Stereoscopic 3D Video Quality of Experience

Impact of Coding, Transmission and Display Technologies

Kun Wang

Mittuniversitetet
MID SWEDEN UNIVERSITY

Department of Information and Communication Systems
Faculty of Science, Technology and Media
Mid Sweden University

Licentiate Thesis No. 98
Sundsvall, Sweden
2013

Mittuniversitetet
Informations och kommunikationssystem
SE-851 70 Sundsvall
SWEDEN

ISBN 978-91-87103-72-8
ISSN 1652-8948

Akademisk avhandling som med tillstånd av Mittuniversitetet framlägges till offentlig granskning för avläggande av teknologie licentiatexamen fredagen den 17 May 2013 i M102, Mittuniversitetet, Holmgatan 10, Sundsvall.

Tryck: Tryckeriet Mittuniversitetet

To My Parents
To My Wife
To My Son

Abstract

Three-Dimensional (3D) videos are using their success from cinema to home entertainment markets such as TV, DVD, Blu-ray, video games, etc. The video quality is a key factor which decides the success and acceptance of a new service. Visual quality will have more severe consequences for 3D than for 2D videos, e.g. eye-strain, headache and nausea.

This thesis addresses the stereoscopic 3D video quality of experience that can be influenced during the 3D video distribution chain, especially in relation to coding, transmission and display stages. The first part of the thesis concentrates upon the 3D video coding and transmission quality over IP based networks. 3D video coding and transmission quality has been studied from the end-users' point of view by introducing different 3D video coding techniques, transmission error scenarios and error concealment strategies. The second part of the thesis addresses the display quality characterization. Two types of major consumer grade 3D stereoscopic displays were investigated: glasses with active shutter (SG) technology based display, and those with passive polarization technology (film patterned retarder,FPR) based display.

The main outcomes can be summarized in three points: firstly the thesis suggests that a spatial down-sampling process working together with high quality video compressing is a efficient means of encoding and transmitting stereoscopic 3D videos with an acceptable quality of experience. Secondly, this thesis has found that switching from 3D to 2D is currently the best error concealment method for concealing transmission errors in the 3D videos. Thirdly, this thesis has compared three major visual ergonomic parameters of stereoscopic 3D display system: crosstalk, spatial resolution and flicker visibility. The outcomes of the thesis may be of benefit for 3D video industries in order to improve their technologies in relation to delivering a better 3D quality of experience to customers.

Keywords: 3D, 3D TV, video quality, Quality of Experience, video distribution, crosstalk, flicker.

Acknowledgements

My journey to the Lic. or later Ph.D degree would never have been possible without the help of many people. It is my great pleasure to take this opportunity to thank them for the support and advice that I have received over the past years.

Firstly, I would like to express my deepest gratitude to Prof. Kjell Brunnström at Acreo Swedish ICT for offering me the opportunity to pursue my doctoral studies under his supervision. His tremendous support and guidance proved to be useful in all aspects of my researches and life.

I would further like to thank my main supervisor Prof. Mårten Sjöström at Mid Sweden University. His wide knowledge and experience have been always of great value for me. His help and advice have provided a good basis for my research work and writing of thesis. I would also like to thank my co-supervisors: Dr. Ulf Jennehag and Dr. Roger Olsson for their fruitful discussions and supervision. Special thanks to Dr. Mikael Gidlund who inspired me to start my PhD journey. Thanks also to my colleagues at Mid Sweden University, Dr. Sylvain Tourancheau, Sebastian Schwarz, Yun Li, Suryanarayana Muddala, Mitra Damghanian, for their help and suggestions on my work.

Special thanks to Assoc. Prof. Marcus Barkowsky and Prof. Patrick Le Callet at University of Nantes, France, for their unreserved knowledge sharing, constructive ideas and enormous support on my research work.

I would like to thank Niclas Rydell, Steve Fuller at TCO development, Valentin Kulyk, Mats folkesson at Ericsson research for their support and precious advice.

My great appreciation to all my colleagues at Acreo Swedish ICT, Dr. Jie Li, Dr. Claus Popp Larsen, Anders Gavler, Dr. Mikhail Popov, Dr. Anders Djupsjöbacka, Andreas Aurelius, Dr. Tianhua Xu, Miu Yoong Leong, Dr. Marco Forzati, Prof.Gunnar Jacobsen, Dr. Evgeny Vanin, Dr. Qin Wang, Dr. Anders Berntson, Dr. Pierre-yves Fonjallaz, Dr. Bertrand Noharet, Dr. Stephane Junique, Dr. Zhangwei Yu, for their inspiring, encouragement, and creation of a pleasant working environment. Special thanks to Börje Andrén for his support, sharing of his expertise in display quality area and his humor.

Furthermore I would like to thank my friends who gave me help and encouragement during my studies: Jianming Zeng, Can Huang, Jia Yang, Bing Zhan, Kees and Petra van Beek, Sjef Nijhuis, Giuliana Macina, Patryk Urban, Shun Yu.

Last but not least, I would like to thank my parents Lili Lu, Bin Wang, my wife Peng Zhao and my son Peisheng Wang for their continuous supporting and encouraging me throughout my whole life.

Contents

List of Papers

This thesis is mainly based on the following papers, herein referred by their Roman numerals:

I M. Barkowsky, **K. Wang**, R. Cousseau, K. Brunnström, R. Olsson, and P. Le Callet. Subjective Quality Assessment of Error Concealment Strategies for 3DTV in the presence of asymmetric Transmission Errors. In *IEEE conference Packet Video Workshop in HongKong*, 2010.

II **K. Wang**, M. Barkowsky, R. Cousseau, K. Brunnström, R. Olsson, P. Le Callet and M. Sjöström. Subjective evaluation of HDTV stereoscopic videos in IPTV scenarios using absolute category rating. In *IS& T/SPIE Electronic Imaging*, pp. 78631T-78631T, 2011.

III **K. Wang**, M. Barkowsky, K. Brunnström, M. Sjöström, R. Cousseau, P. Le Callet. Perceived 3D TV Transmission Quality Assessment: Multi-Laboratory Results Using Absolute Category Rating on Quality of Experience Scale. In *IEEE Transactions on Broadcasting*, vol.PP, no.99, pp.1, 0, 2012.

IV B. Andrén, **K. Wang**, K. Brunnström. Characterizations of 3D TV: Active vs Passive. In *Proceedings of SID Symposium Digest of Technical Papers*, vol. 43, no. 1, pp. 137-140, 2012.

The author has also contributed to the following publications which are not included in this thesis:

1. S.Tourancheau, **K. Wang**, J. Bulat, R. Cousseau, L. Janowski, K. Brunnström, and M. Barkowsky. Reproducibility of crosstalk measurements on active glasses 3D LCD displays based on temporal characterization. In *IS& T/SPIE Electronic Imaging*, pp. 82880Y-82880Y, 2010.
2. M. Barkowsky, S.Tourancheau, K. Brunnström, **K. Wang**, B. Andrén. Crosstalk Measurements of Shutter Glasses 3D Displays. In *Proceedings of SID Symposium Digest of Technical Papers*, Vol. 42, No. 1, pp. 812-815, 2012.
3. B. Andrén, **K. Wang**, K. Brunnström A comparison of visual ergonomic measurements between active and passive 3D TV. In *SID Digest of EuroDisplay*, 109, 2011.

4. K. Brunnström, I. Sedano, **K. Wang**, M. Barkowsky, M. Kihl, B. Andrén, P.Le Callet, M. Sjöström, and A. Aurelius. 2D No-Reference Video Quality Model Development and 3D Video Transmission Quality. In *Proceedings of the 6th International Workshop on Video Processing and Quality Metrics for Consumer Electronics - VPQM*, 2012.

5. A. Perkis, A.; J. You, L. Xing, T. Ebrahimi,F. De Simone, M. Rerabek,P. Nasiopoulos, Z.Mai, M.Pourazad, K. Brunnström, **K. Wang**, B. Andrén. Towards certification of 3D video quality assessment. In *Proceedings of the 6th International Workshop on Video Processing and Quality Metrics for Consumer Electronics - VPQM*, 2012.

6. M. Barkowsky, K.Brunnström, T. Ebrahimi, L. Karam, P. Lebreton,P. Le Callet, A. Perkis, A. Raake,M. Subedar, **K. Wang**, L. Xing, J. You. Book chapter: Subjective and objective visual quality assessment in the context of stereoscopic 3DTV. In *Book 3D-TV System with Depth-Image-Based Rendering: Architecture, Techniques and Challenges*, Springer, 2012.

List of Figures

Terminology

Abbreviations and Acronyms

2D	Two Dimensional
3D	Three Dimensional
ACR	Absolute Category Rating
ACR-HR	Absolute Category Rating with Hiden Reference
ARQ	Automatic Repeat-reQuest
bps	Bits Per Second
BT	Broadcasting service (television)
DSCQS	Double Stimulus Continuous Quality Scale
DSIS	Double Stimulus Impairment Scale
Exp	Experiment
FEC	Forward Error Correction
FPR	Film Pattern Retarder
HD	High Definition
HEVC	High-Efficiency Video Coding
HFES	Human Factors and Engineering Society
HRC	Hypothetical Reference Circuit
HVS	Human Visual System
ICDM	International Committee for Display Metrology
IP	Internet Protocal
ITU	International Telecommunication Union
JND	Just Noticeable Difference
LCD	Liquid Crystal Display
MOS	Mean Opinion Score
MVC	Multi-View Coding
PVS	Processed Video Sequence
QoE	Quality of Experience
QoS	Quality of Service
Ref	Reference video
S3D	Stereoscopic Three Dimension

SG	Shutter Glasses
SRC	Source video
SSQ	Simulator Sickness Questionnaires
SVC	Scalable Video Coding
TSCF	Temporal Contrast Sensitivity Function
VESA	Video electronics standards association

Mathematical Notation

CL	Crosstalk through the Left view
BW	Luminance measured through left channel when the left view input is Black and right view is White
BB	Luminance measured through left channel when both left and right view are Black
WB	Luminance measured through left channel when the left view input is White and right view is Black
i	The grey level of the left channel
j	The grey level of the right channel
$L_{i,j}$	Luminance measured through left channel when left view input is grey level "i",and right view input is grey level "j"
$L_{i,i}$	Luminance measured through left channel when left view input is grey level "i",and right view input is grey level "i"
$L_{j,i}$	Luminance measured through left channel when left view input is grey level "j",and right view input is grey level "i"
C_m	Contrast modulation
L_{255}	Luminance of white image (grey level 255)
L_0	Luminance of black image (grey level 0)
n	Pixel width
n_r	Calculated grille line width in pixels for which the value of Cm is estimated by linear interpolation to be equal to the contrast modulation threshold Ct
$C_m(n)$	Contrast modulation of n pixel width grille
$C_m(n+1)$	Contrast modulation of n+1 pixel width grille
C_t	Threshold contrast modulation
N_{adr}	Number of addressable pixels
R	Resolution

Chapter 1

Introduction

Three-Dimensional (3D) image viewing was first introduced in the 19 century by a British scientist, Sir Charles Wheatstone, who invented the stereoscope. Nowadays inspired by the rapidly increasing popularity of 3D movies, 3D videos and applications have been included within many fields, e.g. telecommunication, video conferencing, advertising and exhibitions, health care, medical diagnosis, city planning, mechanic and architecture designing, etc. This is particularly the case in the present home entertainment market 3D related products, for example 3D TV, DVD, Blu-ray, video games, mobile phones, etc. are becoming more and more popular.

