additive information corresponding to the blocking artifacts shown in Figure 12.2b, for which the HVS is highly sensitive.

12.3.2 Binocular-based cyclopean image quality

In the second stage, we first synthesize the cyclopean image based on a binocular combination (BC) model with weights estimated using local energy map as described in Section 12.3.1.2. These local energy maps simulate strength of the view dominance on BR phenomenon. Then, the quality of the synthesized cyclopean image is assessed using the previously mentioned GM-based 2D-IQA metric (see Section 12.3.1.1). As described in Figure. 12.2.3, many physiology-inspired BC models have been proposed for the formulation of the cyclopean image. In this paper, inspired by the gain-control model proposed in [82], the cyclopean image of a stereopair is synthesized based on a linear summation model

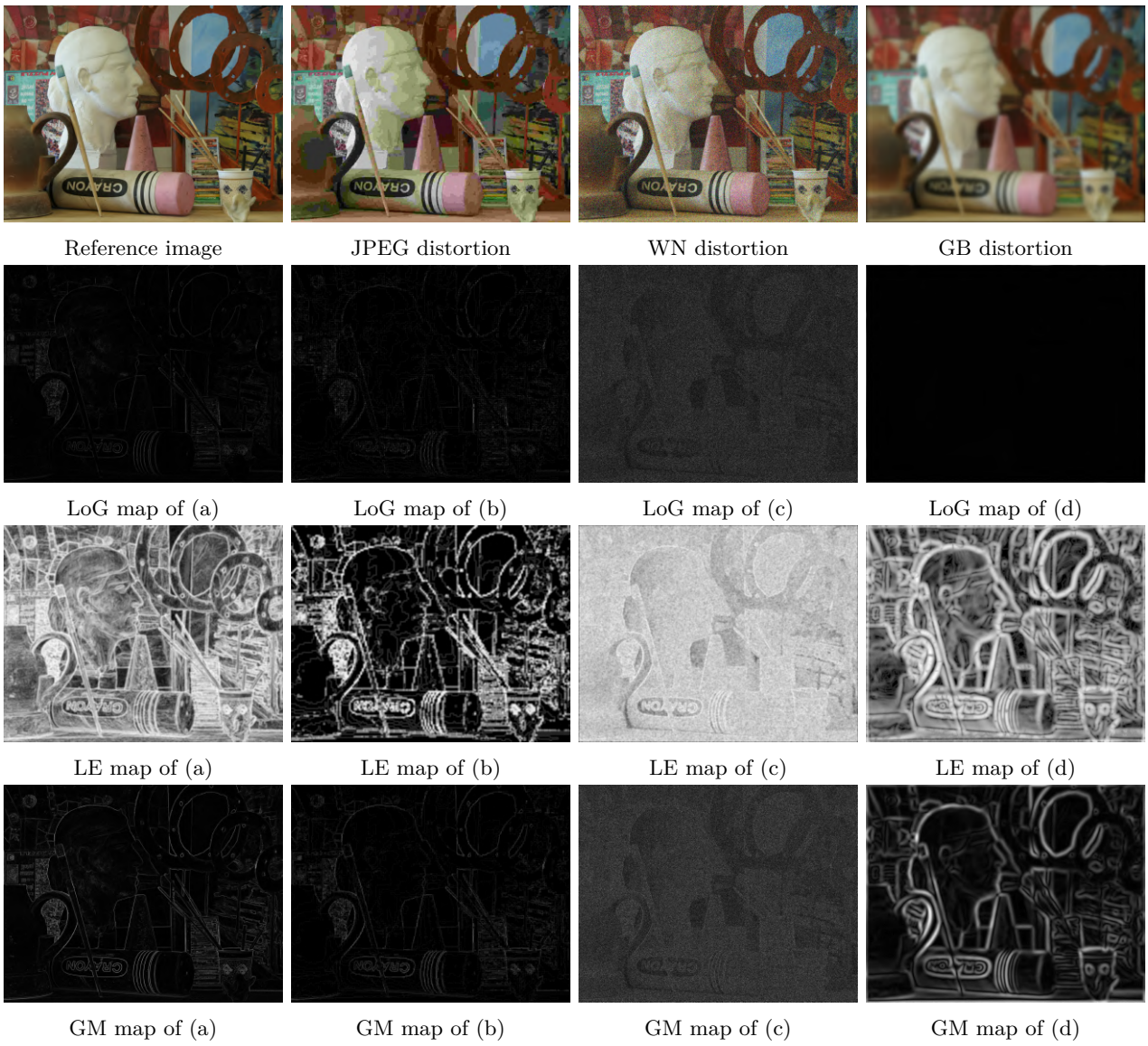


Figure 12.2 – Example of reference and distorted images from the Waterloo-IVC 3D image quality database Phase I [**?**], and their corresponding LoG, LE, and GM maps. Brighter gray level means higher value. WN, JPEG, and GB denote the different distortions respectively injected in the image including the JPEG compression, additive white noise, and Gaussian blur.

defined as follows:

$$C(x,y) = \frac{E_l(x,y)}{E_t} \times I_l(x,y) + \frac{E_r(x-d_l,y)}{E_t} \times I_r(x-d_l,y), \tag{12.14}$$

where E_l and E_r refer to local energy maps of left and right views respectively, which can be estimated by one of LoG, GM and LE maps of the single-view image (see Section 12.3.1.2). And E_t is expressed

as:

$$E_t = E_l(x, y) + E_r(x - d_l, y), \tag{12.15}$$

In addition, the right image I_r and its E_r are warped to their corresponding locations in left image I_r using the disparity map of the left image d_l that denotes the horizontal shift of the pixel from the left to right view. To estimate the disparity maps of a stereopair, we apply a stereo matching algorithm developed by Lee *et al.* [59], which can deliver high performance and deals efficiently with the issues of occlusion and depth discontinuities. As shown in Figure 12.1, to avoid the geometrical information degradation, we utilize the disparity map of the reference stereopair to determine the cyclopean images of both the reference and test stereopairs (i.e., C_R and C_T).

Next, we compute the quality map of C_T (named QM_c) using GM-based IQA metric based on Eqs. 12.2, 12.3 and 12.4. The GMSM metric does not consider the visual spatial sensitivity of the HVS that can probably affect human quality judgments. It is known that JND thresholds indicate the maximum distortion that the HVS cannot detect. Thus, JND thresholds of the reference image/stereopair, which determine the visual sensitivity, is used to weight GM_c. The JND-weighted cyclopean quality Q_c is calculated by:

$$Q_c = \frac{\sum_{x,y}^{N} \left[\frac{1}{JND(x,y)} \times QM_c(x, y) \right]}{\sum_{x,y}^{N} \frac{1}{JND(x,y)}}, \tag{12.16}$$

where N is the number of the pixels in the image. A pixel with a high JND value will tolerate a large distortion, and thus has a low visual importance on the perceived quality. To compare the performance of the proposed SIQA metric integrating JND, two efficient 2D-JND models (i.e., NAMM [46] and JND-TE [47]) and three 3D-JND models (i.e., BJND [48], SJND [49] and DJND [50]) were used in this paper.

12.3.3 Overall 3D quality

Finally, Q_c and Q_{lr} are combined to compute the overall 3D quality Q_{3D} based on a linear pooling model as follows:

$$Q_{3D} = \alpha \times Q_c + (1 - \alpha) \times Q_{lr}, \tag{12.17}$$

where α is the weight to adjust the relative importance of Q_c and Q_{lr} for Q_{3D}. The optimal value α acheving the best performance on specific 3D-IQA database will be further discussed in Section 12.4.3.

12.4 Experimental results

In this section, we first give a summary of existing stereoscopic IQ databases, and briefly describe seven publicly available ones used in experiments. Based on these databases, we then evaluate the performances of the proposed FR-SIQA model using different importance weights for combined monocular quality Q_lr and cyclopean quality Q_c in order to determine the optimal value of α and β. Next, we

Table 12.1 – Summary of stereoscopic image quality databases. **Sym** and **Asym** denote separately the asymmetrically and symmetrically stereopairs. **Mix** represents the multiple distortions in the single-view image of a stereopair. WN, JPEG, JP2K, GB, FF, DB, TR, RE indicate additive white noise, JPEG, JPEG 2000, Gaussian blur, fast fading, downsampling blur, transmission error and rendering error, respectively.

Database	Sym	Asym	Mix	depth map	Publicly availiable	2D (D)MOS	Distortion types
LIVE 3D I [112]	Yes	No	No	Yes	Yes	No	WN, JPEG, JP2K, GB, FF
LIVE 3D II [16]	Yes	Yes	No	Yes	Yes	No	WN, JPEG, JP2K, GB, FF
Waterloo-IVC I [73]	Yes	Yes	No	Yes	Yes	Yes	WN, JPEG, GB
Waterloo-IVC II [73]	Yes	Yes	No	No	Yes	Yes	WN, JPEG, GB
NBU 3D II [113]	Yes	No	No	No	Yes	No	WN, JPEG, JP2K, GB, H.264
NBU-MDSID [114]	Yes	No	Yes	No	Yes	Yes	WN, JPEG, GB
IEEE 3D [115]	Yes	No	No	No	Yes	No	WN, JPEG, JP2K, GB, FF
IRCCyN/IVC 3D [42]	Yes	No	No	No	Yes	No	JPEG, JP2K, GB
MCL-3D [116]	Yes	No	No	Yes	Yes	Yes	WN, JPEG, JP2K, GB, DB, TR, RE
MMSPG 3D IQA [117]	Yes	No	No	No	Yes	No	Inter-camera distances
MICT 3D IQA [118]	Yes	Yes	No	No	No	No	JPEG
SVBL 3D IQA [61]	Yes	No	No	No	No	No	WN, JPEG, JP2K
NBU 3D I [119]	No	Yes	No	No	No	No	WN, JPEG, JP2K, GB

compare the performance of the proposed model using different BR modeling strategies and JND models. Finally, we provide an extensive comparison between our model and competitive state-of-the-art SIQA models on seven databases. Besides, the impacts of different scales on proposed SIQA model's performance will further investigated.

12.4.1 Stereoscopic image quality databases

To the best of our knowledge, there are currently thirteen databases for SIQA from the literature [13, 73], in which ten are publicly available. The summary of 3D image databases is given in Table 12.1. Among the publicly available databases, LIVE 3D I and II, Waterloo-IVC I and II, NBU 3D II, NBU-MDSID, and IEEE 3D are used in the validation of the proposed model. Furthermore,

IRCCyN/IVC 3D, MCL-3D and MMSPG 3D IQA databases have not been considered because 1) IRCCyN/IVC 3D only contains 96 stereopairs that are not sufficient to test SIQA models performance, 2) MCL-3D contains 2D-plus-depth sources and rendering stereopairs that are probably inaccurate, 3) MMSPG contains single camera-setup-related distortion type (different setting of inter-camera distances). Above-mentioned used databases associated with the subjective scores are briefly described in the following:

- LIVE 3D image quality database consists of two phases (Phase I and II). LIVE 3D Phase I [112] contains 20 reference stereo pairs and 365 symmetrically distorted stereo pairs, including five distortion types: additive white noise (WN), JPEG, JPEG 2000 compression (JP2K), Gaussian blur (GB), and fast fading (FF). LIVE 3D Phase II [16] contains the same five distortion types as described in Phase I, 8 reference stereo pairs, 120 symmetrically and 240 asymmetrically distorted stereopairs. Symmetric/asymmetric distortion in a stereopair mean that the left and right images are degraded by the same/different distortion levels. The resolution of all single-view images is 640×360 in LIVE 3D database.

- Waterloo-IVC 3D image quality databases [73] were created in two phases (Phase I and II). Waterloo-IVC phase I contains 6 reference stereopairs, 330 distorted stereopairs with 78 symmetric and 252 asymmetric distortions, including WN, JPEG, and GB. In addition, Phase I also provides the ground-truth disparity map of each reference stereopair. Waterloo-IVC phase II contains 10 reference stereopairs, 130 symmetrically and 330 asymmetrically distorted stereopairs corresponding to the same three distortion types. In addition to different distortion levels, an asymmetrically distorted stereopair from Phase I and II can be impaired by different types of distortion. Both Phases I and II provide the quality ground truth of each single-view image. All single-view images have HD resolution (1920×1080).

- NBU 3D IQA database Phase II [113] is composed of 312 symmetrically distorted stereopairs generated from 12 reference stereopairs. Five distortion types in this database are WN, JPEG, JP2K, GB and H.264 compression, respectively. The resolutions of all single-view images range between 480×270 and 1024×768.

- NBU-MDSID database [114] consists of 10 reference stereopairs and 270 multiply distorted stereopairs with symmetric distortions including WN, JPEG, and GB. The mixed distortion refers to a stereopair impaired by at least two types of distortions. In addition to multiply-distorted stereopairs, this database also contains 90 singly-distorted stereopairs and their corresponding 2D mean opinion scores (MOS). We only used the 270 multiply distorted stereopairs for the validation because of the challenge brought for QA. The resolution of all single-view images is HD (1920×1080) in this database.

- IEEE standard association stereo image database [115] is composed of 26 reference stereopairs, and 650 symmetrically distorted stereopairs that correspond to 130 stereopairs for each distortion

type including WN, JPEG, JP2K, GB, and FF. The resolution per view is also HD (1920×1080) in this database.

12.4.2 Experimental implementation and performance evaluation

In the experimental implementation, we tested and compared the performance of the proposed SIQA algorithm using different parameters values and component types. Table 12.2 lists the optimal values of the parameters and most appropriate components for each database. The performance with different parameters and components will be further discussed in the following sections.

To fairly evaluate the correlation and difference between the predicted quality and the subjective score, i.e., MOS or differential mean opinion score (DMOS), the predicted scores are mapped on the same scales as the DMOS based on a five-parameters logistic function [120] as follows:

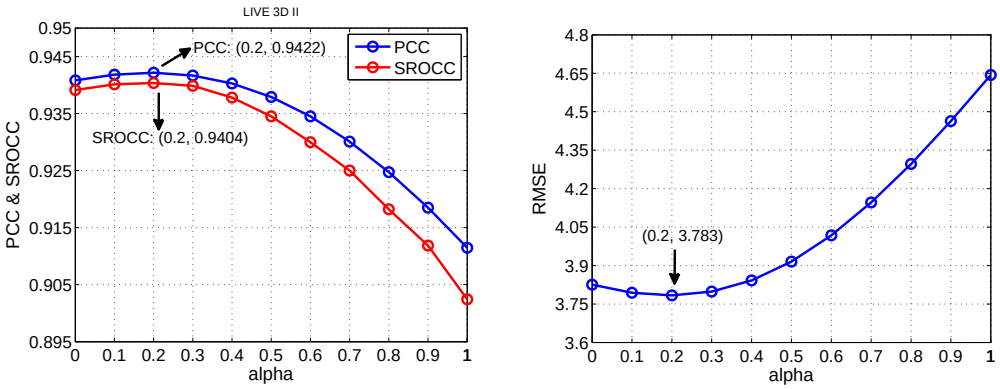
$$Q_p = p_1 \cdot \left[\frac{1}{2} - \frac{1}{e^{(p_2 \cdot (Q_{3D} - p_3))}}\right] + p_4 \cdot Q_{3D} + p_5, \tag{12.18}$$

where p_1, p_2, p_3, p_4 and p_5 are the regression parameters selected based on the subjective scores and estimated quality that is represented by Q_{3D}. Q_p denotes the predicted score after non-linear regression. Once obtaining Q_p, the performance of the proposed SIQA metric was evaluated by three statistic-based criteria: Pearson linear correlation coefficient (PCC) for the estimation of prediction accuracy, Spearman rank order correlation coefficient (SROCC) for monotonicity prediction and Root-Mean-Square Error (RMSE) for prediction consistency. Higher PCC and SROCC values, or lower RMSE values indicate better performance in terms of correlation with the human judgment of quality.

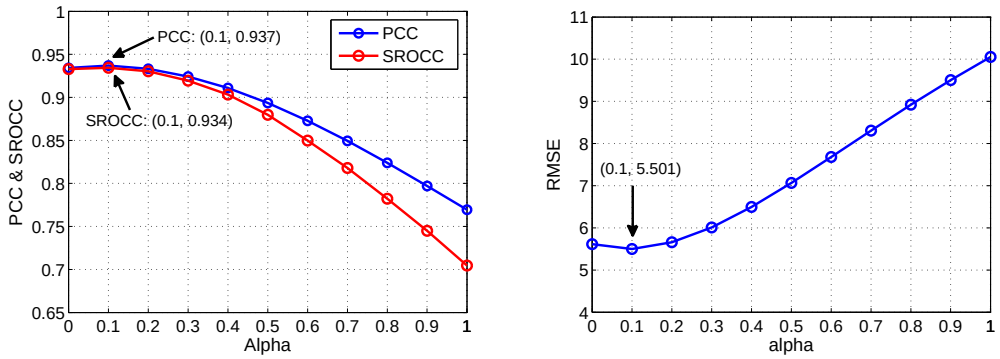
Table 12.2 – Parameter values in the implementation on seven databases. α is the weighting parameter described in Eq. 12.17. Visual saliency (VS) component, stimulus strength modeling, and JND model are related to Eq. 12.7, 12.8 and 12.16, respectively. The downsampling factor represents the image scaling with different values for view dominance strength modeling (see Figure 12.1).

Database	α	**JND model**	**VS**	**Stimulus strength**	**Downsampling factor**
LIVE 3D I	0.4	BJND [48]	Yes	LoG	1
LIVE 3D II	0.2	BJND [48]	without	LoG	2
Waterloo I	0.1	DJND [50]	without	LoG	1, 2 and 4
Waterloo II	0	N/A	with	GM	2 and 4
NBU II	0.2	NAMM [46]	with	LoG	4
NBU-MDSID	0.2	BJND [48]	with	LoG	4
IEEE 3D	0.1	BJND [48]	with	LoG	4

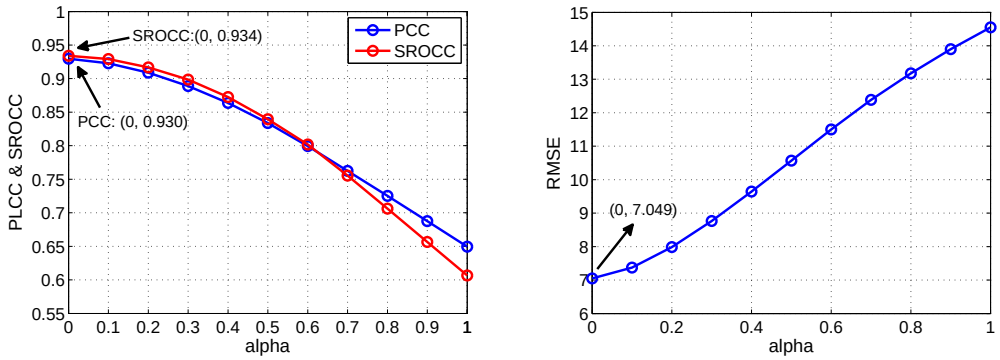
12.4.3 Contribution of the monocular and cyclopean images qualities



PCC, SROCC and RMSE values on LIVE 3D Phase II database

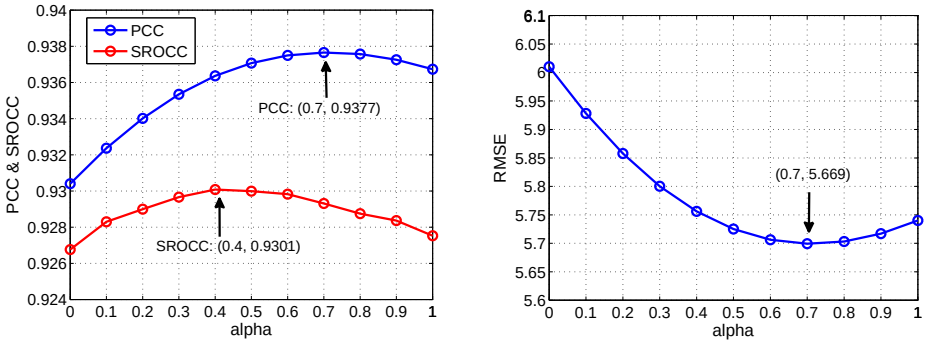


PCC, SROCC and RMSE values on Waterloo-IVC Phase I database

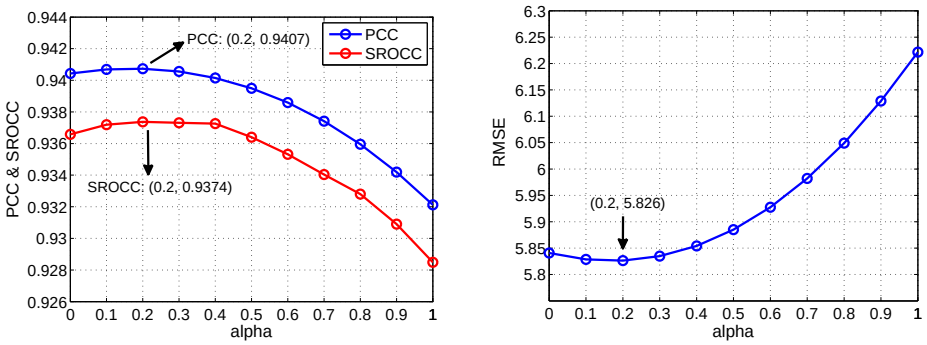


PCC, SROCC and RMSE values on Waterloo-IVC Phase II database

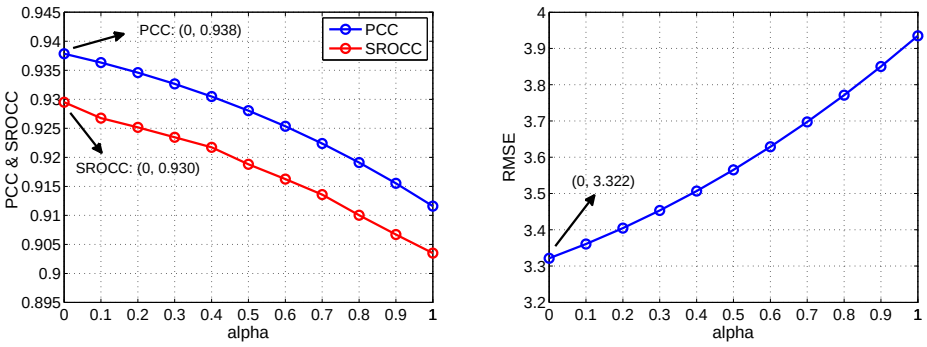
Figure 12.3 – Performance of MOBIQUM on three databases with different α values (see Eq. 12.17). Best results for each criterion are marked by arrows.