In relation to presenting 3D videos, a number of techniques have been invented [Alatan *et al.*, 2007], e.g. stereoscopic [Bruls *et al.*, 2007], multi-view [Merkle *et al.*, 2007], volumetric [Favalora, 2005], holograph [Yoshikawa & Yamaguchi, 2012]. The stereoscopic 3D (S3D), the focus of this thesis, is most widely used in current movie industry and 3DTV broadcasting. A 3DTV broadcasting service has already been introduced in several countries over recent years. Internet online video service providers, e.g. Youtube, have also offered S3D video services over the internet. The 3D video is ever closer to the lives of ordinary people.

1.1 Background and Problem Motivation

1.1.1 3D visual perception

The perception of 3D depth from stereoscopic 3D videos is based on the manner in which the human brain and eyes work. Most humans beings have two eyes looking at the world from two slightly different, horizontally spaced angles [Wheatstone, 1838]. In a similar manner, as Fig. 1.1 shows, the S3D videos present viewers with two images (i.e. two perspectives of the same scene) having a slight spatial shift of viewpoint, a.k.a. binocular disparity. Each eye will only see one of the two pictures, the Human Visual System (HVS) will then make use of the disparity, to create a sen-

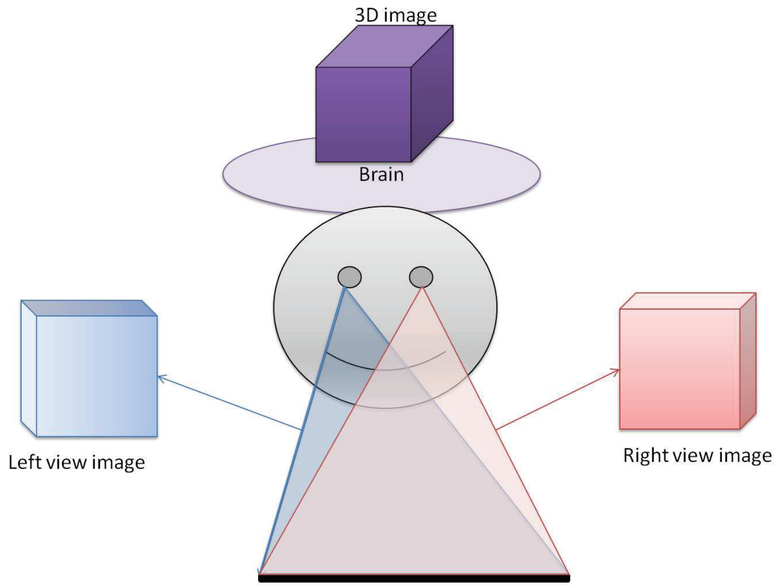


Figure 1.1: Depth perception from stereoscopic 3D. Two images representing two perspective of the same objects are shown to human eyes separately, human visual system will interpret the difference (disparity) between these two images and generate depth sensation in the brain.

sation of 3D depth in the brain. However this stereoscopic viewing is an unnatural viewing situation and any incorrect presentation will cause visual fatigue or even a headache and nausea etc. [Hoffman *et al.*, 2008].

1.1.2 3D video distribution chain

A whole 3D TV end-to-end system can be illustrated as a chain as shown in Fig. 1.2.

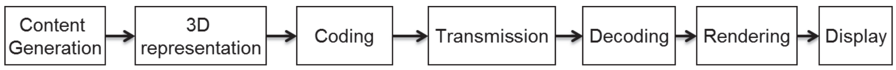


Figure 1.2: 3D video distribution chain.

Firstly, a 3D content can be generated from either the capturing of a real world scene by cameras, or by computer generated imagery, or by being converted from 2D content [Grau, 2004; Tam & Zhang, 2006]. Secondly, the content can be stored and represented in different 3D formats, e.g. stereoscopic [Bruls *et al.*, 2007], holograph [Yoshikawa & Yamaguchi, 2012] and multi-view(multiple stereoscopic view pairs) [Alatan *et al.*, 2007; Merkle *et al.*, 2007; Ozaktas & Onural, 2007]. The transmission of 3D content will require more bandwidth than 2D videos, and the network itself also has different constraints on the bandwidth depending on network conditions, which may vary locally and over time, therefore it is necessary to compress the formatted 3D video before sending it through the network. After transmission, the video will be decoded and rendered according to the 3D format, then it will be shown on 3D displays and, finally, users will perceive the videos through HVS and thus generate the 3D depth sensation in the brain.

1.1.3 Video Quality

The video quality is important when delivering videos, especially for new services such as 3DTV, and an experience of any unpleasant visual quality will have more severe consequences for 3D than 2D e.g. eye-strain, headache and nausea (”sea-sickness”) [Ukai & Howarth, 2008]. Therefore, the quality of experience (QoE) is a key parameter which will dominate customers' acceptance of 3DTV.

QoE is defined in [Le Callet *et al.*, 2012]: ”the degree of delight or annoyance of the user of an application or service. It results from the fulfillment of his or her expectations with respect to the utility and / or enjoyment of the application or service in the light of the user's personality and current state.”

In relation to 3D video distribution, the QoE can be influenced by quality artefacts introduced in each stage of the above mentioned 3D video distribution chain. For instance, in the content capturing stage, the misalignment of cameras and improper calibrations may give rise to artefacts such as geometry distortions, temporal offset and colour mismatch, etc. [Docherty & Koch, 1993]. During the coding process, artefacts mainly are caused by coding techniques and the degree of compressing [Punchihewa & Bailey, 2002]. The transmission errors, e.g. packet loss and delay, will certainly affect the delivered videos quality [Mu *et al.*, 2009; Meenowa *et al.*, 2010; Pinson *et al.*, 2010]. Finally, the displays will introduce artefacts depending on the technologies, such as crosstalk [Woods, 2011], flicker [Watson & Ahumada, 2012], and resolution [Heynderickx & Kaptein, 2009] [Kim & Banks, 2012] etc.

1.2 Overall Aim

The aim of this thesis is to investigate stereoscopic 3D video quality of experience that can be influenced during the 3D video distribution chain, especially in the coding, transmission and display stages. The outcome of this thesis is to produce tools or recommendations for the 3D video industries in order to improve their technologies such that a better delivery of the 3D quality of experience to customers can be achieved.

1.3 Scope

This thesis has its focus only on the stereoscopic 3D video quality. The first part of the thesis is restricted to the 3D video coding and transmission quality over IP based networks, which refines the related research work referring to the source coding, transmission, and decoding stages in Fig. 1.1. The second part of the thesis addresses the last stage of the 3D distribution system chain, specifically the display quality characterization. Two types of major consumer grade 3D stereoscopic displays (active shutter eye-glass technology based and passive polarized eye-glass film type patterned retarder (FPR) technology based display) are investigated in this work.

1.4 Concrete and Verifiable Goals

While the additional dimension of a 3D video provides the user with an immersed experience it also brings new issues and challenges.

1.4.1 Stereoscopic 3D video coding and transmission quality

Problem 1: Today the network bandwidth and resources for transmitting videos normally is limited, especially for full HD resolution S3D videos (close to double amount of the data transmission bitrate required for HD 2D videos). Therefore, how to efficiently deliver a 3DTV service has become one of the major concerns. Compress coding, on one hand, is commonly used to reduce the required video transmission bitrate; on the other hand, it inevitably introduces visual artefacts. Therefore, there is always a trade-off between the amount of bandwidth that can be saved (compression degree) and the amount of degradation on the visual quality.

Goal 1: To investigate the impact of coding artefacts on users' 3D QoE and identify an efficient coding approach and compression degree for 3D video delivery with an acceptable QoE.

Problem 2: Apart from the artefacts introduced by the source coding, the transmission network in itself often introduces errors e.g. delay or packet loss. The impacts of network errors on 2D video quality have been discussed in many studies e.g. [Mu *et al.*, 2009; Meenowa *et al.*, 2010; Pinson *et al.*, 2010]. In relation to 3D case there is still a lack of research. As the perception of S3D videos is different from that for 2D videos, the influence of network errors on the user experience with regards to 3D quality will be more complex than is the case for 2D videos.

Goal 2: To investigate the impact of transmission errors on the 3D video quality of experience and propose an suitable error concealment method for 3D videos.

Problem 3: Another challenge for 3D videos is how to accurately measure and evaluate the perceived video quality by the end-users. For 2D videos, many subjective measurement methods have been standardized [Recommendation, 2012b; ITU-T, 2008]. However when dealing with 3D videos, the concept of QoE is required to be expanded by considering multi-dimensional aspects. In addition to the traditional concept of video quality, other aspects such as depth quality, visual comfort, etc. must also be considered. Therefore, it has become necessary to develop a reliable means of measuring 3D QoE.

Goal 3: To investigate the subjective methods for measuring users' 3D QoE and work towards establishing a reliable subjective test method for 3D videos.

1.4.2 Stereoscopic 3D display Quality

Problem 4: The traditional concept of sharpness, colour, resolution, etc in 2D displays are not any more equal than that perceived by a human in relation to the 3D watching condition; there are, additionally, new issues specifically introduced by 3D display technologies, for example crosstalk and flicker. These will all affect the performance of 3D displays and thus the user's 3D experience. Therefore all these display characteristics must be given significant attention and be addressed with regards to 3D displays.

Goal 4: To characterize the stereoscopic 3D display system quality, investigate and verify the measurement methods for 3D display systems.

1.5 Outline

This thesis is organized as follows: an introduction of 3D video quality and evaluation methods is provided in Chapter 2. The investigation of coding and transmission impact on the 3D video quality of experience is discussed in Chapter 3, the display quality study of glasses with both active shutter based 3DTV and also with passive polarization based frame pattern retarder 3DTV is described in Chapter 4. A conclusion and future work is given in Chapter 5. Author's publications in the related fields are appended at the end of thesis.

1.6 Description of Contributions

Four publications (one journal paper, three conference papers) with the author's contributions are included in this thesis.

- Paper I: The author actively contributed to the development of the ideas, preparation and design of the subjective experiment, and data analysis, 30% of the writing of the manuscript and the final presentation at conference were conducted by the author.

- Paper II: The author was involved in the idea development, video processing, simulation, subjective experiment, data analysis and additionally the paper was written by the author with suggestions and comments from co-authors .

- Paper III: Based on the experiment results from paper I and II, the author contributed to the idea development, data analysis and writing of the paper with suggestions from co-authors.
- Paper IV: The author contributed to the display system measurement, data analysis and the 50% of the writing of the manuscript.

Chapter 2

3D Image and Video Quality

This chapter starts with an introducing to the basic knowledge with regards to the 3D perception related human vision system in section 2.1, then follows with video and image quality model and its evaluation methodologies in section 2.2 and section 2.3.

2.1 3D perception of Human visual system

Human perception of a 3D world is based on various depth cues which can be divided into monocular and binocular cues.