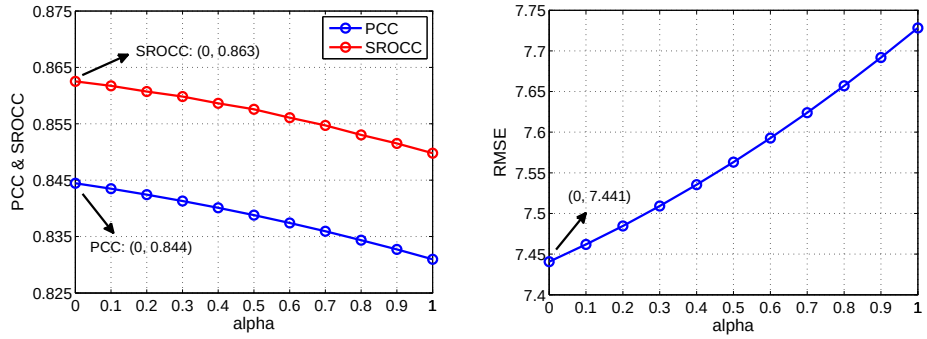
PCC, SROCC and RMSE values on LIVE 3D Phase I database



PCC, SROCC and RMSE values on NBU II database



PCC, SROCC and RMSE values on NBU-MDSID database



PCC, SROCC and RMSE values on IEEE 3D database

Figure 12.4 – Performance of MOBIQUM on four databases with different α value (see Eq. 12.17).

Based on these databases, we then evaluate the performance using different weight values α (used in Eq. 12.17) for combining monocular quality Q_{lr} and cyclopean quality Q_c in order to determine the optimal relationship between both qualities. Figures 12.3 and 12.4 illustrate the PCC, SROCC and RMSE values with different α on seven databases. The best-performance results are shown along with α values and the corresponding values. As shown in Figure 12.3, the proposed SIQA metric delivers the best performance with $\alpha = 0.2$, $\alpha = 0.1$ and $\alpha = 0$ on LIVE 3D II, Waterloo I and II databases, respectively. Figure 12.4 shows that the best results are obtained with $\alpha = 0.4$ for SROCC, and $\alpha = 0.7$ for LCC and RMSE on LIVE 3D I database. It should be noted that the results with different α values here were computed based on our SIQA metric without using JND map (see Figure 12.1) to modulate the cyclopean image quality Q_c. The proposed metric using the JND map achieves better results with $\alpha = 0.4$ than $\alpha = 0.7$. Therefore, we fix $\alpha = 0.4$ on LIVE 3D I for further performance comparison. Following this idea, we set $\alpha = 0.2$ and $\alpha = 0.1$ (see Table 12.2) for NBU-MDSID and IEEE 3D databases, even though $\alpha = 0$ results in the best performances illustrated in Figure 12.4. Moreover, for NBU II, $\alpha = 0.2$ leads to the best performance in terms of PCC, SROCC, and RMSE. In sum, the results in Figures 12.3 and 12.4 can conclude that the BS-inspired combined monocular quality has much more impact on the overall 3D quality Q_{lr} in contrast with the cyclopean image quality Q_c.

12.4.4 Impact of the different components of MOBIQUM

In this section, we discuss the effectiveness of each individual component of the proposed SIQA metric and compare the model's performance with different component-related models. Specifically, we first evaluate the 3D quality prediction performance using the proposed metric without/with different JND models as shown in Figure 12.1. After obtaining the best choice of the JND model for each database, we additionally explore the method performance with different strategies modeling the stimulus strength in BR phenomenon so as to demonstrate the effectiveness of image LoG response.

12.4.4.1 Impact of different 2D- and 3D-JND models on MOBIQUM

Tables 12.3 and 12.4 list the performance evaluation results using the proposed SIQA metric with different 2D- and 3D-JND models in terms of SROCC and RMSE indicators on all databases except Waterloo II, because the latter delivers the best performance without using JND model as described in Table 12.2. The results from Figure 12.3 show that the proposed metric without integrating JND or with integrating BJND achieve promising results on LIVE3D II. Moreover, DJND-based performs quite well on Waterloo I database. This can be explained by the fact that BJND and DJND models take the disparity/depth information into account, which undoubtedly correspond better to the way of judging 3D quality of the HVS. In addition, SJND-based and JND-TE-based perform also competitively well for LIVE 3D II and Waterloo I. This finding, which validates the effectiveness of SJND and JND-TE models, confirms the results from previous work [57, 121]. Besides, the use of

Table 12.3 – Performance evaluation of MOBIQUM with different 2D- and 3D-JND models based on LIVE 3D Phase II, Waterloo-IVC Phase I databases. The best results for each criterion are highlighted in boldface.

Strategy		LIVE 3D II		Waterloo I	
		SROCC	RMSE	SROCC	RMSE
	Without JND	0.940	**3.783**	0.934	5.501
With 2D-JND	NAMM [46]	0.940	3.792	0.934	5.494
	JND-TE [47]	0.940	3.788	0.926	**5.487**
With 3D-JND	BJND [48]	**0.941**	3.790	0.934	5.728
	SJND [49]	0.940	3.788	0.934	5.494
	DJND [50]	0.940	3.789	**0.935**	5.491

Table 12.4 – Performance evaluation of MOBIQUM with different 2D- and 3D-JND models based on LIVE 3D Phase I, NBU 3D II, NBU-MDSID, and IEEE 3D databases. The best results for each criterion are highlighted in boldface.

Strategy		LIVE 3D I		NBU-II		NBU-MDSID		IEEE 3D	
		SROCC	RMSE	SROCC	RMSE	SROCC	RMSE	SROCC	RMSE
	Without JND	0.930	5.777	0.949	4.833	0.931	3.281	0.891	6.574
With 2D-JND	NAMM [46]	0.930	5.775	0.949	**4.815**	0.931	3.283	0.891	6.576
	JND-TE [47]	0.929	5.824	0.949	4.821	0.931	3.286	0.891	6.575
With 3D-JND	BJND [48]	**0.934**	**5.565**	0.949	4.888	**0.934**	**3.236**	**0.892**	**6.542**
	SJND [49]	0.930	5.796	0.949	4.816	0.931	3.282	0.891	6.576
	DJND [50]	0.390	5.801	0.949	4.820	0.931	3.288	0.891	6.579

BJND outperforms all others JND models on LIVE 3D I, NBU-MDSID and IEEE 3D, as shown in Table 12.4. Furthermore, the NAMM-based performs the best on NBU-II, and delivers competitive performance on other databases. The proposed metric without JND can achieve acceptable results efficiently compared with the metrics with JND, because such JND-TE- or BJND-based metrics are relatively costly.

In summary, the proposed SIQA metric with BJND is the best compromise for the used 3D quality databases containing only symmetrically distorted stereopairs, whereas the metric without JND or with DJND performs well for databases containing both symmetric and asymmetric distortions.

12.4.4.2 Impact of different strategies modeling view dominance strength on MOBIQUM

Furthermore, we analyze the impact of different strategies (i.e., LoG, GM and LE maps) used to model the view dominance strength as shown in Figure 12.1. Table 12.5 details the performance of the proposed algorithm with each strategy on four databases. We can see that our method using LoG

Table 12.5 – Performance evaluation with LoG or GM or LE to simulate view dominance strength based on LIVE 3D Phase I, II, Waterloo-IVC II, and NBU-MDSID databases. The best results among LoG, GM and LE are highlighted in boldface.

Database	PCC			SROCC			RMSE		
	LoG	**GM**	**LE**	**LoG**	**GM**	**LE**	**LoG**	**GM**	**LE**
LIVE 3D I	**0.947**	0.940	0.941	**0.942**	0.934	0.934	**5.273**	5.587	5.541
LIVE 3D II	**0.942**	0.901	0.837	**0.941**	0.889	0.819	**3.790**	4.892	6.182
Waterloo II	0.921	**0.930**	0.739	0.924	**0.934**	0.719	7.477	**7.049**	12.898
NBU-MDSID	**0.941**	0.940	0.940	**0.938**	0.930	0.933	**3.236**	3.270	3.269

map outperforms the ones using GM and LE maps in terms of PCC, SROCC, and RMSE on LIVE 3D I, II and NBU-MDSID. Moreover, the LoG-based method delivers promising results on Waterloo II. The finding explains why we used the LoG in our metric design in order to accurately and efficiently simulate the visual stimulus strength of BR behavior. In addition, the GM-based method performs better than LE-based method on LIVE 3D II and Waterloo II, which contains both symmetrically and asymmetrically distorted stereopairs. Although LE-based method slightly outperforms the GM-based method on LIVE 3D I and NBU-MDSID, it remains less efficient in terms of computational runtime.

12.4.5 Overall performance

To evaluate the algorithm performance, we extensively compare our proposed method with competitive state-of-the-art SIQA methods, including the 2D-extended FR-IQA metrics, FR- and NR- 3D IQA metrics. For 2D-extended SIQA methods, we assess the 3D quality by averaging the quality scores of the left- and right-view images. Tables 12.6 and 12.7 shows the overall performance of the SIQA methods in terms of PCC, SROCC and RMSE results on three databases (containing both symmetric and asymmetric distortions) and four databases (containing only symmetric distortion), respectively. It can be seen from Table 12.6 that the proposed method outperforms most other SIQA methods on LIVE 3D II and Waterloo II databases, and delivers highly competitive performance on Waterloo I. Although Liu's method [93] achieves better performance than our proposed method on LIVE 3D II, his method is relatively more complicated due to the consideration of statistical features of the stereopair, cyclopean image, and binocular product image. Obviously, Wang's [73], Geng's [20], Fezza's [35], and Yao's [38] methods perform better than Benoit's [42] and You's [43] methods thanks to binocular vision properties consideration. In addition, we can observe that the performance results on Waterloo I and II are generally lower compared with LIVE II. This is mainly due to the fact that It is more challenging to assess 3D images with asymmetrical mixed distortions in Waterloo I and II. From Table 12.7, we observe that our SIQA method performs the best among all methods on LIVE 3D I, NBU 3D II and IEEE 3D. Moreover, the proposed method can yield promising results on NBU-MDSID. Among the 2D-based SIQA metrics, the GMSD-based method achieves the best performance on LIVE

Table 12.6 – Overall performance of SIQA methods on LIVE 3D Phase II, Waterloo-IVC Phase I and II databases. The ranking 1st and 2nd for each criterion are highlighted with red and blue bold texts, respectively.

SIQA	LIVE 3D II			Waterloo-IVC I			Waterloo-IVC II		
method	PCC	SROCC	RMSE	PCC	SROCC	RMSE	PCC	SROCC	RMSE
UQI [63]	0.864	0.842	5.677	0.753	0.640	10.360	0.679	0.591	14.056
VIF [88]	0.839	0.819	6.182	0.808	0.801	8.571	0.735	0.708	12.987
GMSD [90]	0.843	0.827	6.073	0.748	0.741	10.439	0.742	0.758	12.836
GMSM [90]	0.817	0.796	6.506	0.770	0.763	10.034	0.757	0.761	12.508
Benoit [42]	0.764	0.748	7.281	0.680	0.585	-	0.547	0.534	-
You [43]	0.721	0.721	7.141	0.679	0.597	-	0.679	0.622	-
Fezza [86]	0.871	0.862	5.553	0.692	0.553	11.357	0.561	0.447	15.846
Chen [16]	0.910	0.905	4.693	0.657	0.510	11.910	0.613	0.367	16.345
Wang [73]	0.916	0.919	-	0.930	0.918	-	0.892	0.869	-
Geng [20]	0.921	0.919	5.400	0.846	0.810	9.469	-	-	-
Khan [25]	0.932	0.922	-	0.934	0.925	-	0.910	0.905	-
Fezza [35]	0.925	0.908	**3.018**	0.904	0.898	-	0.890	0.886	-
Liu [93]	0.932	0.930	4.120	**0.956**	**0.937**	**4.960**	**0.922**	**0.911**	**7.377**
Yao [38]	**0.941**	**0.933**	**3.757**	0.932	0.907	-	0.869	0.868	-
Proposed	**0.942**	**0.941**	3.790	**0.937**	**0.935**	**5.491**	**0.930**	**0.934**	**7.049**

3D I and NBU 3D II, whereas the MS-SSIM-based method performs best on NBU-MDSID and IEEE 3D. Compared with Table 12.6, all 2D-based SIQA methods achieve reasonably accurate prediction results and better performance on the databases containing only the symmetric distortion in Table 12.7.