The monocular depth cues refer to the depth sensation that can be perceived by a human even with only one eye. The depth perception from 2D images or videos are mainly due to monocular depth cues. The major monocular cues are:

- Motion parallax: when a subject moves with respect to the environment or vice versa, a different angular velocity exists between the line of sight to a fixed object and the line of sight to any other objects in the visual field [Graham, 1965]. For example when running in the street, the closer objects pass quickly while far away objects appear to be stationary.
- Occlusion: closer objects block other objects that are behind them. One of the important occlusion cues is T-junction, that is edges of farther objects disappear when closer objects are in front.
- Accommodation: the ability of the eyes to change the optical power (focus length) which is used by human brain for interpreting depth.
- Perspectives: Parallel lines appear to converge at a faraway point i.e. vanishing point. For example looking at railway lines, the two lines appear to be closer and closer with the increasing visual distance.
- Shadow: the shadow provides cues with regards to the shape of the objects and their position in space.

- Size: closer objects appear bigger than distant objects.

Fig. 2.1 shows an example of 3D depth perception from monocular cues. The perspective cue and size cue cause people to feel that the bigger objects at the left side are closer to them, and from left to right objects are becoming further and further inwards the paper. The shadow cue provides the feeling that the text objects have thickness (depth) and are standing onside the paper. The occlusion cue (the small green tree logo at right side is slightly blocked by the text ”o”) provides the feeling that the text ”o” is in front of the tree.

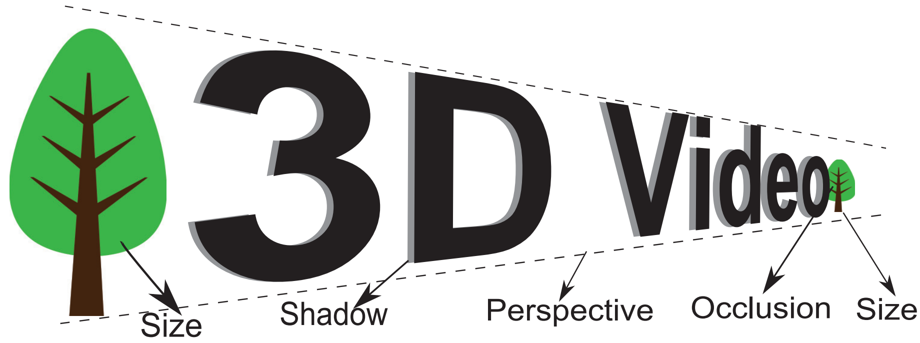


Figure 2.1: An example of depth perception from monocular cues: perspectives,shadows and size.

When a comparison is made in relation to the monocular cues described above, the more powerful depth sensation is constructed by binocular cues [King, 2009] that are perceived by two eyes looking at the scene from two slightly different, horizontally spaced angles. The stereoscopic 3D videos make use of this principle by presenting viewers with two images which have a slight spatial shift in relation to their viewpoint i.e. two perspectives of the same view, that give rise to visual retinal disparity. Each eye will only see one of the two pictures. The Human Visual System (HVS) will then group objects together in the two images, extract distance information from them and then create a sensation of 3D depth. However this type of stereoscopic viewing condition provided from the current S3D system is an unnatural viewing situation as compared to reality, and any improper presenting may lead to visual discomfort such as eye-strain, nausea,etc. one of the major differences between the natural 3D viewing condition and stereoscopic 3D is the conflict between accommodation and vergence [Schreer *et al.*, 2005], as shown in Fig. 2.2.

If the display is taken plane as a reference position, when a 3D depth sensation is perceived in front of the display (pop out from the screen, crossed disparity), two eyes move inward to each other, this is called convergence. Whereas the 3D objects are perceived behind or inside the display (uncrossed disparity), two eyes move outwards to each other so as to compare the eye position on the reference plane, i.e. eye diverging from the display plane. In a natural viewing situation when eyes verge to an object, the accommodation is also adjusted to focus on the object they have fixed on at the same point. However for stereoscopic 3D view condition, this becomes

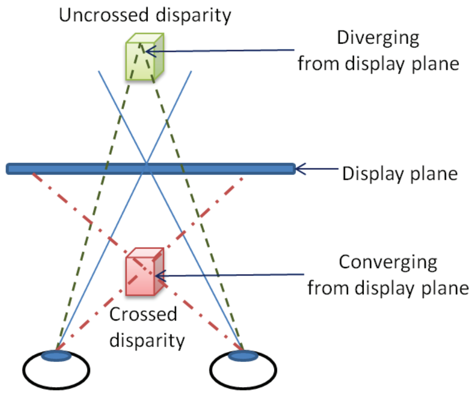


Figure 2.2: Human vision system accommodation and vergence in a 3D watching condition.

a conflict, because the accommodation makes the eyes focus on the display plane while the vergence of the object is in front of or behind the display [Suryakumar, 2005]. On the contrary a recent research from Japan [Hori *et al.*, 2011] shows the conflict between accommodation and vergence does not occur and therefore there is no difference between natural 3D vision and stereoscopic 3D vision in terms of accommodation and convergence.

In addition to the depth sensation and visual comfort issues raised by S3D videos, the traditional concept of image and video quality is different from the 2D to the 3D case. For example, the perceived resolution or detailedness of S3D videos may affect the users' experience differently when compared to the same image characteristics in traditional 2D videos. The research conducted by Heynderickx [Heynderickx & Kaptein, 2009]shows that a human can perceive more details from 3D images than 2D images with the same spatial resolution.

2.2 Image and Video quality model

Image or video quality is the degree of excellence of the image or video [Dictionary, 1989]. The evaluation of the quality can normally be divided into two classes: subjective evaluation and objective evaluation, as illustrated in Fig. 2.3.

The subjective evaluation of video quality is based on the end-user's subjective experience of the watched video, called quality of experience (QoE) [Le Callet *et al.*, 2012]. Although the individual subjective opinions from different users may vary, the mean of a panel of observers are usually a stable and reliable measure. However the subjective method is time consuming and expensive. The Objective quality evaluation method is a mathematical model which can automatically calculate, evaluate and predict video quality. It is developed, on the one hand, to overcome the disadvantage of the subjective method, while on the other hand, it is to approximate the accuracy and robustness of the subjective results as much as is possible.

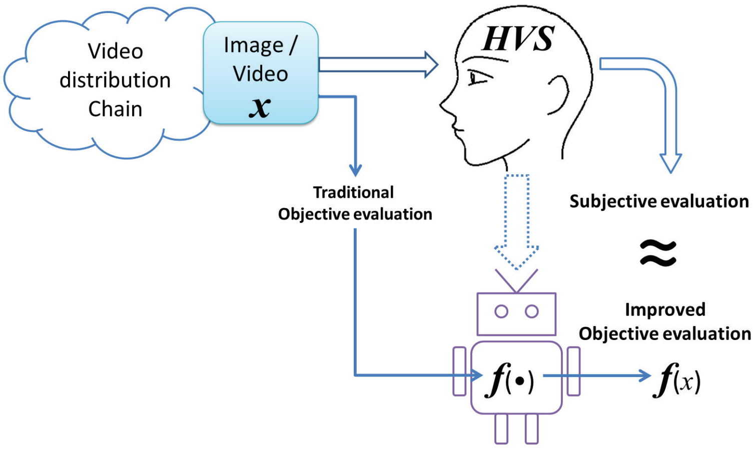


Figure 2.3: An overview of video quality evaluation methods, subjective evaluation, traditional objective evaluation and improved objective evaluation with consideration of human vision system.

The traditional objective model, as shown in Fig. 2.3, measures video quality from the physical properties of the video itself, and thus the measuring is directly based on the video. Therefore, the model is not complete as it misses the consideration of the human visual system. An improved version is to add an additional model which could simulate HVS, and takes the video physical properties as inputs, thus the output of the improved objective model, ideally, can be close to the subjective results.

Engeldrum [Engeldrum, 2000] developed a model to evaluate image / video quality. The model describes a relationship between quality of experience and technology variables into several measurable steps as shown in Fig. 2.4. The QoE can be influenced by the technology variables during each stage of 3D video distribution chain. For instance, in the capturing stage the technology variables can be the camera type, the baseline distance of stereoscopic camera, etc. The coding technology variables can be the coding techniques, and the degree of compression. In transmission stage, the technology variables can be different networks, streaming technologies and transmission systems. For the display, the variables could involve autostereoscopic technology, using glasses which are active shutter based or passive polarization based stereoscopic display technology, etc. In order to discover a reliable way of evaluating perceived video/image quality, it is important to determine a relationship between the quality and technology variables. However it is difficult to find a direct and reliable link between them because the QoE involves the interaction between human visual characteristics and technology variables. Engeldrum's model breaks the relationship between QoE and technology variables into several measurable steps, as shown by the square block shown in Fig. 2.4. There are also three ellipses that describe the relationship between the steps (square blocks). These ellipses can normally be mathematically modelled. For example, the system model

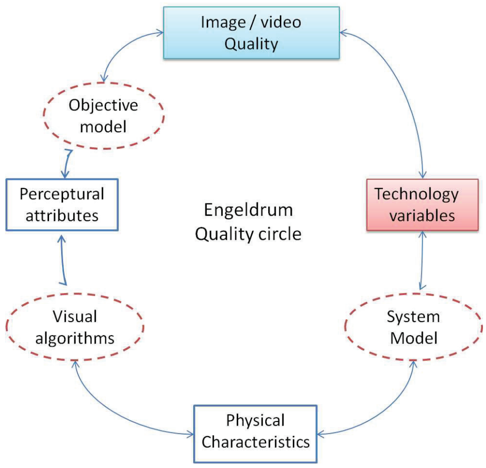


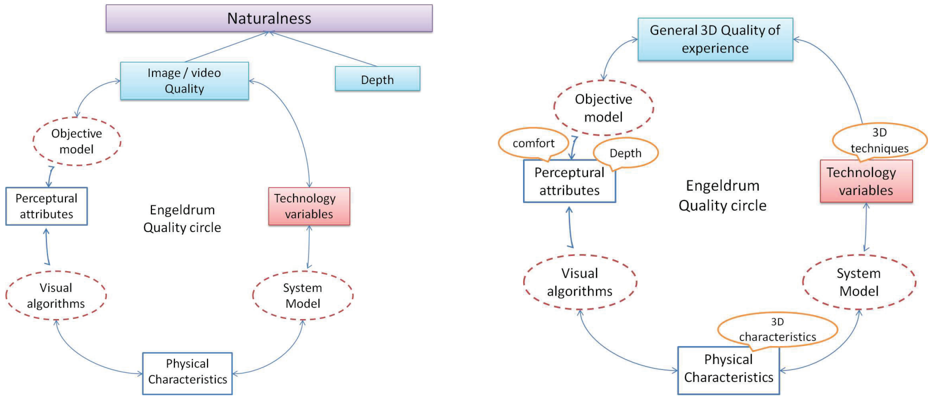
Figure 2.4: Engeldrum's image/video quality circle model breaks the relationship between video quality of experience and technology variables into two measurable steps(square blocks): physical characteristics and perceptual attributes.

can be used to predict physical parameters from the technology variables; according the knowledge of HVS, visual algorithms can be developed to simulate how a human visual system translates physical image characteristics into perceptual attributes; finally, objective quality model can be developed to predict customer's quality rating by taking the input from the subjective perceptual attributes.

S3D video is based on a pair of 2D videos, therefore it has certain similarities with the 2D video. Many of the developed 2D quality algorithms can also be partially applied to S3D videos. However, due to the 3D depth perception of HVS, there are differences between 2D videos and S3D videos. The new features and issues that 3D videos have brought must be addressed carefully when developing S3D quality models. Some new features in 3D video, particularly depth perception due to binocular disparity, may bring positive add-on values to the perceived video quality. This added value may be expressed in terms of sense of presence and naturalness [IJsselsteijn *et al.*, 2000; Seuntiëns *et al.*, 2005]. Some negative values can be also brought in by S3D such as visual discomfort [Ukai & Howarth, 2008]. Therefore it is important for 3D quality model to include all these 3D features.

The studies [IJsselsteijn *et al.*, 2000; Seuntiens *et al.*, 2003, 2006, 2007] showed that during the subjective evaluation, when using the traditional scale "image/video quality", the added value of 3D depth was not visible from viewer's voting. Kaptein etc. [Kaptein *et al.*, 2008] based on Engeldrum's model, proposed a new model for S3D quality evaluation as shown in Fig. 2.5(a).

In Kaptein's model "naturalness" is used to assess the overall quality of S3D images or videos. The traditional term of "image quality" from the original Engeldrum's model becomes more related to the 2D quality rather than an overall 3D experience, therefore, in addition to the "image / video quality", other aspects of



(a) (b)

Figure 2.5: Quality models for 3D images/videos. (a)Kaptein's model: it separates general 3D QoE into several individual scales, it keeps original Engeldrum's model for presenting 2D QoE, and lists add-on 3D features separately. The "Naturalness" presents general 3D QoE. (b)General 3D quality experience model: it uses only one scale to assess the general 3D QoE, 3D features are added into each steps of Engeldrum's model.

S3D videos, such as 3D depth, have to be considered separately.

An alternative means of assessing the S3D video quality of experience is shown in Fig. 2.5(b). The term of "General 3D quality of experience" is used in the quality circle to express the overall performance of S3D videos. The 3D features are integrated as a part of the perceptual attributes, and together with other 2D attributes, end-users can evaluate their overall video experience according to these perceptual attributes. These quality models can be used for the development of subjective measurement methods for the evaluation S3D video QoE, and the detailed discussion is described in section 2.3.2.

2.3 Subjective Evaluation

In this section, some major standard subjective evaluation methods for 2D videos are described in 2.3.1, and then the development of subjective methods for 3D videos are discussed in 2.3.2.

2.3.1 Standard Subjective Evaluation

Subjective assessment is commonly used to measure users' quality of experience. For the evaluation in 2D, many standards, which specify test procedures, viewing conditions and data analysis methods have been used over the years in small and large scale evaluations, e.g. by the Video Quality Experts Group (VQEG) [VQEG

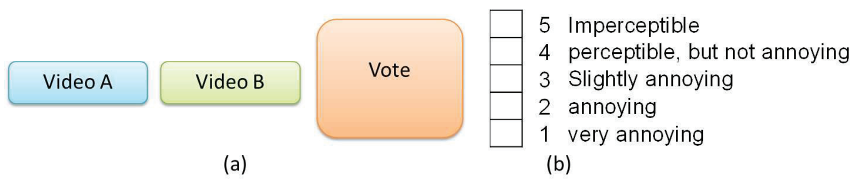


Figure 2.6: DSIS subjective method. (a)presentation sequence: the reference video, "Video A", is always shown to subjects in the first order then follows with "video B" which need to be assessed by the subjects. (b)voting scale: subjects vote on the "video B" and compare with "video A", the five-level scale ranges from 1 ("very annoying") to 5 ("imperceptible" difference).

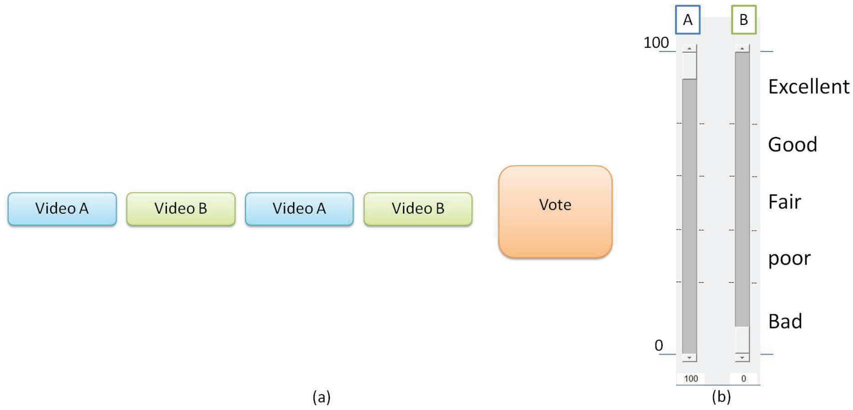


Figure 2.7: DSCQS subjective method. (a)presentation sequence, the videos are shown to the subjects sequentially with one repetition without notice of which one is reference video. (b)voting scale, subjects give their scores on both videos, the scale ranges from 0 (worst QoE) to 100 (best QoE)

et al., 2009], ITU-R Recommendation BT.500-13 [Recommendation, 2012b] and ITU-T P.910 [ITU-T, 2008]. The three most common used test procedures are: Double Stimulus Impairment Scale (DSIS), Double Stimulus Continuous Quality Scale (DSCQS), and Absolute Category Rating (ACR). These are described as follows:

- The Double Stimulus Impairment Scale (DSIS)is defined in the ITU-R Recommendation BT.500- 13 [Recommendation, 2012b] as indicated in Fig. 2.6. Pairs of videos are presented to subjects. The first "Video A" in Fig. 2.6 is always an unimpaired video as a reference and "video B" is an impaired video which is to be judged against the reference. Subjects rate the quality of "video B" according to the amount of impairment they perceive. The rating normally uses a five level impairment scale, where the highest score, five, means the processed video is indistinguishable from the reference, and the lowest score indicates very annoying impairment.

MOS	Quality
5	Excellent
4	Good
3	Fair
2	Poor
1	Bad

Figure 2.8: ACR subjective method voting scale, subjects votes on every watched videos in a five-level scale.

- The Double Stimulus Continuous Quality Scale (DSCQS) as specified in ITU-R BT.500-13 [Recommendation, 2012b] is illustrated in Fig. 2.7. In contrast to DSIS, where the unimpaired reference video is always presented first, in the DSCQS method subjects see a degraded sequence and its corresponding unimpaired reference sequence in a random order and the observers are not informed which one is the reference in the video. The videos are shown to the subjects sequentially with one repetition as shown in Fig. 2.7. After watching the video pairs twice, observers are asked to give their opinion scores on each of the videos. The opinions are on a continuous quality scale from 0, representing the ”worst” quality, to 100, representing the most ”excellent” quality.

- The Absolute Category Rating (ACR) method is standardized in ITU-T P.910 [ITU-T, 2008]. Unlike DSCQS and DSIS, ACR is a single stimulus method. Viewers only watch a video once and then give their rating on the watched videos. ACR-HR (absolute category rating with hidden reference video) method is a variation of the standard ACR. The hidden unimpaired reference video is mixed together with other processed videos, then all the Processed Video Sequences (PVS) are presented to viewers without knowing which one is the reference. The voting scale of ACR method is illustrated in Fig. 2.8, where ”1” presents bad quality and ”5” presents excellent quality.

  Since each video in the test set plays only once, the single stimulus based method is time-saving as compared to double stimulus based methods, and therefore is capable of collecting more votes from subjects. This allows for the testing of a broader variety of different stimuli conditions. The double stimulus based methods are time consuming because the same reference sequences are played many times, however, it make it easier for the viewers to be able to distinguish small differences in video quality. Hence, there is a trade-off between the accuracy gain from double stimulus methods and the time gain from single stimulus methods. The selection of test methods should be made according to the different application requirements.

2.3.2 Subjective Evaluation for 3D Video

Although it is the case that for 2D videos many subjective evaluation methods have been standardized, in relation to 3D videos,determining a reliable way of measuring end-users 3D experience is still an ongoing process in many standards organizations. One international recommendation BT 1438 [Recommendation, 2000] concerning the subjective assessment of stereoscopic television pictures was published in 2000. This document suggests that BT.500-13 [Recommendation, 2012b] can be applied for S3D subjective assessment methods, but it does not identify and address the type and visibility of artefacts peculiar to stereoscopic images. Another recommendation ITU-R BT.2160 [Recommendation, 2012a] suggests that visual comfort, image quality and depth quality are major perceptual dimensions which must be assessed. However, it does not offer any means of measuring these three major perceptual dimensions.

The present subjective evaluation of 3D images and video sequences today is, in general, based on two trends.

- The first evaluation method evaluates the overall visual experience of the 3D videos, where subjects consider both 2D and 3D perceptual attributes all together as judgment criteria. Hence, only one dimension scale is voted on by the subjects. This method matches the quality model proposed in Fig. 2.5(b). The one scale model, ideally, should measure the complete general 3D video quality of experience, which is useful, particularly when assessing the degree to which 3D technologies outperform or underperform the 2D technology in general. However, in reality, it is difficult to determine what is actually being measured and what quality aspects subjects are taking into account when they are voting during the subjective test. 3D videos have only recently been introduced to the consumer market, and for the majority of end-users 3D is still new, thus the concept of video quality may still stay, in relation to their perception, with 2D video quality with which they are familiar. Therefore, another method for guiding subjects to discover new features in 3D videos is described below.

- The other method of assessing 3D videos uses multiple scales, for example, the three major perceptual attributes: image/video quality, depth quality and visual comfort. This matches Kaptein's quality model described in Fig. 2.5(a). The scale in relation to the video quality in second method is closer to 2D videos quality which are generally familiar to the majority of subjects. The 3D features e.g. depth quality and visual comfort can be evaluated by subjects in separated scales or by means of questionnaires. The multi-scale quality model, on the one hand, can clearly indicate the features and the added values of 3D video, while on the other hand, it is still not clear how each individual scale is weighted, i.e. how important these 3D features are, and how much they can contribute to the general video QoE.

Chapter 3

3D coding and transmission quality

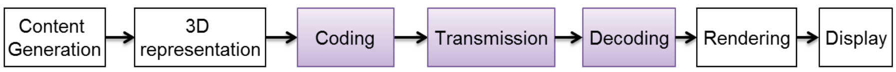


Figure 3.1: 3D video distribution chain–coding and transmission.

In chapter 2 the video quality model was briefly introduced as were the methods to measure the QoE, while in this chapter use is made of the knowledge gleaned from the previous chapter, and an investigation is conducted in relation to 3D video QoE and how QoE can be influenced by coding and transmission stages in the 3D video distribution chain (see Fig. 3.1). This chapter starts with an introduction of the S3D signal format, 3D coding schemes, coding artefacts and transmission artefacts on videos, after which the lists author's major contributions are listed in section 3.5.