In addition to the overall performance evaluation mentioned above, we provide in Figure 12.5 the scatter distributions of subjective quality scores (i.e., DMOS) versus predicted scores by the proposed metric, as well as the non-linear fitting curves (marked in red) on different databases. DMOS values are provided by each database, and the fitting curves are determined via Eq. 12.18. Note that a better convergence of the data points in the scatter plot corresponds to a better consistency with the subjective quality scores. In spite of the presence of some outliers in Figure 12.5, the scatter points are well concentrated around the fitting curves, which indicates a high correlation between the human quality judgment and MOBIQUM.

12.4.6 Performance on symmetric and asymmetric distortions

As mentioned previously, the asymmetrically distorted stereopairs are more challenging for IQA task than the symmetrically distorted ones. Therefore, in addition to the above overall performance evaluation, we further tested the SIQA methods performances on separated symmetric and asymmetric

Table 12.7 – Overall performance of SIQA methods on LIVE 3D Phase I, NBU 3D II, NBU-MDSID, and IEEE 3D databases. The ranking 1st and 2nd for each criterion are highlighted with red and blue bold texts, respectively.

SIQA	LIVE 3D I			NBU 3D II			NBU-MDSID			IEEE 3D		
method	PCC	SROCC	RMSE	PCC	SROCC	RMSE	PCC	SROCC	RMSE	PCC	SROCC	RMSE
SSIM [62]	0.877	0.877	7.889	0.814	0.906	9.969	0.938	0.927	3.325	0.812	0.820	8.104
MS-SSIM [122]	0.930	0.924	6.026	0.840	0.924	9.313	**0.951**	**0.941**	**2.972**	0.821	0.833	7.925
UQI [63]	0.939	0.933	5.645	0.792	0.790	10.48	0.837	0.821	5.232	0.734	0.744	9.434
VIF [88]	0.925	0.920	6.237	0.869	0.864	8.495	0.691	0.574	6.916	0.652	0.655	10.53
GMSM [90]	0.938	0.933	5.676	0.838	0.937	9.368	0.867	0.858	4.774	0.801	0.816	8.316
GMSD [90]	0.943	**0.937**	5.472	0.938	**0.943**	5.976	0.897	0.873	4.227	0.795	0.804	8.429
Fezza [86]	0.821	0.922	9.358	0.867	0.937	8.548	**0.953**	**0.941**	**2.914**	0.832	0.845	**7.771**
Chen [16]	0.922	0.914	6.351	0.862	0.921	8.718	0.938	0.924	3.316	0.825	0.839	7.884
Shao [18]	0.935	0.925	5.816	0.941	**0.941**	**5.800**	0.919	0.905	3.687	-	-	-
Geng [20]	0.943	0.932	5.514	-	-	-	-	-	-	0.853	0.862	-
Wang [24]	0.924	0.916	6.272	0.925	0.928	6.541	-	-	-	-	-	-
Oh [34]	0.943	0.935	-	-	-	-	-	-	-	**0.855**	**0.927**	-
Jiang [100]	**0.946**	0.934	**5.276**	0.933	0.921	6.363	-	-	-	-	-	-
Jiang [123]	0.930	0.912	6.031	0.936	0.931	6.019	-	-	-	-	-	-
Proposed	**0.947**	**0.942**	**5.273**	**0.960**	**0.949**	**4.815**	0.941	**0.934**	3.236	**0.882**	**0.892**	**6.542**

distortions for LIVE 3D II, Waterloo I and II databases as displayed in Tables 12.8 and 12.9, respectively. The results from Table 12.8 demonstrate that MOBIQUM (i.e., the proposed method) significantly outperforms most of the FR 3D-based (Chen's [16] , Wang's [73] and Geng's [20]) methods for asymmetric distortions, and performs quite well for symmetric distortions. Moreover, Zhang's [17] and Fezza's [35] methods deliver competitive results for asymmetric distortion because both methods take BF and BR into account. Obviously, 2D-based SIQA methods perform quite well and even better than 3D-based methods for symmetric distortions, but they both generally achieve low performance for asymmetric distortions. This is mainly due to the fact that 2D-based methods predict the perceived 3D quality neither considering depth/disparity information nor the binocular visual properties. Besides, as shown in Table 12.9, our proposed method delivers the best performance on Waterloo-IVC II for both symmetric and asymmetric distortions, and perform better than recently proposed competitive SIQA methods i.e., Fezza's [35] and Khan's [25] methods on Waterloo-IVC I. It is worth noting that the proposed SIQA method computes 2D image quality using GMSM, which yields promising results for symmetric distortions.

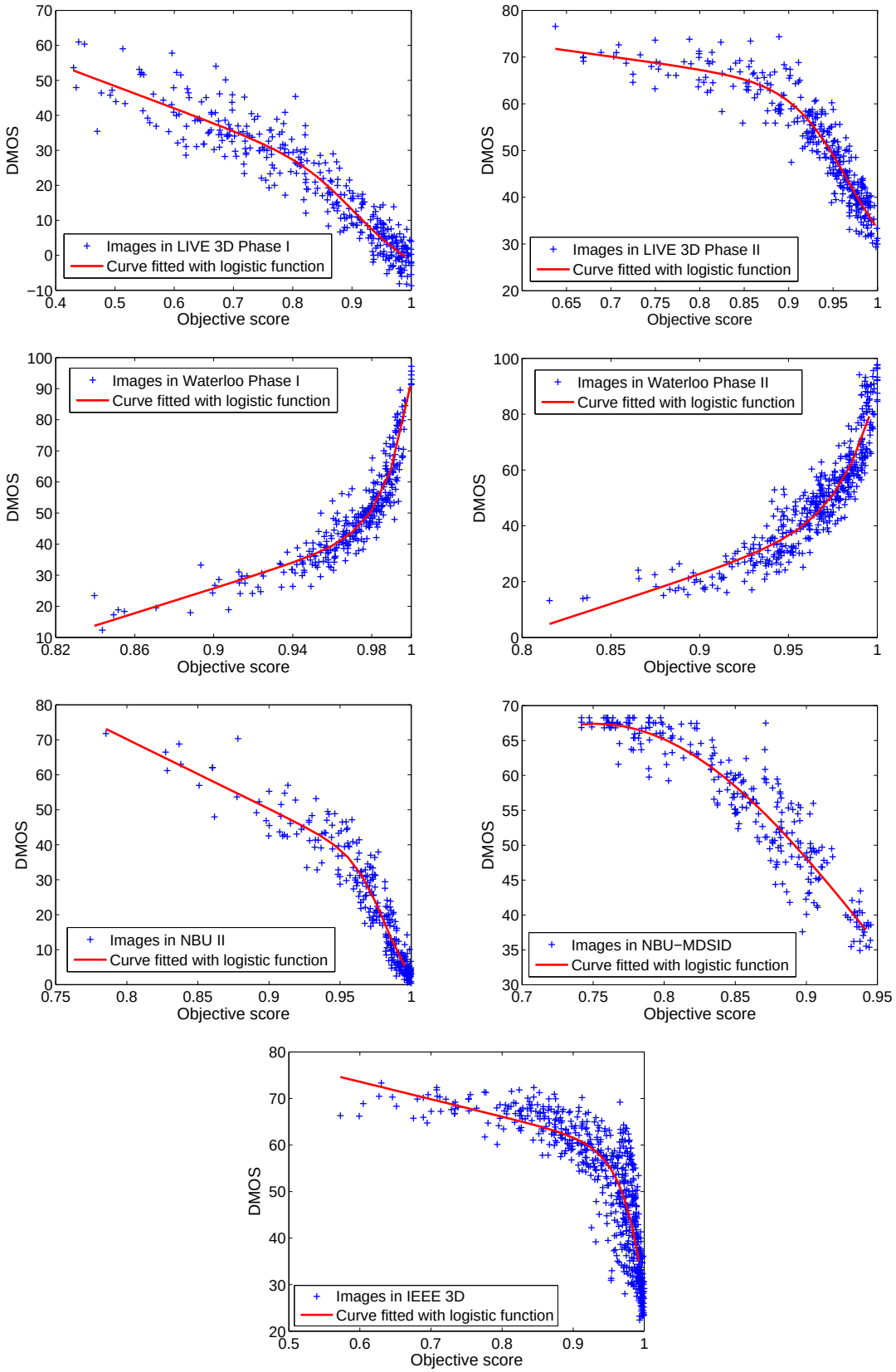


Figure 12.5 – Scatter plots of objective scores predicted by the proposed SIQA metric against subjective quality scores (DMOS) on different 3D-IQA databases.

Table 12.8 – Performance of SIQA methods on symmetric and asymmetric distortions of LIVE 3D Phase II database. The ranking 1st and 2nd for each criterion are highlighted with red and blue bold texts, respectively.

SIQA method	Symmetric			Asymmetric		
	PCC	SROCC	RMSE	PCC	SROCC	RMSE
UQI [63]	**0.941**	0.939	4.213	0.795	0.755	6.154
VIF [88]	0.928	0.916	4.652	0.777	0.732	6.382
GMSM [90]	**0.954**	**0.940**	**3.754**	0.734	0.688	6.885
GMSD [90]	0.937	0.923	4.350	0.775	0.755	6.404
Benoit [42]	0.921	0.910	5.712	0.746	0.732	6.976
You [43]	0.911	0.898	7.128	0.659	0.604	8.009
Fezza [86]	0.930	0.921	4.576	0.820	0.796	5.801
Chen [16]	0.940	0.927	4.277	0.878	0.859	4.846
Wang [73]	0.937	0.923	-	0.898	**0.902**	-
Zhang [17]	**0.954**	**0.947**	3.900	0.903	0.895	**4.371**
Geng [20]	0.938	0.929	4.414	0.877	0.868	5.667
Fezza [35]	0.935	0.928	-	**0.957**	0.892	-
Appina [124]	0.907	0.857	-	0.811	0.872	-
Proposed	**0.954**	**0.940**	**3.763**	**0.928**	**0.926**	**3.767**

Table 12.9 – Performance of SIQA methods on symmetric and asymmetric distortions of Waterloo-IVC Phase I and II databases.

SIQA method	Waterloo-IVC I				Waterloo-IVC II			
	PCC		SROCC		PCC		SROCC	
	S	**As**	**S**	**As**	**S**	**As**	**S**	**As**
UQI [63]	0.814	0.724	0.635	0.631	0.759	0.629	0.565	0.566
VIF [88]	0.918	0.788	0.913	0.755	0.887	0.653	0.830	0.631
GMSM [90]	**0.964**	0.683	**0.957**	0.705	**0.960**	0.666	**0.954**	0.683
GMSD [90]	0.951	0.641	0.934	0.676	**0.951**	0.638	0.940	0.677
Benoit [42]	0.850	0.697	0.728	0.577	0.755	0.555	0.571	0.454
You [43]	0.868	0.709	0.752	0.571	0.763	0.686	0.560	0.600
Fezza [86]	0.881	0.611	0.782	0.484	0.778	0.474	0.620	0.392
Chen [16]	0.869	0.592	0.774	0.442	0.736	0.449	0.512	0.341
Wang [73]	**0.964**	**0.929**	0.948	0.910	0.938	0.880	0.905	0.848
Khan [25]	-	**0.935**	-	**0.923**	-	**0.915**	-	**0.905**
Fezza [35]	0.910	0.882	0.902	0.869	0.914	0.845	0.915	0.804
Proposed	**0.970**	0.927	**0.963**	**0.926**	**0.960**	**0.921**	**0.953**	**0.925**

12.4.7 Performance on individual distortion type

Table 12.10 – Performance of SIQA methods for different types of distortion on LIVE 3D Phase I database. The ranking 1st and 2nd for each criterion are highlighted with red and blue bold texts, respectively.