3.1 Stereoscopic 3D signal formats

Stereoscopic 3D video signals have several formats, in general they can be divided into two categories:

- full resolution stereo (e.g. frame packing, frame sequential) ;
- frame compatible format (e.g. side-by-side, top-and-bottom);

The full resolution format 3.2(a) allows each individual view of the S3D to have full of their original resolution, the amount of data carried in each frame is approximately twice that for a 2D video frame. As for the frame compatible format, shown in Fig. 3.2(b) and (c), the left and right images are grouped into a single 2D video

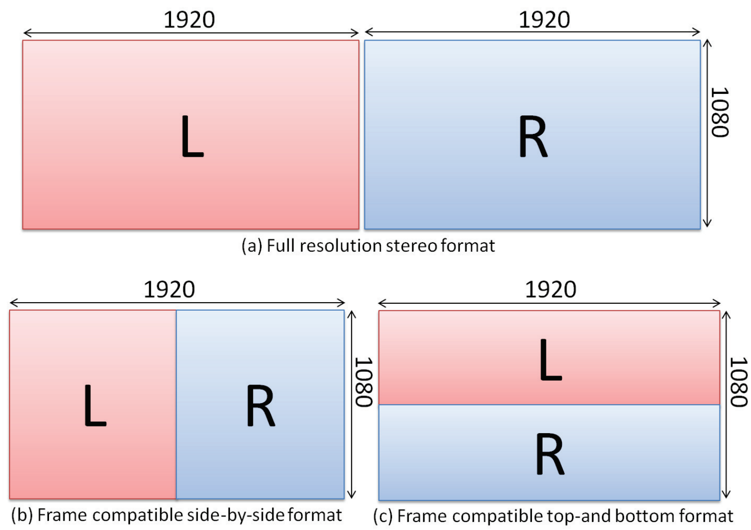


Figure 3.2: Stereoscopic video formats. (a)Full resolution S3D, (b)Frame compatible 3D side-by-side format, each view only has half of original horizontal resolution (c)Frame compatible top-and-bottom each view only has half of original vertical resolution

frame, and hence the spatial resolution is halved either in horizontal or vertical directions, but the amount of video data requiring to be processed and transmitted has remained the same as for the 2D videos. The frame compatible format offers the possibility of using an existing 2D video transmission and encoding/decoding system for delivering S3D videos without the necessity for any hardware upgrading.

3.2 Coding Schemes for Stereoscopic 3D videos

At the present time, in relation to 2D video services, the H.264/AVC coding scheme [Recommendation, 2007] is often employed . For 3D video broadcasting services, frame compatible format is used in the majority of cases, to which H.264 is applied as a 2D encoding algorithm. For full resolution stereoscopic 3D, two coding schemes are currently available namely the H.264 simulcast coding and H.264 multi-view coding (MVC). In the H.264 simulcast coding scheme it encodes the left and right views of stereoscopic videos independently, hence it allows the broadcaster to use the legacy H.264 coding equipment. This scheme generates approximately double the amount of encoded data as compared to that for 2D videos since two view channels must be coded separately. However, as the images from the different views are highly correlated, a great deal of information between the two views is redundant. H.264 MVC is one of the compression standards that uses this redundancy in order to improve the coding efficiency by introducing inter-view prediction, where images are not only predicted from spatially or temporally neighbouring image regions but also from corresponding images in adjacent views.

Apart from the above mentioned video coding schemes, various data processing operations such as temporal and spatial resampling are frequently used to improve the efficiency. Scalable video coding (SVC) [Schwarz *et al.*, 2007; Saygili *et al.*, 2009] provides the possibility of spatial, temporal and quality scalability, which can make video content delivery adaptive to variant network conditions and applications. Another method is to exploit the performance of the HVS in terms of binocular fusion and disparity sensitivity in the context of asymmetric coding [Saygılı *et al.*, 2010]. Furthermore, a new coding scheme for both 2D and 3D videos, High-Efficiency Video Coding (HEVC, also called H.265) [Sullivan *et al.*, 2012], will be standardized in 2013. The HEVC can improve the compression performance significantly when comparing it to the existing standards.

3.3 Coding Artefacts

Lossy compression may introduce visual artefacts depending on the coding techniques and compression ratio. A number of coding artefacts on the 2D image quality such as blocking artefacts, blur, ringing and staircase effects, colour bleeding, can be found in [Winkler, 2005]. In relation to S3D videos, due to the binocular vision of HVS, the coding artefacts can have a different impact on the perceptual experience of 3D videos in comparison to 2D videos. The coding artefacts degrade the left and the right views differently and when the difference between the left and right views videos is large, the depth perception can be hindered [Schertz, 1992; Meesters *et al.*, 2004]. In some circumstances this large difference can even cause visual discomfort due to binocular rivalry.

3.4 Transmission Artefacts

Apart from the artefacts introduced by the source coding, the transmission network itself often introduces errors due to delay, packet loss etc. Error concealment algorithms are normally integrated in the system or decoders to repair the transmission errors. Therefore, the visual quality depends on both the transmission artefacts and the error concealment approach. The impacts of network errors on 2D video quality have been discussed in many studies [Mu *et al.*, 2009; Meenowa *et al.*, 2010; Pinson *et al.*, 2010], and a review of error concealment methods for 2D videos can be found in [Wang & Zhu, 1998]. In the 3D case, a transmission distortion in one view or in both views is perceived differently. A degradation in one view or a temporal misalignment between the left and the right view leads to binocular rivalry. This binocular rivalry strongly degrades the quality of experience as it exhibits visual discomfort which might lead to a headache or nausea [Ukai & Howarth, 2008].

3.5 Contributions

The author's contributions from three papers are included in this thesis:

- Paper I: "Subjective Quality Assessment of Error Concealment Strategies for 3DTV in the presence of asymmetric Transmission Errors"
- Paper II: "Subjective evaluation of HDTV stereoscopic videos in IPTV scenarios using absolute category rating"
- Paper III: "Perceived 3D TV Transmission Quality Assessment: Multi-Laboratory Results Using Absolute Category Rating on Quality of Experience Scale"

The first paper addresses impacts of network transmission errors on the perceived S3D video quality of experience; the second paper focuses on coding and transmission efficiency of S3D videos; the third paper presents a joint extensive analysis of the experiment results from the first and second paper.

3.5.1 Paper I

The transmission of 3D sequences over packet based networks may result in degradations of the video quality due to packet loss. The main goal of paper I is the quality of experience of a stereoscopic 3D video in a simulated error-prone transmission channel. Three packet loss conditions and four error concealment strategies for 3D videos were compared and evaluated. A subjective experiment was performed using the ACR-HR method to evaluate the perceived video quality of experience and visual comfort. The selection of ACR method enabled a broad range of video processing conditions to be tested.

Novelty

There are three major novelties in this contribution:

- Four error concealment strategies were proposed and tested for S3D video transmission.
- The impacts of network errors on 3D video experience were investigated from both quality of experience aspects and visual discomfort aspects.
- This paper suggests a suitable bandwidth for transmit full HD resolution S3D videos.

Results

This paper proposes four error concealment methods for S3D videos.

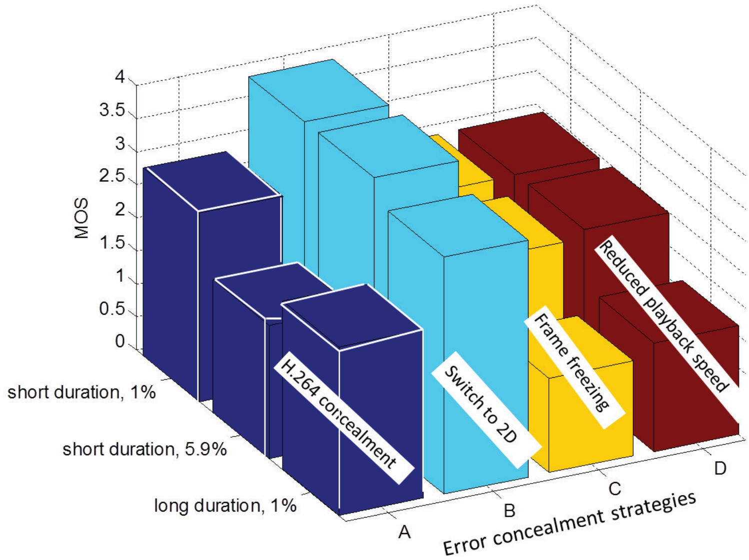


Figure 3.3: Subjective evaluation of four error concealment strategies for S3D videos in three packet loss scenarios

A 2D error concealment strategy: an implementation from H.264 decoder itself consists of directly playing back the decoded video.

B Switching to a 2D presentation: since one of the views in 3D video is distorted, the other view is still undistorted and this undistorted view is then displayed to both eyes when an error occurs in one view, thus leading to a 2D impression.

C Frame freezing: when transmission errors exist, the last frames that were correctly received for both views are displayed, and then the scene suddenly jumps to the next received error-free 3D frames and continues playing.

D Reduced playback speed: assuming that a buffer of video frames exists, when a transmission error detected, subjects would see that the playback slows down, jumps and then continues at normal speed when error-free.

The results show (see Fig. 3.3) that the subjects generally preferred the ″B″ method (Switching to a 2D presentation), and this is probably because this method did not interrupt the smooth playback of the videos, and the concealed video in both stereoscopic views were same, therefore, there was no binocular rivalry between the left and right views. The H.264 concealment algorithm ″A″ has a drawback in the S3D case, since only one view was distorted and even after concealment of the distorted area in the video, it appears differently to that of the corresponding area in the other view; this mismatch might cause visual discomfort and reduce the quality rating. The ″A″ results also indicate that the larger amount of packet loss had a greater effect

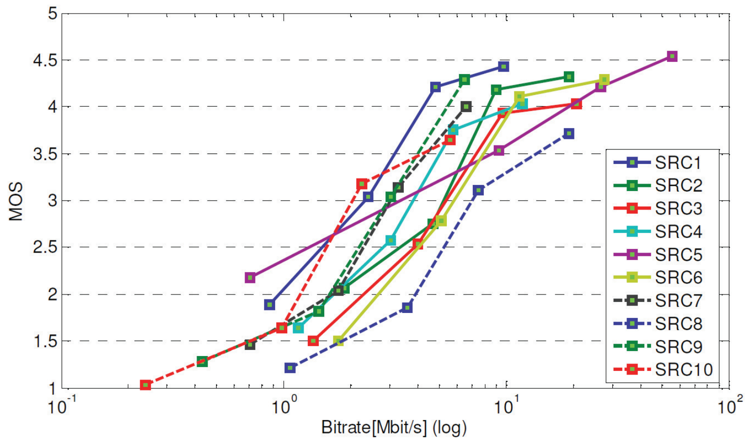


Figure 3.4: Full resolution 3D video transmission bandwidth for 10 video source contents.

on the visual discomfort due to the increased number of mismatches between two views. Remaining in an undistorted 3D presentation mode but, pausing or slowing the playback, as was tested in cases ”C” and ”D” usually perform more poorly than the ”A” method with the exception of the scenario involving the large amount of burst errors.

The study also shows that, in general, the packet loss errors spreading over a longer period of sequences proved to be more annoying than burst errors happening over a short period. For visual comfort, it has been determined that the QoE value and the visual discomfort depends strongly on the properties of the source reference sequence, e.g. the 3D scene content, the camera capturing, and the native resolution,etc . The coding artefacts and transmission errors influence not only the QoE value but also the visual comfort. A good MOS value is directly linked to the absence of visual disturbance.

The subjective experiment results also indicate that, for a given QoE level, the required transmission bandwidth varies significantly for different source content (SRC) as shown in Fig. 3.4. A transmission bandwidth of 10Mbps is necessary for full HD resolution S3D videos in order to secure a user quality mark of ”fair” or higher.

3.5.2 Paper II

Paper II addresses the coding and transmission efficiency of full resolution S3D videos. In order to study the effect on the quality of experience, two codecss were selected, namely the H.264 simulcast and MVC as they are, presently, the most commonly used. The videos were encoded at four different compression rate. In addition the effect of pre-process techniques such as temporal down-sampling and spatial down-sampling prior to the coding process are also studied.

Novelty

Two major novelties are included in this paper:

- Evaluation results of coding and transmission efficiency of full HD resolution S3D videos.
- Development of subjective measurement method for 3D videos.

Results

The results show that, at a common QoE level, a pre-processing technique using a resolution reduction of four (spatial down-sampling) on a full HD resolution S3D content may result in higher coding and transmission efficiency when H.264 video coding is used (see table 4 in Paper II). The spatial down-sampling technique not only saves the transmission bandwidth but also reduces the amount of processing work for encoding and decoding. The reduction of the frame rate did not save a significant amount of bitrate but it did reduce the video quality of experience to a large extent.

The subjective measurement method for 3D videos was experimentally developed in this paper. The lab environment was according to the ITU-BT.500-13 [Recommendation, 2012b]. The test was based on the standard ACR-HR method with an additional indicator of visual comfort as shown in Fig. 3.5 shows. For every se-

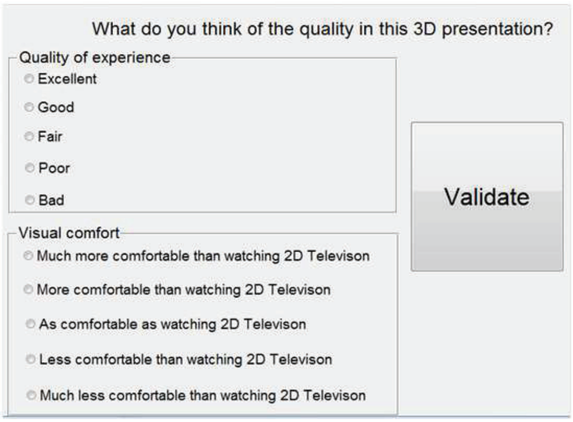


Figure 3.5: Subjective experiment voting interface, the Quality of Experience scale represents the general 3D QoE which considers both 2D and 3D atrributes, the additional visual comfort asks subjects to compare their comfort experience with 2D TV watching condition as a reference.

quence, the subjects were asked to vote for the overall quality of experience of 3D videos in the upper part of the voting interface with standard ACR quality scales.

The rating of overall quality experience includes the consideration of both 2D and 3D perceptual attributes, as described in model (Fig. 2.5(b)) in section 2.3.2. In the lower part of the voting window subjects were asked to use a scale of visual comfort in order to evaluate the visual comfort associated with the visualization of the sequences compared to viewing on a conventional 2DTV. Questionnaires were given to subjects to ask about their psycho-physical conditions both before and after the test.

3.5.3 Paper III

Paper III summarized the main findings from paper I II, then added an extensive analysis. The perceived video quality of experience was evaluated by a joint analysis of the data collected from three individual experiments conducted at two laboratories. The first experiment (Exp.1) was conducted at one lab in order to assess the network impact on quality. The second and third experiments (Exp.2 and Exp.3) were based on the same video sets and were to assess the coding and transmission efficiency, but they were conducted at two different labs. The assessment methods in this study provided the possibility to analyze combined cross-lab and cross-experiment data.

Novelty

There are three novelties included in this paper.

- Assessment method for cross-lab and cross-experiment analysis.
- Comparison of 3D QoE to 2D QoE
- Bitrate redundancy indicator for S3D videos.

Results

Cross-lab analysis was made based on common video sequence sets used in two different labs (Exp2 and Exp3). Fig. 3.6 shows the scatter-plots for the mean opinion scores (MOS) results from the two laboratories. The regression curve indicates that there is a slight offset and gradient. The detailed analysis with regard to lab difference can be found in paper III section IV. In order to remove the majority of the difference and to analyze data from two labs as a whole set, a linear transformation of the data from Lab1 (Acreo lab in Sweden) to Lab2 (IRCCyN lab in France) was conducted. In a similar manner to that for the cross-lab transformation, cross-experiment analysis was made based on common video sets between Exp.1 and the aligned results of Exp.2 and 3. Another linear transformation was performed by mapping Exp.1 results to the combined results of Exp.2 and Exp.3, and finally the combined cross-lab and cross-experiment data were analyzed as one data set.

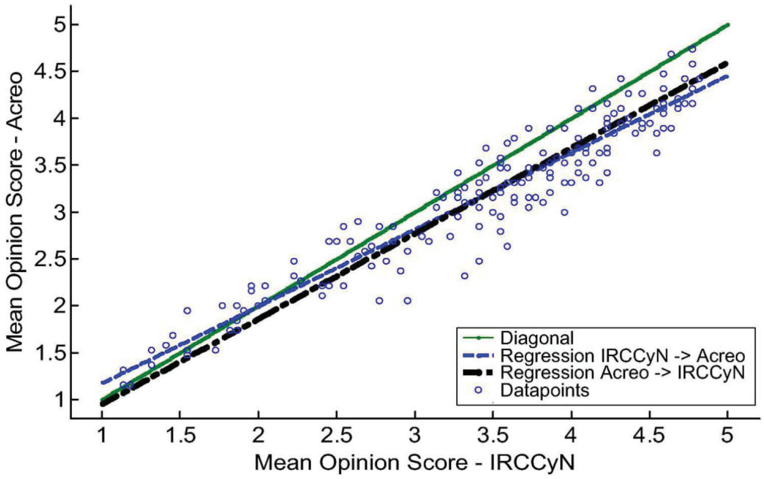


Figure 3.6: Cross-lab comparison: scatter plot of data from the two labs with linear regression. X-axis is the subjects' mean opinion score at lab2, and Y-axis is the MOS for lab1, each scatter point presents an individual subject's vote on a specific processed video.

Fig 3.7 shows a comparison between the undistorted 2D and 3D presentation of a video. It can be seen that there is a marginal preference for the 2D over the 3D presentation although the difference is not statistically significant for the majority of cases. This preference varies for different video source contents. Two of eleven SRCs (SRC4 and 10) had a statistical preference for their 2D presentations to their 3D presentations.

One of the goals of this study is to work towards establishing a reliable subjective test method for 3D videos. Questions have arisen regarding the correct interpretation of the scale ”Quality of Experience”. There could be many reasons regarding why the subjects did not feel that the 3D outperformed the 2D.

Firstly, while the ACR methodology appears to provide stable results across labs and experiments, it misses reference when suddenly viewing of 2D content in the context of 3D. Therefore the single stimulus method misses the accuracy of comparing 2D and 3D presentations. Secondly, in our experiments one scale model was used to evaluate the ”overall 3D quality of experience”. Although the ”visual comfort” was listed as an additional scale, the visual comfort attribute was not excluded when subjects were voting on the ”overall 3D quality of experience” scale. The overall QoE result is not clear that whether or not the added value of depth was considered in the final ”3D QoE” scores. Kaptein's model, described in section 2.2, could provide an alternative in order to make 3D added values visible. Thirdly, the mental reference of the observers, especially for naïve observers, is biased towards 2D viewing more than 3D since they are more used to 2D videos. Last but not least, while current S3D technology does offer the added values e.g. depth, it is also associated with some problems e.g. visual discomfort. Therefore, the overall quality of expe-

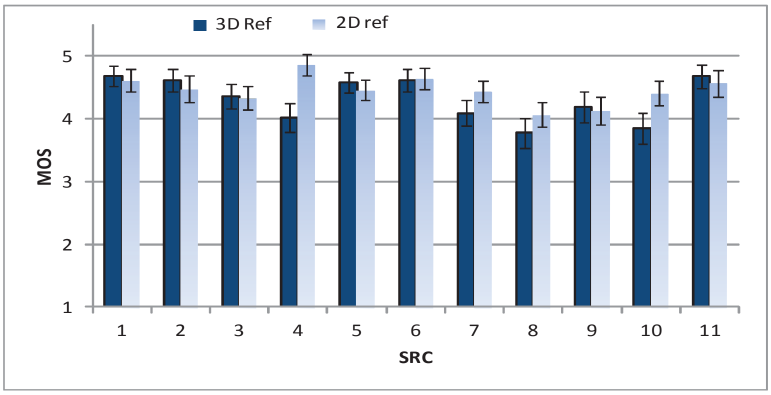


Figure 3.7: A comparison between 2D and 3D QoE, X-axis is a list of source content(SRC), Y-axis is the subjects' mean opinion score. for each SRC the left column presents 3D reference video and the right column presents 2D reference video.

rience that can be provided by the current 3D technology is not sufficiently good to cause the subjects to have a significant preference in relation to 3D.

A ”bitrate redundancy indicator” was introduced to investigate the interest of protecting the bitstream against transmission errors. This indicator compares three video transmission scenarios.

1. Less compressed video (bitstream has higher bitrate) directly transmits in an error-prone transmission environment, and then conceals errors at receiver side.
2. More compressed video (bitstream has lower bitrate) directly transmits in an error-free network.
3. More compressed video with error protection and correction method(additional bitrate overhead, e.g. FEC(forward error correction)) transmits together in an error-prone transmission environment.

Based on the same perceived quality level, the ”bitrate redundancy indicator” indicates how much bandwidth can be saved from the first scenario to the second transmission scenario. The gained bandwidth may be used for error protection and correction methods, e.g. ARQ (Automatic Repeat-reQuest) and FEC(forward error correction), which is the third transmission scenario. The study shows the ”bitrate redundancy indicator” ranges from 2 to 18 on average (the detailed results can be found in Paper III tableVI). It means that the second scenario is able to save at least half of the transmission bandwidth for the first scenario when both achieve the same QoE. Although the error-free transmission channel is an ideal case, if any error protection method (the overhead bitrate for error protection) uses less bandwidth than the saved bandwidth, then the third scenario is more bitrate efficient.

Chapter 4

Stereoscopic 3D Display Quality

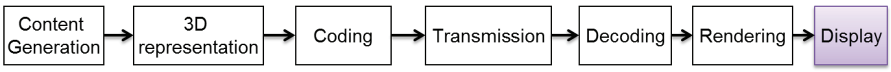


Figure 4.1: 3D video distribution chain–Display system.