SIQA	JPEG		JP2K		WN		GB		FF	
method	PCC	SROCC	PCC	SROCC	PCC	SROCC	PCC	SROCC	PCC	SROCC
SSIM [62]	0.475	0.435	0.858	0.857	0.944	0.939	0.907	0.879	0.670	0.584
MS-SSIM [122]	0.633	0.612	0.930	0.892	0.952	0.942	0.944	0.925	0.803	0.722
GMSM [90]	0.724	0.676	0.944	0.908	0.941	0.936	0.952	0.922	0.812	0.723
GMSD [90]	0.696	0.639	0.940	0.910	0.943	0.938	0.958	**0.939**	0.847	0.800
VIF [88]	0.681	0.580	0.937	0.900	0.931	0.931	0.965	0.934	0.862	0.804
UQI [63]	**0.756**	**0.729**	0.949	0.908	0.926	0.925	0.959	0.925	**0.884**	0.821
Chen [16]	0.528	0.530	0.920	0.888	0.955	**0.948**	0.943	0.924	0.776	0.700
Lin [19]	0.755	0.716	**0.952**	**0.913**	0.927	0.929	0.958	0.933	0.862	**0.829**
Shao [18]	0.520	0.495	0.921	0.895	0.945	0.941	0.959	**0.940**	0.859	0.796
Shao [22]	0.665	0.634	0.936	0.900	0.944	0.943	0.954	0.924	0.831	0.781
Geng [20]	0.719	0.653	0.942	0.905	**0.963**	**0.956**	0.962	0.931	0.867	0.816
Wang [24]	0.617	0.513	0.909	0.881	0.949	0.944	0.951	0.931	0.778	0.686
Khan [25]	0.711	0.606	0.951	0.907	0.947	0.939	0.959	0.930	0.858	0.809
Chen [14]	0.695	0.617	0.907	0.863	0.917	0.919	0.917	0.878	0.735	0.652
Jiang [100]	0.707	0.609	0.942	0.884	0.950	0.925	**0.968**	0.922	0.868	0.796
Yue [125]	0.744	0.595	0.934	0.833	**0.962**	0.932	**0.971**	0.857	0.854	0.779
Proposed	**0.779**	**0.744**	**0.957**	**0.921**	0.947	0.942	0.958	0.926	**0.879**	**0.823**

In addition to the exploration made above, we also examine the performance of SIQA methods for different distortion types on four databases including LIVE 3D I and II, NBU 3D II and IEEE 3D. The results are separately reported in Tables 12.10, 12.11, 12.12 and 12.13. For LIVE 3D Phase I, MOBIQUM outperforms most of the other SIQA methods for JPEG, JP2K, and FF distortions, and achieves high performance for WN and GB distortions. In Table 12.11, it can be noticed that our method delivers the best performance for WN distortions and yields promising results for JPEG, FF and GB distortions. Moreover, it provides competitive results for JP2K distortion, even though it performs less efficiently than some 3D-based SIQA methods such as Wang's [24] and Jiang's [100] methods. Interestingly, some 2D-based SIQA methods perform quite well for specific distortion types. For instance, GMSM provides high performance for JPEG and WN distortions, UQI delivers the best performance for FF distortion and VIF yields the best results for GB distortion. The UQI-based method performs better than other methods for FF distortion on LIVE 3D I (see Table 12.10) and II databases. For NBU 3D II database, the proposed SIQA method still performs competitively

Table 12.11 – Performance of SIQA methods for different types of distortion on LIVE 3D Phase II database. The ranking 1st and 2nd for each criterion are highlighted with red and blue bold texts, respectively.

SIQA	JPEG		JP2K		WN		GB		FF	
method	PCC	SROCC	PCC	SROCC	PCC	SROCC	PCC	SROCC	PCC	SROCC
SSIM [62]	0.664	0.678	0.730	0.703	0.931	0.922	0.848	0.838	0.868	0.834
MS-SSIM [122]	0.859	0.847	0.824	0.798	0.951	0.946	0.798	0.799	0.871	0.832
GMSM [90]	**0.924**	**0.900**	0.886	0.877	0.956	**0.953**	0.900	0.874	0.908	0.874
GMSD [90]	0.898	0.887	0.871	0.868	0.932	0.934	0.944	0.921	**0.968**	0.935
VIF [88]	0.824	0.778	0.831	0.822	0.834	0.822	**0.987**	**0.951**	0.947	0.934
UQI [63]	0.799	0.739	0.841	0.831	0.851	0.840	0.961	**0.954**	**0.980**	**0.956**
Chen [16]	0.862	0.843	0.834	0.814	**0.957**	0.940	0.963	0.908	0.901	0.884
Shao [22]	0.851	0.834	0.877	0.854	0.934	0.933	0.945	0.924	0.933	0.941
Voo [126]	0.679	0.737	0.814	0.895	0.937	0.933	0.872	0.879	0.911	0.905
Wang [24]	0.909	0.826	**0.953**	0.773	0.891	**0.953**	0.952	0.770	0.921	0.831
Ma [23]	0.899	0.879	0.887	0.878	0.957	0.949	**0.978**	0.906	0.901	0.893
Chen [14]	0.901	0.867	0.899	0.950	0.947	0.950	0.941	0.900	0.932	0.933
Appina [124]	0.829	0.839	0.867	0.864	0.920	0.932	0.878	0.846	0.836	0.860
Jiang [100]	**0.924**	0.889	**0.942**	**0.909**	0.917	0.873	0.920	0.865	0.919	0.887
Liu [127]	0.904	0.858	0.765	**0.908**	0.817	0.940	0.975	0.901	0.931	0.924
Proposed	**0.916**	**0.896**	0.894	0.879	**0.965**	**0.959**	**0.978**	0.896	0.955	**0.945**

well. In particular, our method achieves high performance for JPEG, JP2K and GB distortions, and delivers promising results for WN and H.264 distortions. Similar to the results in Table 12.11, VIF performs also quite well for GB distortion. Besides, the results from Table 12.13 demonstrates that the proposed method achieves high performances for JPEG, FF distortion and competitive results for WN and GB distortions compared with other SIQA methods. Overall, MOBIQUM can achieve remarkable performance for JPEG and WN distortions across the used four databases. In summary, extensive results from previous results comprehensively validate the effectiveness of the proposed method, and show high correlation with the human judgment of quality.

12. Paper VII: Stereoscopic Image Quality Assessment based on Monocular and Binocular Visual Properties

Table 12.12 – Performance of SIQA methods for different types of distortion on NBU 3D II database.

SIQA	JPEG		JP2K		WN		GB		H.264	
method	PCC	SROCC	PCC	SROCC	PCC	SROCC	PCC	SROCC	PCC	SROCC
SSIM [62]	**0.956**	**0.949**	0.681	0.901	0.745	0.855	0.950	0.949	0.751	0.921
MS-SSIM [122]	0.937	0.883	0.937	**0.952**	0.930	0.933	0.968	0.948	0.924	0.943
GMSM [90]	0.821	0.911	0.938	**0.950**	0.929	**0.947**	0.972	0.967	0.958	**0.959**
GMSD [90]	0.916	0.913	**0.953**	0.947	0.916	0.937	0.967	0.961	0.943	0.941
VIF [88]	0.952	0.915	0.823	0.939	0.928	0.920	**0.980**	**0.980**	0.951	0.928
UQI [63]	0.869	0.768	0.725	0.754	0.642	0.662	0.940	0.938	0.746	0.740
Chen [16]	0.949	0.908	0.891	0.924	0.906	0.928	0.955	0.948	0.926	0.927
Shao [18]	0.928	0.935	**0.953**	**0.952**	**0.981**	0.945	0.963	**0.977**	**0.962**	**0.955**
Wang [24]	0.909	0.826	**0.953**	0.773	0.891	0.953	0.952	0.770	0.921	0.831
Jiang [100]	0.924	0.889	0.942	0.909	0.917	0.873	0.920	0.865	0.919	0.887
Jiang [123]	0.920	0.924	0.895	0.914	**0.953**	**0.955**	0.925	0.874	0.936	0.922
Li [36]	0.900	0.883	0.906	0.942	0.911	0.926	0.950	0.966	-	-
Proposed	**0.953**	**0.952**	**0.954**	0.947	0.952	0.918	**0.973**	0.970	**0.968**	0.937

Table 12.13 – Performance of SIQA methods for different types of distortion on IEEE 3D database. The ranking 1st and 2nd for each criterion are highlighted with red and blue bold texts, respectively.

SIQA	JPEG		JP2K		WN		GB		FF	
method	PCC	SROCC	PCC	SROCC	PCC	SROCC	PCC	SROCC	PCC	SROCC
SSIM [62]	**0.944**	0.905	0.862	0.851	0.785	0.786	**0.927**	0.858	0.965	0.946
MS-SSIM [122]	0.942	0.903	0.884	0.871	0.739	0.746	0.919	0.844	**0.974**	**0.957**
GMSM [90]	0.930	0.902	0.823	0.813	0.804	0.802	0.885	0.840	0.910	0.892
GMSD [90]	0.933	0.884	**0.919**	**0.899**	0.741	0.754	**0.921**	0.852	0.955	0.945
VIF [88]	0.942	**0.926**	0.890	0.879	0.837	0.830	0.920	**0.871**	0.954	0.945
UQI [63]	0.902	0.869	0.569	0.548	0.887	0.865	0.905	0.840	0.944	0.936
Chen [16]	0.940	0.906	0.886	0.875	0.727	0.748	0.917	0.842	**0.972**	**0.957**
Lin [87]	0.869	0.844	0.762	0.748	0.869	0.841	0.794	0.743	0.837	0.812
Fezza [35]	0.933	0.925	0.839	0.805	**0.902**	**0.923**	0.848	0.808	0.889	0.907
Oh [34]	0.906	0.891	**0.925**	**0.948**	0.823	**0.905**	0.839	0.837	0.871	0.941
Proposed	**0.948**	**0.943**	0.710	0.717	**0.940**	0.902	0.913	**0.862**	0.970	**0.949**

12.5 Conclusion and future work

In this paper, we propose a FR-SIQA system considering the qualities of (1) left- and right-views images (related to monocular vision) and (2) a cyclopean-view image (related to BV). Specifically, we first compute the quality of left and right image separately based on gradient magnitude similarity metric (GMSM), and then linearly combine the qualities of both views into a 2D quality with the weights modeling the related stimulus strength of each view. Next, using the GMSM metric, we estimate the quality of the cyclopean image, which is derived from a BF-inspired combination model. In addition, the cyclopean quality of the test stereo pair is weighted with a JND map of the reference stereo pair that reveals different visual sensitivities on the image's region degradation. Finally, the overall 3D quality score is computed by integrating 2D monocular image quality with 3D binocular-based JND-weighted cyclopean image quality. Experimental results on extensive databases show that the proposed SIQA metric delivers high quality prediction accuracy and outperforms than many other SIQA methods.

12. Paper VII: Stereoscopic Image Quality Assessment based on Monocular and Binocular Visual Properties

Bibliography

[1] Marc Lambooij, Marten Fortuin, Ingrid Heynderickx, and Wijnand IJsselsteijn. Visual discomfort and visual fatigue of stereoscopic displays: a review. *Journal of Imaging Science and Technology*, 53(3):30201–1, 2009. 224

[2] Wa James Tam, Filippo Speranza, Sumio Yano, Koichi Shimono, and Hiroshi Ono. Stereoscopic 3D-TV: visual comfort. *IEEE Transactions on Broadcasting*, 57(2):335–346, 2011. 224

[3] J. Wang, S. Wang, and Z. Wang. Asymmetrically Compressed Stereoscopic 3D Videos: Quality Assessment and Rate-Distortion Performance Evaluation. *IEEE Transactions on Image Processing*, 26(3):1330–1343, March 2017. 224, 227

[4] J. Wang, S. Wang, K. Ma, and Z. Wang. Perceptual depth quality in distorted stereoscopic images. *IEEE Transactions on Image Processing*, 26(3):1202–1215, March 2017. 224

[5] W. Hong and L. Yu. A spatio-temporal perceptual quality index measuring compression distortions of three-dimensional video. *IEEE Signal Processing Letters*, 25(2):214–218, Feb 2018. 224

[6] F. Battisti, M. Carli, P. Le Callet, and P. Paudyal. Toward the assessment of quality of experience for asymmetric encoding in immersive media. *IEEE Transactions on Broadcasting*, 64(2):392–406, June 2018. 224, 227

[7] IT Union. Subjective methods for the assessment of stereoscopic 3DTV systems. *Recommendation ITU-R BT*, 2021, 2015. 224

[8] Zhou Wang and Alan C Bovik. Mean squared error: Love it or leave it? a new look at signal fidelity measures. *IEEE Signal Processing Magazine*, 26(1):98–117, 2009. 224

[9] Alan Conrad Bovik. Automatic prediction of perceptual image and video quality. *Proceedings of the IEEE*, 101(9):2008–2024, 2013. 224

[10] Deepti Ghadiyaram and Alan C Bovik. Massive online crowdsourced study of subjective and objective picture quality. *IEEE Transactions on Image Processing*, 25(1):372–387, 2016. 224

[11] Jongyoo Kim, Hui Zeng, Deepti Ghadiyaram, Sanghoon Lee, Lei Zhang, and Alan C Bovik. Deep convolutional neural models for picture-quality prediction: Challenges and solutions to data-driven image quality assessment. *IEEE Signal Processing Magazine*, 34(6):130–141, 2017. 224

[12] Kede Ma, Wentao Liu, Kai Zhang, Zhengfang Duanmu, Zhou Wang, and Wangmeng Zuo. End-to-end blind image quality assessment using deep neural networks. *IEEE Transactions on Image Processing*, 27(3):1202–1213, 2018. 224

[13] Che-Chun Su, Anush Krishna Moorthy, and Alan Conrad Bovik. Visual quality assessment of stereoscopic image and video: challenges, advances, and future trends. In *Visual Signal Quality Assessment*, pages 185–212. Springer, 2015. 224, 236

[14] Ming-Jun Chen, Lawrence K Cormack, and Alan C Bovik. No-reference quality assessment of natural stereopairs. *IEEE Transactions on Image Processing*, 22(9):3379–3391, 2013. 224, 225, 227, 248, 249

[15] Jan Brascamp, Hansem Sohn, Sang-Hun Lee, and Randolph Blake. A monocular contribution to stimulus rivalry. *PNAS*, 110(21):8337–8344, 2013. 224, 227

[16] Ming-Jun Chen, Che-Chun Su, Do-Kyoung Kwon, Lawrence K Cormack, and Alan C Bovik. Full-reference quality assessment of stereopairs accounting for rivalry. *Signal Processing: Image Communication*, 28(9):1143–1155, 2013. 225, 227, 228, 229, 232, 236, 237, 244, 245, 247, 248, 249, 250

[17] Yi Zhang and Damon M Chandler. 3d-mad: A full reference stereoscopic image quality estimator based on binocular lightness and contrast perception. *IEEE Transactions on Image Processing*, 24(11):3810–3825, 2015. 225, 226, 228, 229, 232, 245, 247

[18] Feng Shao, Kemeng Li, Weisi Lin, Gangyi Jiang, Mei Yu, and Qionghai Dai. Full-reference quality assessment of stereoscopic images by learning binocular receptive field properties. *IEEE Transactions on Image Processing*, 24(10):2971–2983, 2015. 225, 228, 229, 245, 248, 250

[19] Yancong Lin, Jiachen Yang, Wen Lu, Qinggang Meng, Zhihan Lv, and Houbing Song. Quality index for stereoscopic images by jointly evaluating cyclopean amplitude and cyclopean phase. *IEEE Journal of Selected Topics in Signal Processing*, 11(1):89–101, 2017. 225, 232, 248

[20] Xianqiu Geng, Liquan Shen, Kai Li, and Ping An. A stereoscopic image quality assessment model based on independent component analysis and binocular fusion property. *Signal Processing: Image Communication*, 52:54–63, 2017. 225, 227, 243, 244, 245, 247, 248

[21] Jian Ma, Ping An, Liquan Shen, and Kai Li. Full-reference quality assessment of stereoscopic images by learning binocular visual properties. *Applied Optics*, 56(29):8291–8302, 2017. 225, 228, 229, 232

[22] Feng Shao, Wanting Chen, Gangyi Jiang, and Yo-Sung Ho. Modeling the perceptual quality of stereoscopic images in the primary visual cortex. *IEEE Access*, 5:15706–15716, 2017. 225, 232, 248, 249

[23] Jian Ma, Ping An, Liquan Shen, and Kai Li. Joint binocular energy-contrast perception for quality assessment of stereoscopic images. *Signal Processing: Image Communication*, 65:33–45, 2018. 225, 228, 232, 249

[24] Xu Wang, Lin Ma, Sam Kwong, and Yu Zhou. Quaternion representation based visual saliency for stereoscopic image quality assessment. *Signal Processing*, 145:202–213, 2018. 225, 245, 248, 249, 250

[25] S. Khan and S. S. Channappayya. Estimating depth-salient edges and its application to stereoscopic image quality assessment. *IEEE Transactions on Image Processing*, 27(12):5892–5903, Dec 2018. ISSN 1057-7149. doi: 10.1109/TIP.2018.2860279. 225, 228, 230, 244, 245, 247, 248

[26] Feng Qi, Debin Zhao, and Wen Gao. Reduced reference stereoscopic image quality assessment based on binocular perceptual information. *IEEE Transactions on Multimedia*, 17(12):2338–2344, 2015. 225

[27] Xu Wang, Qiong Liu, Ran Wang, and Zhuo Chen. Natural image statistics based 3D reduced reference image quality assessment in contourlet domain. *Neurocomputing*, 151:683–691, 2015. 225

[28] Lin Ma, Xu Wang, Qiong Liu, and King Ngi Ngan. Reorganized DCT-based image representation for reduced reference stereoscopic image quality assessment. *Neurocomputing*, 215:21–31, 2016. 225

[29] Jiachen Yang, Bin Jiang, Yafang Wang, Wen Lu, and Qinggang Meng. Sparse representation based stereoscopic image quality assessment accounting for perceptual cognitive process. *Information Science*, 430:1–16, 2018. 225

[30] Jian Ma, Ping An, Liquan Shen, and Kai Li. Reduced-reference stereoscopic image quality assessment using natural scene statistics and structural degradation. *IEEE Access*, 6:2768–2780, 2018. 225

[31] Che-Chun Su, Lawrence K Cormack, and Alan C Bovik. Oriented correlation models of distorted natural images with application to natural stereopair quality evaluation. *IEEE Transactions on Image Processing*, 24(5):1685–1699, 2015. 225, 228

[32] Feng Shao, Weijun Tian, Weisi Lin, Gangyi Jiang, and Qionghai Dai. Toward a blind deep quality evaluator for stereoscopic images based on monocular and binocular interactions. *IEEE Transactions on Image Processing*, 25(5):2059–2074, 2016. 225

[33] Wujie Zhou, Lu Yu, Yang Zhou, Weiwei Qiu, Ming-Wei Wu, and Ting Luo. Blind quality estimator for 3d images based on binocular combination and extreme learning machine. *Pattern Recognition*, 71:207–217, 2017. 225, 232

[34] Heeseok Oh, Sewoong Ahn, Jongyoo Kim, and Sanghoon Lee. Blind deep S3D image quality evaluation via local to global feature aggregation. *IEEE Transactions on Image Processing*, 26 (10):4923–4936, 2017. 225, 245, 250

[35] Sid Ahmed Fezza, Aladine Chetouani, and Mohamed-Chaker Larabi. Using distortion and asymmetry determination for blind stereoscopic image quality assessment strategy. *Journal of Visual Communication and Image Representation*, 49:115–128, 2017. 225, 243, 244, 245, 247, 250

[36] Fucui Li, Feng Shao, Qiuping Jiang, Randi Fu, Gangyi Jiang, and Mei Yu. Local and global sparse representation for no-reference quality assessment of stereoscopic images. *Information Science*, 422:110–121, 2018. 225, 250

[37] Jiachen Yang, Kyohoon Sim, Bin Jiang, and Wen Lu. No-reference stereoscopic image quality assessment based on hue summation–difference mapping image and binocular joint mutual filtering. *Applied optics*, 57(14):3915–3926, 2018. 225

[38] Yang Yao, Liquan Shen, and Ping An. Bivariate analysis of 3D structure for stereoscopic image quality assessment. *Signal Processing: Image Communication*, 65:128–140, 2018. 225, 243, 244

[39] Feng Shao, Ying Gao, Qiuping Jiang, Gangyi Jiang, and Yo-Sung Ho. Multi-stage pooling for blind quality prediction of asymmetric multiply-distorted stereoscopic images. *IEEE Transactions on Multimedia*, 2018. 225

[40] Julesz Bela. *Foundations of cyclopean perception.* U. Chicago Press, Oxford, England, 1971. 225

[41] Aldo Maalouf and Mohamed-Chaker Larabi. Cyclop: A stereo color image quality assessment metric. In *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 1161–1164. IEEE, 2011. 225, 229

[42] Alexandre Benoit, Patrick Le Callet, Patrizio Campisi, and Romain Cousseau. Quality assessment of stereoscopic images. *EURASIP Journal on Image and Video Processing*, 2008(1):659024, Jan 2009. ISSN 1687-5281. doi: 10.1155/2008/659024. URL https://doi.org/10.1155/2008/659024. 225, 228, 236, 243, 244, 247

[43] Junyong You, Liyuan Xing, Andrew Perkis, and Xu Wang. Perceptual quality assessment for stereoscopic images based on 2D image quality metrics and disparity analysis. In *Proc. Int. Workshop Video Process. Quality Metrics Consum. Electron*, volume 9, pages 1–6, 2010. 225, 228, 243, 244, 247

[44] Ee-Leng Tan and Woon-Seng Gan. Computational models for just-noticeable differences. In *Perceptual Image Coding with Discrete Cosine Transform*, pages 3–19. Springer, Singapore, 2015. doi: 10.1007/978-981-287-543-3_2. 225

[45] Nikil Jayant, James Johnston, and Robert Safranek. Signal compression based on models of human perception. *Proceedings of the IEEE*, 81(10):1385–1422, 1993. 225

[46] XK Yang, WS Ling, ZK Lu, Ee Ping Ong, and SS Yao. Just noticeable distortion model and its applications in video coding. *Signal Processing: Image Communication*, 20(7):662–680, 2005. 225, 235, 238, 242

[47] Anmin Liu, Weisi Lin, Manoranjan Paul, Chenwei Deng, and Fan Zhang. Just noticeable difference for images with decomposition model for separating edge and textured regions. *IEEE Transactions on Circuits and Systems for Video Technology*, 20(11):1648–1652, 2010. 225, 235, 242

[48] Yin Zhao, Zhenzhong Chen, Ce Zhu, Yap-Peng Tan, and Lu Yu. Binocular just-noticeable-difference model for stereoscopic images. *IEEE Signal Processing Letters*, 18(1):19–22, 2011. 225, 235, 238, 242

[49] Feng Qi, Debin Zhao, Xiaopeng Fan, and Tingting Jiang. Stereoscopic video quality assessment based on visual attention and just-noticeable difference models. *Signal, Image and Video Processing*, 10(4):737–744, 2016. 225, 235, 242