3D display system is the final stage of the 3D video distribution chain 4.1, and it is the direct interface between the systems and a human. Therefore, the quality of display system has a strong influence on the perceived 3D quality of experience. There are many 3D display technologies at the present time, e.g. eye-displays based on glasses [Kim *et al.*, 2012; Lee *et al.*, 2009], auto-stereoscopic (glasses-free) displays [Dodgson, 2005], and volumetric 3D displays [Favalora, 2005], etc. The majority of the 3D display technologies are not compatible with each other as they are dependent on input 3D representations, i.e. for certain display, only the content can be shown, which has been specially prepared for that display technology. This incompatibility issue seriously limits the spread of the 3D displays especially when the amount of available 3D content is already limited at the present time.

There are two types of 3D display that are common at the present time, and they have already been commercially available on the consumer market for a number of years.

- Temporal multiplexed 3D display using active shutter lasses (SG type).
- Spatial multiplexed 3D display with Film-type Patterned Retarder technology, using passive polarized glasses (FPR type).

These two types of display require glasses and they only work with the stereoscopic 3D video format, which is widely used in current 3DTV broadcasting services and in movie theaters. This chapter addresses the comparison of these two type of stereoscopic 3D display systems and investigates the 3D display quality by objectively measuring important visual ergonomic parameters such as crosstalk, flicker and resolution. This chapter starts with an introduction to basic knowledge of current generation stereoscopic 3D displays in section 4.1, then major visual ergonomic parameters of S3D display are discussed in section 4.2. Section 4.3 lists the author's research contributions in the relevant area.

4.1 Stereoscopic 3D display system

The whole 3D display system consists of display equipment, synchronization device, if required and a viewing assistant device e.g. eye-glasses, different 3D display technology may use different types of glasses. Therefore, when evaluating the display system quality, the performance of a synchronization device and the properties of eye-glasses should also be considered.

4.1.1 Temporal multiplexed 3D display system using active shutter glasses(SG)

The Human Visual System (HVS) is capable of remembering a vivid image that was perceived by the eyes a short while ago and this phenomenon is called persistence of vision. When discrete images are presented to the human eyes above the flicker fusion frequency [Brundrett, 1974], the human perceives the discrete pictures as motion pictures. The active type 3D display, also called temporally or sequentially multiplexed display, as shown in Fig 4.2(a), is making use of this principle and displaying the left and right images of the S3D video sequentially to each eye at a fast refresh rate. The active shutter glass is a piece of liquid crystal display, synchronized with the display, when the display is showing left view images, only the left eye-glass is open for receiving the image from the display while the right eye-glass is closed, and vice versa.

Common 2D Liquid Crystal Displays (LCD) update at a refresh rate of 60 Hz [1], in order to temporally multiplex two views, a 3D active display must run at a refresh rate of 120 Hz, so that each eye will perceive 60 Hz. Some newer active 3D displays temporally insert black frames in between the left and right views, therefore a refresh rate for one stereo-pair frame becomes 240 Hz [Kim *et al.*, 2012].

The main advantage of the SG type display system is that it can offer both left and right view images at full original resolution, however, the temporal multiplexing might also cause a flicker perception when the left and right images are alternating.

[1]For TV systems 60 Hz for the countries using NTSC TV system and 50 Hz for PAL & SECAM; for Computer monitors, 60 Hz is common all over the world

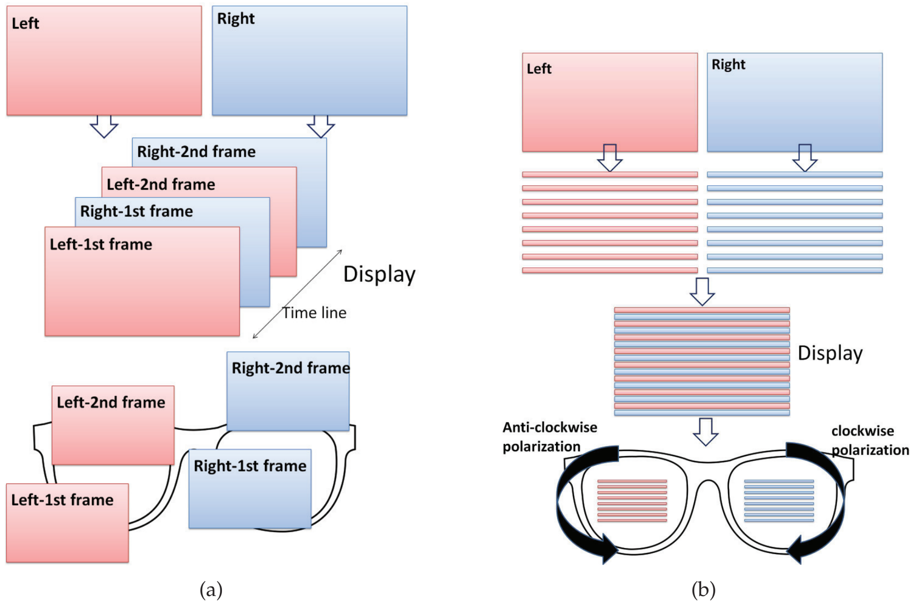


Figure 4.2: Two types of current generation stereoscopic 3D display system. (a)Temporal multiplexed display using active shutter glasses: the left and right channel video frames are sequentially shown to people, with the help of shutter glasses the left and right eye alternately receives video frames from corresponding channel (b) Spatial multiplexed display using passive polarized glasses: the left and right channel video frames are interleaved and shown to people at same time on the screen, with the help of polarized glasses each eye received video from corresponding channel

4.1.2 Spatial multiplexed 3D display system with film-type patterned retarder technology, using passive polarized glasses

The left and right view of the S3D video is spatially multiplexed in the passive type display as shown in Fig. 4.2(b). The pattern retarder is attached and aligned on the LCD panel, which makes each horizontal row with the opposite circular polarization on the display and hence presents the stereo-pairs at two different polarization states. The passive polarized eye-glass, one with a clockwise polarization and the other with an anticlockwise polarization, will filter out the corresponding images for each eye. Since both left and right view images are showing on the LCD panel at the same time, and every second line represents one of the views, the vertical resolution of each view is halved (see detail discussion in section 4.2.2). However there is no flicker problem in FPR type display since it does not use temporal multiplexing. In addition, the passive polarized glasses have the advantage of being light weight, low cost and have a high transmittance when comparing them to active shutter glasses.

4.2 Visual Ergonomic parameters

4.2.1 Crosstalk

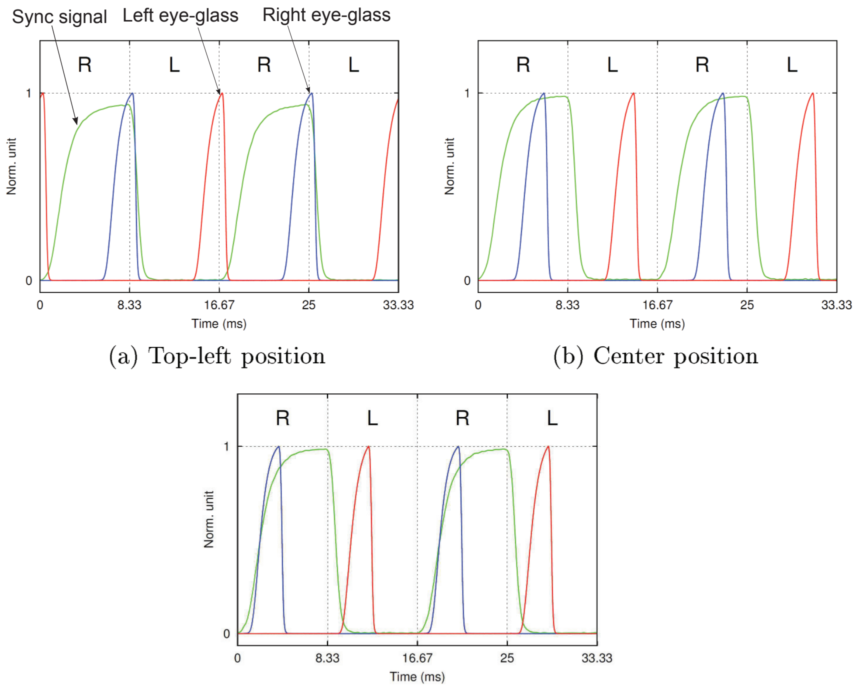
The term crosstalk has been used in many fields, it means the physical ”leakage” of information from one channel to an other. In the 3D video and imaging field, crosstalk refers to the incomplete insulation of the left and right image channels [Woods, 2011], and the ”leakage” from left to right or vice-versa can result in a perceived double image, contours or shadows, a phenomenon called ”image ghosting”. Crosstalk is one of the main causes of visual/perceptual impairments such as ghosting, which contributes to the decrease in perceived S3D video quality and visual comfort [Meesters *et al.*, 2004]. Crosstalk is affected by many factors such as contrast, disparity of the 3D video source, glasses, display technology, viewing distance and angles, etc.

There are many 3D crosstalk mathematical definitions [Woods, 2011] , one of the commonly used crosstalk formula defined by [Liou *et al.*, 2009] is shown in equation 4.1

$$CL = \frac{BW - BB}{WB - BB} \tag{4.1}$$

, where CL represents the crosstalk value measured through the left view. BW is the luminance measured through the left channel when the left view input is black (0) and the right view is white (255); and WB is when the left view input is white and the right view input is black; the BB is the luminance of the left view when both left and right view inputs are black.

The equation 4.1 only measures the crosstalk condition when full-white in one view and full-black in the other view, which is supposed to give the maximum crosstalk values. However, this is no longer true for displays that exhibit non-linear and non-additive processes, for example time-sequential multiplexed 3D LCD displays [Woods, 2012]. Therefore grey-to-grey crosstalk was proposed as a metrics of



(a) Top-left position

(b) Center position

(c) Bottom-right position

Figure 4.3: Synchronization between display and shutter glasses measured at different screen positions. i=0 and j=255. (a) in top-left position; (b) in bottom-right position; (c) in bottom right position. Luminance of the display has been normalized with respect to the white(255,255) luminance. Transmittance functions of the shutter.glasses have been normalized between 0 and 1.

crosstalk values for all grey level transition combinations.

$$CL(i,j) = \frac{L_{i,j} - L_{i,i}}{L_{j,i} - L_{i,i}} \tag{4.2}$$

, where i, j represent the grey level (ranges from 0-255) of the left channel and right channel respectively. $CL(i,j)$ is the crosstalk measured through the left view when the left view input is the grey level i and the right view input is grey level j. $L_{i,j}$ is the luminance measured through left channel when left view input is the grey level i and the right view input grey level is j.

The LCD pixels take different times for transiting from one grey level to the other (pixel response rate). For the SG type display system, when the left and right views alternate the display might fail to reach a correct luminance within a short period. Therefore the grey-to-grey response varies, which leads to a varying grey-to-grey crosstalk. The variation of crosstalk ratio is hardly predictable [Tourancheau *et al.*, 2012].

The black frame insertion technique, as mentioned in section 4.1.1, can assist the SG type 3D display to reduce crosstalk [Kim *et al.*, 2012], however, it will reduce the brightness. Therefore, there is a trade off between the reduced crosstalk and the amount of luminance of stereoscopic videos passing through the eye-glasses.