[50] Fei Xue, Cheolkon Jung, and Joongkyu Kim. Disparity-based just-noticeable-difference model for perceptual stereoscopic video coding using depth of focus blur effect. *Displays*, 42:43–50, 2016. 225, 235, 238, 242

[51] Xu Wang, Sam Kwong, and Yun Zhang. Considering binocular spatial sensitivity in stereoscopic image quality assessment. In *IEEE International Conference on Visual Communications and Image Processing (VCIP)*, pages 1–4, Taiwan, 2011. IEEE. 225, 227, 228

[52] Walid Hachicha, Azeddine Beghdadi, and Faouzi Alaya Cheikh. Stereo image quality assessment using a binocular just noticeable difference model. In *20th IEEE International Conference on Image Processing (ICIP)*, pages 113–117, Melbourne, 2013. IEEE. 225, 227, 228

[53] Feng Shao, Weisi Lin, Shanbo Gu, Gangyi Jiang, and Thambipillai Srikanthan. Perceptual full-reference quality assessment of stereoscopic images by considering binocular visual characteristics. *IEEE Transactions on Image Processing*, 22(5):1940–1953, 2013. 225, 227, 228

[54] Sid Ahmed Fezza, Mohamed-Chaker Larabi, and Kamel Mohamed Faraoun. Stereoscopic image quality metric based on local entropy and binocular just noticeable difference. In *IEEE International Conference on Image Processing (ICIP)*, pages 2002–2006. IEEE, 2014. 225, 227, 228

[55] Wujie Zhou, Gangyi Jiang, Mei Yu, Feng Shao, and Zongju Peng. PMFS: a perceptual modulated feature similarity metric for stereoscopic image quality assessment. *IEEE Signal Processing Letters*, 21(8):1003–1006, 2014. 225, 227, 228

[56] Yu Cao, Wenhao Hong, and Lu Yu. Full-reference perceptual quality assessment for stereoscopic images based on primary visual processing mechanism. In *IEEE International Conference on Multimedia and Expo (ICME)*, pages 1–6. IEEE, 2016. 225, 227, 228, 230

[57] Yu Fan, Mohamed-Chaker Larabi, Faouzi Alaya Cheikh, and Christine Fernandez-Maloigne. Stereoscopic image quality assessment based on the binocular properties of the human visual system. In *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 2037–2041. IEEE, 2017. 225, 232, 241

[58] Yu Fan, Mohamed-Chaker Larabi, Faouzi Alaya Cheikh, and Christine Fernandez-Maloigne. Full-reference stereoscopic image quality assessment accounting for binocular combination and disparity information. In *IEEE International Conference on Image Processing (ICIP)*, pages 760–764. IEEE, 2017. 225, 232

[59] Patrizio Campisi, Patrick Le Callet, and Enrico Marini. Stereoscopic images quality assessment. In *15th European Signal Processing Conference*, pages 2110–2114. IEEE, 2007. 226, 227, 235

[60] Paul Gorley and Nick Holliman. Stereoscopic image quality metrics and compression. In *Stereoscopic Displays and Applications XIX*, volume 6803. International Society for Optics and Photonics, 2008. 226, 227

[61] Jiachen Yang, Chunping Hou, Yuan Zhou, Zhuoyun Zhang, and Jichang Guo. Objective quality assessment method of stereo images. In *3DTV Conference: The True Vision-Capture, Transmission and Display of 3D Video, 2009*, pages 1–4. IEEE, 2009. 226, 236

[62] Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing*, 13(4): 600–612, 2004. 227, 245, 248, 249, 250

[63] Zhou Wang and Alan C Bovik. A universal image quality index. *IEEE signal processing letters*, 9(3):81–84, 2002. 227, 229, 244, 245, 247, 248, 249, 250

[64] M. Carnec, P. Le Callet, and D. Barba. An image quality assessment method based on perception of structural information. In *IEEE International Conference on Image Processing (ICIP)*, volume 3, pages III–185, Sep. 2003. doi: 10.1109/ICIP.2003.1247212. 227

[65] Zhou Wang and Eero P Simoncelli. Reduced-reference image quality assessment using a wavelet-domain natural image statistic model. In *Human Vision and Electronic Imaging X*, volume 5666, pages 149–160, 2005. 227

[66] David G Lowe. Object recognition from local scale-invariant features. In *Proceedings of the Seventh IEEE International Conference on Computer Vision*, volume 2, pages 1150–1157, Sep. 1999. 227

[67] Martin A Fischler and Robert C Bolles. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, 24(6):381–395, 1981. 227

[68] Pieter Seuntiens, Lydia Meesters, and Wijnand Ijsselsteijn. Perceived quality of compressed stereoscopic images: Effects of symmetric and asymmetric jpeg coding and camera separation. *ACM Transactions on Applied Perception (TAP)*, 3(2):95–109, 2006. 227

[69] Sid Ahmed Fezza and Mohamed-Chaker Larabi. Perceptually driven nonuniform asymmetric coding of stereoscopic 3d video. *IEEE Transactions on Circuits and Systems for Video Technology*, 27(10):2231–2245, 2017. 227, 232

[70] V De Silva, H Kodikara Arachchi, Erhan Ekmekcioglu, Anil Fernando, Safak Dogan, Ahmet Kondoz, and S Savas. Psycho-physical limits of interocular blur suppression and its application to asymmetric stereoscopic video delivery. In *19th International Packet Video Workshop*, pages 184–189, May 2012. 227

[71] Jeremy M Wolfe. Stereopsis and binocular rivalry. *Psychological review*, 93(3):269, 1986. 227

[72] Randolph Blake and Nikos K Logothetis. Visual competition. *Nature Reviews Neuroscience*, 3 (1):13–21, 2002. 227

[73] Jiheng Wang, Abdul Rehman, Kai Zeng, Shiqi Wang, and Zhou Wang. Quality prediction of asymmetrically distorted stereoscopic 3D images. *IEEE Transactions on Image Processing*, 24 (11):3400–3414, 2015. 227, 232, 236, 237, 243, 244, 245, 247

[74] Gregory C DeAngelis, Izumi Ohzawa, and Ralph D Freeman. Depth is encoded in the visual cortex by a specialized receptive field structure. *Nature*, 352(6331):156, 1991. 227

[75] BG Cumming. An unexpected specialization for horizontal disparity in primate primary visual cortex. *Nature*, 418(6898):633, 2002. 227

[76] Jae-Jeong Hwang and Hong Ren Wu. Stereo image quality assessment using visual attention and distortion predictors. *TIIS*, 5(9):1613–1631, 2011. 227, 228

[77] Alex J Smola and Bernhard Schölkopf. A tutorial on support vector regression. *Statistics and computing*, 14(3):199–222, 2004. 228

[78] Dwight W Curtis and Stanley J Rule. Binocular processing of brightness information: A vector-sum model. *Journal of Experimental Psychology: Human Perception and Performance*, 4(1): 132, 1978. 228

[79] Yun Liu, Jiachen Yang, Qinggang Meng, Zhihan Lv, Zhanjie Song, and Zhiqun Gao. Stereoscopic image quality assessment method based on binocular combination saliency model. *Signal Processing*, 125:237–248, 2016. 228

[80] Alexander I Cogan. Human binocular interaction: Towards a neural model. *Vision research*, 27 (12):2125–2139, 1987. 228

[81] Jiachen Yang, Yafang Wang, Baihua Li, Wen Lu, Qinggang Meng, Zhihan Lv, Dezong Zhao, and Zhiqun Gao. Quality assessment metric of stereo images considering cyclopean integration and visual saliency. *Information Sciences*, 373:251–268, 2016. 228

[82] Jian Ding and George Sperling. A gain-control theory of binocular combination. *Proceedings of the National Academy of Sciences*, 103(4):1141–1146, 2006. 228, 233

[83] Jian Ding, Stanley A Klein, and Dennis M Levi. Binocular combination of phase and contrast explained by a gain-control and gain-enhancement model. *Journal of vision*, 13(2):13–13, 2013. 228

[84] Jiawei Zhou, Mark A Georgeson, and Robert F Hess. Linear binocular combination of responses to contrast modulation: Contrast-weighted summation in first-and second-order vision. *Journal of vision*, 14(13):24–24, 2014. 228

[85] Mark A Georgeson, Stuart A Wallis, Tim S Meese, and Daniel H Baker. Contrast and lustre: A model that accounts for eleven different forms of contrast discrimination in binocular vision. *Vision research*, 129:98–118, 2016. 228

[86] Sid Ahmed Fezza and Mohamed-Chaker Larabi. Stereoscopic 3d image quality assessment based on cyclopean view and depth map. In *IEEE Fourth International Conference on Consumer Electronics–Berlin (ICCE-Berlin)*, pages 335–339. IEEE, 2014. 229, 244, 245, 247

[87] Yu-Hsun Lin and Ja-Ling Wu. Quality assessment of stereoscopic 3D image compression by binocular integration behaviors. *IEEE Transactions on Image Processing*, 23(4):1527–1542, 2014. 229, 232, 250

[88] Hamid R Sheikh and Alan C Bovik. Image information and visual quality. In *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, volume 3, pages iii–709. IEEE, 2004. 229, 244, 245, 247, 248, 249, 250

[89] Lin Zhang, Lei Zhang, Xuanqin Mou, David Zhang, et al. FSIM: a feature similarity index for image quality assessment. *IEEE transactions on Image Processing*, 20(8):2378–2386, 2011. 229, 230

[90] Wufeng Xue, Lei Zhang, Xuanqin Mou, and Alan C Bovik. Gradient magnitude similarity deviation: A highly efficient perceptual image quality index. *IEEE Transactions on Image Processing*, 23(2):684–695, 2014. 229, 230, 244, 245, 247, 248, 249, 250

[91] Anmin Liu, Weisi Lin, and Manish Narwaria. Image quality assessment based on gradient similarity. *IEEE Transactions on Image Processing*, 21(4):1500–1512, 2012. 230

[92] Anish Mittal, Anush Krishna Moorthy, and Alan Conrad Bovik. No-reference image quality assessment in the spatial domain. *IEEE Transactions on Image Processing*, 21(12):4695–4708, 2012. 230

[93] Lixiong Liu, Bao Liu, Che-Chun Su, Hua Huang, and Alan Conrad Bovik. Binocular spatial activity and reverse saliency driven no-reference stereopair quality assessment. *Signal Processing: Image Communication*, 58:287–299, 2017. 230, 232, 243, 244

[94] Christopher C Yang and Sai Ho Kwok. Efficient gamut clipping for color image processing using LHS and YIQ. *Optical Engineering*, 42(3):701–712, 2003. 230

[95] Ke Gu, Leida Li, Hong Lu, Xiongkuo Min, and Weisi Lin. A fast reliable image quality predictor by fusing micro-and macro-structures. *IEEE Transactions on Industrial Electronics*, 64(5):3903–3912, 2017. 231

[96] Judith MS Prewitt. Object enhancement and extraction. *Picture processing and Psychopictorics*, 10(1):15–19, 1970. 231

[97] Paul B Hibbard. Binocular energy responses to natural images. *Vision Research*, 48(12):1427–1439, 2008. 231

[98] Rafik Bensalma and Mohamed-Chaker Larabi. A perceptual metric for stereoscopic image quality assessment based on the binocular energy. *Multidimensional Systems and Signal Processing*, 24 (2):281–316, 2013. 231

[99] Chathura Galkandage, Janko Calic, Safak Dogan, and Jean-Yves Guillemaut. Stereoscopic video quality assessment using binocular energy. *IEEE Journal of Selected Topics in Signal Processing*, 11(1):102–112, 2017. 231

[100] Gangyi Jiang, Haiyong Xu, Mei Yu, Ting Luo, and Yun Zhang. Stereoscopic image quality assessment by learning non-negative matrix factorization-based color visual characteristics and considering binocular interactions. *Journal of Visual Communication and Image Representation*, 46:269–279, 2017. 232, 245, 248, 249, 250

[101] Yucheng Zhu, Guangtao Zhai, Ke Gu, and Min Liu. Blindly evaluating stereoscopic image quality with free-energy principle. In *Circuits and Systems (ISCAS), 2016 IEEE International Symposium on*, pages 2222–2225. IEEE, 2016. 232

[102] Daniel V Meegan, Lew B Stelmach, and W James Tam. Unequal weighting of monocular inputs in binocular combination: Implications for the compression of stereoscopic imagery. *Journal of Experimental Psychology: Applied*, 7(2):143, 2001. 232

[103] Jonathan Harel, Christof Koch, and Pietro Perona. Graph-based visual saliency. In *Advances in neural information processing systems*, pages 545–552, 2007. 232

[104] Hamed Rezazadegan Tavakoli, Esa Rahtu, and Janne Heikkilä. Fast and efficient saliency detection using sparse sampling and kernel density estimation. In *Scandinavian Conference on Image Analysis*, pages 666–675. Springer, 2011. 232

[105] Lijuan Duan, Chunpeng Wu, Jun Miao, Laiyun Qing, and Yu Fu. Visual saliency detection by spatially weighted dissimilarity. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 473–480. IEEE, 2011. 232

[106] Xiaodi Hou, Jonathan Harel, and Christof Koch. Image signature: Highlighting sparse salient regions. *IEEE transactions on pattern analysis and machine intelligence*, 34(1):194–201, 2012. 232

[107] Ke Gu, Guangtao Zhai, Weisi Lin, Xiaokang Yang, and Wenjun Zhang. Visual saliency detection with free energy theory. *IEEE Signal Processing Letters*, 22(10):1552–1555, 2015. 232

[108] Yuzhen Niu, Yujie Geng, Xueqing Li, and Feng Liu. Leveraging stereopsis for saliency analysis. In *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*, pages 454–461. IEEE, 2012. 232

[109] Yuming Fang, Junle Wang, Manish Narwaria, Patrick Le Callet, and Weisi Lin. Saliency detection for stereoscopic images. *IEEE Trans. Image Processing*, 23(6):2625–2636, 2014. 232

[110] Claude Elwood Shannon. A mathematical theory of communication. *ACM SIGMOBILE mobile computing and communications review*, 5(1):3–55, 2001. 233

[111] Jagat Narain Kapur, Prasanna K Sahoo, and Andrew KC Wong. A new method for gray-level picture thresholding using the entropy of the histogram. *Computer vision, graphics, and image processing*, 29(3):273–285, 1985. 233

[112] Anush Krishna Moorthy, Che-Chun Su, Anish Mittal, and Alan Conrad Bovik. Subjective evaluation of stereoscopic image quality. *Signal Processing: Image Communication*, 28(8):870–883, 2013. 236, 237

[113] Junming Zhou, Gangyi Jiang, Xiangying Mao, Mei Yu, Feng Shao, Zongju Peng, and Yun Zhang. Subjective quality analyses of stereoscopic images in 3DTV system. In *Visual Communications and Image Processing (VCIP), 2011 IEEE*, pages 1–4. IEEE, 2011. 236, 237

[114] Feng Shao, Weijun Tian, Weisi Lin, Gangyi Jiang, and Qionghai Dai. Learning sparse representation for no-reference quality assessment of multiply distorted stereoscopic images. *IEEE Trans. Multimedia*, 19(8):1821–1836, 2017. 236, 237

[115] IEEE Standards Association Stereo Image Database. `http://http://grouper.ieee.org/groups/3dhf/`. 236, 237

[116] Rui Song, Hyunsuk Ko, and CC Kuo. MCL-3D: A database for stereoscopic image quality assessment using 2D-image-plus-depth source. *arXiv preprint arXiv:1405.1403*, 2014. 236

[117] Lutz Goldmann, Francesca De Simone, and Touradj Ebrahimi. Impact of acquisition distortion on the quality of stereoscopic images. In *Proceedings of the International Workshop on Video Processing and Quality Metrics for Consumer Electronics*, number MMSPL-CONF-2009-022, 2010. 236

[118] ZM Parvez Sazzad, Shouta Yamanaka, Yoshikazu Kawayokeita, and Yuukou Horita. Stereoscopic image quality prediction. In *Quality of Multimedia Experience, 2009. QoMEx 2009. International Workshop on*, pages 180–185. IEEE, 2009. 236

[119] Xu Wang, Mei Yu, You Yang, and Gangyi Jiang. Research on subjective stereoscopic image quality assessment. In *Multimedia Content Access: Algorithms and Systems III*, volume 7255, page 725509. International Society for Optics and Photonics, 2009. 236

[120] Hamid R Sheikh, Muhammad F Sabir, and Alan C Bovik. A statistical evaluation of recent full reference image quality assessment algorithms. *IEEE Transactions on image processing*, 15(11): 3440–3451, 2006. 238

[121] Yu Fan, Mohamed-Chaker Larabi, Faouzi Alaya Cheikh, and Christine Fernandez-Maloigne. On the performance of 3d just noticeable difference models. In *IEEE International Conference on Image Processing (ICIP)*, pages 1017–1021. IEEE, 2016. 241

[122] Zhou Wang, Eero P Simoncelli, and Alan C Bovik. Multiscale structural similarity for image quality assessment. In *The Thrity-Seventh Asilomar Conference on Signals, Systems & Computers, 2003*, volume 2, pages 1398–1402. Ieee, 2003. 245, 248, 249, 250

[123] Qiuping Jiang, Feng Shao, Weisi Lin, and Gangyi Jiang. Learning a referenceless stereopair quality engine with deep nonnegativity constrained sparse autoencoder. *Pattern Recognition*, 76:242–255, 2018. 245, 250

[124] Balasubramanyam Appina, Sameeulla Khan, and Sumohana S Channappayya. No-reference stereoscopic image quality assessment using natural scene statistics. *Signal Processing: Image Communication*, 43:1–14, 2016. 247, 249

[125] Guanghui Yue, Chunping Hou, Qiuping Jiang, and Yang Yang. Blind stereoscopic 3D image quality assessment via analysis of naturalness, structure, and binocular asymmetry. *Signal Processing*, 150:204–214, 2018. 248

[126] Kenny HB Voo and David BL Bong. Quality assessment of stereoscopic image by 3d structural similarity. *Multimedia Tools and Applications*, 77(2):2313–2332, 2018. 249

[127] Lixiong Liu, Bing Yang, and Hua Huang. No-reference stereopair quality assessment based on singular value decomposition. *Neurocomputing*, 275:1823–1835, 2018. 249

Résumé: Les grandes avancées des technologies stéréoscopiques/3Ds conduisent à une croissance remarquable de la quantité de contenue 3D dans diverses applications (par exemple, les domaines de divertissement et médicaux) grâce à une expérience visuelle d'utilisateur réaliste et immersive. Cependant, l'avènement de ces technologies a également apporté quelques défis techniques et des problèmes telles que l'évaluation de la qualité et la compression dû aux processus perceptuels de la perception binoculaire. Visant à évaluer et optimiser les performances des systèmes d'imagerie 3D en ce qui concerne leur capacité de stockage et leur qualité d'expérience, cette thèse se concentre sur l'investigation de la perception binoculaire selon deux perspectives. Dans la première partie, afin d'améliorer la compression et l'évaluation de la qualité de d'image 3D, notre travail de rechercher vise à explorer et modéliser la sensibilité du système visuel aux dégradations de l'image. Dans la deuxième partie, le travail de recherche vise à étudier les facteurs monoculaires et binoculaires affectant le jugement humain de la qualité 3D, puis à imiter ce dernier pour proposer une méthodologie robuste d'EQIS. Il est bien connu que le SVH ne peut pas détecter les modifications dans une image compressée si ces modifications sont inférieures aux seuils de différence juste notable (JND). Nous fournissons dans la première partie de la thèse une étude approfondie et une comparaison complète sur les modèles JND-3D existants basés sur l'analyse théorique, les expériences psychophysiques et l'application dans EQIS. De plus, nous proposons un nouveau modèle JND-3D basé sur des expériences psychophysiques, en tenant compte de la disparité binoculaire et des effets de masquage spatial du SVH. Des expériences subjectives confirment que le modèle proposé atteint une meilleure performance par rapport aux autres modèles JND-3D en termes de qualité perceptuelle sous le même niveau de bruit. La deuxième partie de cette thèse explore de nouvelles approches d'EQIS considérant les comportements de perception binoculaire (c.-à-d., la fusion binoculaire et la rivalité binoculaire) et la sensibilité visuelle de la SVH. Basés sur ces investigations, nous proposons des méthodes d'EQIS à partir de deux aspects différents. Tout d'abord, nous proposons deux métriques d'EQIS avec référence. La première métrique tient compte de la qualité de l'image cyclopéenne basée sur la perception binoculaire et de la qualité d'image de disparité. La seconde compte la qualité d'image cyclopéenne et la qualité d'image monoculaire basée sur d'une paire stéréoscopique. Deuxièmement, nous développons deux métriques d'EQIS sans référence basés sur les caractéristiques statistiques monoculaires du contraste local de l'image avec ou sans information de disparité. Des expériences approfondies sur diverses bases de données 3D démontrent que les quatre métriques proposées surpassent la plupart de l'état de l'art de l'EQIS, et atteignent une grande précision de prédiction de la qualité.

Mots clés: Contenue 3D, qualité d'expérience, perception binoculaire, système visuel humain, évaluation de la qualité de l'image stéréoscopique, différence juste notable.

Abstract: The great advances of stereoscopic or 3D technologies lead to a remarkable growth of the amount of 3D content in various applications (e.g., entertainment and medical domains) thanks to a realistic and immersive user viewing experience. However, the advent of these technologies has also brought some technical challenges and issues such as quality assessment and compression due to the complex perceptual processes of the binocular perception. Aiming to evaluate and optimize the performance of 3D imaging systems with respect to their storage capacity and quality of experience (QoE), this thesis focuses on the investigation of binocular perception from two different perspectives. In the first part, in order to improve the 3D image compression and quality assessment, our research work aims to explore and model the sensitivity of the human eyes to image impairments. In the second part, the research work is dedicated to the investigation of monocular and binocular factors affecting the human judgment of 3D quality, and then to mimic this judgment to propose a robust SIQA methodology. It is well-known that the HVS cannot detect the changes in a compressed image if these changes are lower than the just noticeable difference (JND) threshold. We provide in the first part of the dissertation a comprehensive overview and an exhaustive comparison on existing 3D-JND models based on theoretical analysis, psychophysical experiments and application in SIQA. In addition, we further propose a new 3D-JND model based on psychophysical experiments, accounting for monocular visual masking effects, binocular disparity, and visual saliency. Subjective experiments validate that the proposed model achieves better performance compared to other 3D-JND models in terms of perceptual quality under the same noise level. The second part of this dissertation explores new SIQA approaches considering binocular perception behaviors (i.e., binocular fusion and binocular rivalry) and visual sensitivity of the HVS. Based on these investigations, we proposed the SIQA methods from two different aspects. First, we propose two full-reference SIQA metrics. The first metric considers binocular-based cyclopean image quality and disparity map quality, the second accounts for binocular-based cyclopean image quality and stereopair-based monocular image quality. Second, we develop two no-reference SIQA metrics based on monocular statistical features of the image local contrast with and without disparity information. Comprehensive and thorough experiments on various publicly available SIQA databases demonstrate that the proposed four metrics outperform state-of-the-art SIQA methods, and achieve high prediction accuracy.

Keywords: 3D content, quality of experience, binocular perception, human visual system, stereoscopic image quality assessment, just noticeable difference.