For the SG type display, the synchronization between the shutter-glasses and the display also contributes to crosstalk [Tourancheau *et al.*, 2012]. The shutter open period cannot be equally synchronized with the whole display because of the temporal delay between the first and last line update. Fig. 4.3 illustrates this difference for three different screen locations.

4.2.2 Resolution and detail

Video electronics standards association (VESA) [FPDM *et al.*, 2005] defines: ”the resolution as a measure of the ability to discriminate picture detail”. It often refers to a number of pixel lines that could just be distinguished by the human eye. People usually see the display resolution in ”width x height” pixels, in the products specification, for example, ”1920 x 1080”, which means the number of pixels in each dimension and the human eye should be able to resolve 1920 pixels in the horizontal direction and 1080 pixels in the vertical direction.

In relation to the perceived resolution, it is inevitably related to the properties of the human vision system and viewing conditions, e.g. the spatial contrast sensitivity function of the eye [Barten, 1999], the viewing distance, environment luminance etc. The perceived spatial resolution is measured by visual acuity, also called resolving power. A vision of 20/20 [2], also known as ”1.0” in decimal notation, is considered to be the nominal performance for human distance vision. At such distance (20 feet)

[2] the ”20” means the distance between the subject and the visual acuity chart in feet. ”20/20” means that the subject can read the chart (from 20 feet away) as well as a normal person could read the same chart from 20 feet away.

a human eye with nominal performance is able to resolve lines with a gap of one arc minute. i.e. 1.75 mm apart.

The measurement of display resolution, according to the Information Display Measurement Standard (IDMS) [ICDM, 2012], is based on threshold contrast modulation associated grille patterns. The threshold value is the minimum modulation required for the detection of the grille pattern (distinguish black and white lines). The contrast modulation (Michelson contrast) is shown in equation 4.3 .

$$C_m = \frac{L_{255} - L_0}{L_{255} + L_0} \tag{4.3}$$

, where C_m is the contrast modulation, L_{255} is the luminance of white and $L0$ is the luminance of black. The requirement of the threshold contrast modulation may differ depending on the display applications. Text usage requires a high visibility for example, the character edges must be clear. IDMS [ICDM, 2012] defines in relation to texts, the maximum number of alternating black and white lines in the grille pattern that can be displayed with a threshold contrast modulation of 50%. For image applications, normally, sharp luminance changes in an image occur infrequently, therefore it has lower requirement as compared to that for text applications, IDMS suggests a minimum threshold of 25% for image applications. The resolution is defined as equation 4.4:

$$n_r = n + \frac{C_t - C_m(n)}{C_m(n+1) + C_m(n)} \quad R = \frac{N_{adr}}{n_r} \tag{4.4}$$

, where n is the pixel width, e.g. 1 pixel width grille patter; $C_m(n)$ is contrast modulation of n pixel width grille; $C_m(n+1)$ is the contrast modulation of n+1 pixel width grille; C_t is threshold contrast modulation. n_r is calculated grille line width in pixels for which the value of C_m is estimated by linear interpolation to be equal to the contrast modulation threshold C_t. N_{adr} refers to the number of addressable pixel lines that can be separately controlled e.g. 1080 vertical pixel lines. When $C_m(1)$ is larger than C_t, then nr is equal to 1 and the resolution,R, is equal to the number of addressable lines.

Active shutter glasses based on a 3D display are temporally multiplexed, as shown in Fig 4.2(a), therefore each eye is receiving the full resolution of the original left and right view images. While the passive polarized glasses based on a 3D display are spatially multiplexed, as shown in Fig. 4.2(b). Each eye is receiving only half of the display pixel lines when playing stereoscopic videos, hence the spatial resolution perceived by each of the eyes for the FPR display is lower than the SG display.

For stereoscopic 3D video or images, the perception of 3D depth is according to the binocular fusion of the perceived image in the left and right views. For the FPR type display, the left and right view images have a difference of one pixel line shifted and a simply summation of the left and right views will not provide a fully compensated pixel line resolution. The human binocular vision will, most probably, make a small vertical alignment of the 1 pixel shifted rows in the two views, and combine them as one line seen in the binocular vision [Kim & Banks, 2012].

Barten [Barten, 1999] and Campbell [Campbell & Green, 1965] stated that in binocular vision, the spatial contrast sensitivity increases by a factor of the square root of 2 as compared to that for monocular viewing, however, this is valid only if the information for both eyes is completely combined and the contrast sensitivity is not limited by external noise, which is the noise already presented in the images.

Hakkinen [Häkkinen *et al.*, 2008] showed that 3D videos sometimes slso have more detail but he also reported that some people feel that 3D videos are more blurred than 2D. Heynderickx [Heynderickx & Kaptein, 2009] pointed out that the blur observed in the Hakkinen study might be due to other binocular artefacts e.g. crosstalk. The crosstalk can lead to double image contour which decreases the perceived sharpness. Heynderickx's subjective experiment results also found that the perceived amount of detail is higher in 3D than in 2D images with the same spatial resolution, and the concept of detailedness may not be equal to sharpness in the 3D case.

4.2.3 Flicker

Flicker is defined by Watson [Watson & Ahumada, 2012]: "a perceptual attribute of displays that consists of an apparent fluctuation in brightness of a surface subject to rapid periodic modulation of luminance". The study of flicker can be traced back to the invention of the movie and television. In comparison to outdated CRT displays, the present LCD displays have significant improvements with regards to flicker visibility, however when stereoscopic 3D with temporal multiplexing technology is introduced, flicker artefacts become one of the most important artefacts that affect the perceived QoE.

The temporal contrast sensitivity function (TCSF), which describes temporal frequency of luminance modulation, is normally referred to as the human visual sensitivity to the luminance changes over time. Fig. 4.4 shows the TCSF described in the temporal sensitivity chapter of the book by [Boff *et al.*, 1986]. As can be seen, the TCSF falls rapidly at high temporal frequencies. Thus, flicker is usually avoided if the display is refreshed at a rate above the critical flicker frequency. The critical flicker frequency is the threshold frequency of the flicker that can be observed. Flicker is more visible for larger fields and for a higher average luminance [Boff *et al.*, 1986].

Display flicker measurement is standardized in HFES (Human Factors and Engineering Society) 100 display standard [Factors & Society, 2007; Farrell *et al.*, 1987] and VESA [FPDM *et al.*, 2005]. The HFES compare human TCSF with the measured fundamental frequency of the display luminance modulation, and the VESA multiplies the TCSF on the measured amplitude of the first several harmonics of the temporal modulation.

Andrew Watson [Watson & Ahumada, 2012] proposed a measurement method which uses JND (just noticeable difference) as a unit to represent the flicker visibility. It measures the luminance as a function of time. The fundamental frequency components can be calculated from the measurement data, and then the JND is calculated

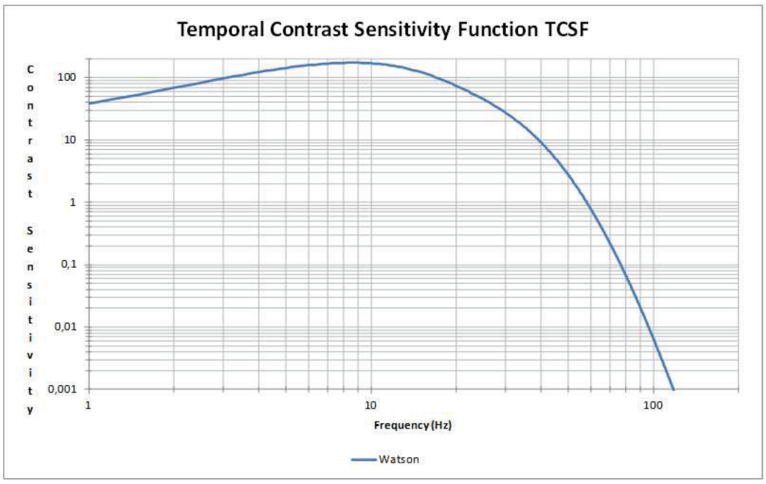


Figure 4.4: Human temporal contrast sensitivity function, the sensitivity function falls sharply when the temporal frequency above 10Hz, and when frequency reach about 60Hz, the sensitivity falls below 1 which is the human contrast threshold.

by multiplex contrast and the TSCF. This method of measuring flicker visibility is included in the IDMS [ICDM, 2012].

4.3 Contributions

The author's contribution in the field of display quality and characterization has been summarized in paper 4 IV, "Characterizations of 3D TV: Active vs Passive". Two types of stereoscopic 3D displays: active SG type and passive FPR type were compared with respect to some important visual ergonomic parameters such as angular dependent crosstalk, flicker and resolution.

4.3.1 Novelty

There are two major novelties included in this paper:

- Measured major visual ergonomic parameters for 3D display e.g. angular dependent crosstalk, resolution, flicker, etc.

- Methods development for 3D displays measuring.

4.3.2 Results

Angular dependent crosstalk was measured according to equation 4.1 for both SG and FPR 3D displays due to the measurement time constraints. The measurement set up is shown in Fig. 4.5.

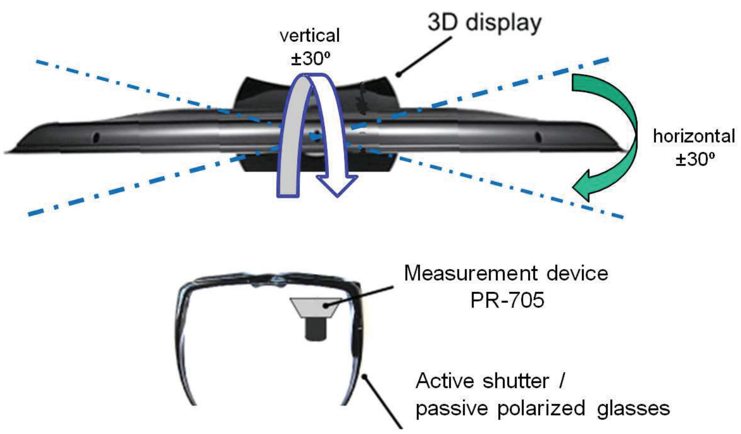
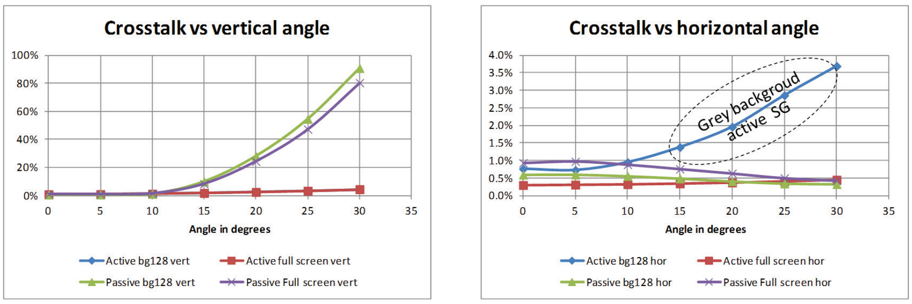


Figure 4.5: Display measurement setup.



(a) (b)

Figure 4.6: Angular dependent crosstalk results, X-axis is measurement angles, Y-axis is crosstalk values. (a)measurement angle varies in vertical direction, Y-axis ranges from 0% to 100%,(b)measurement angle varies in horizontal direction, Y-axis ranges from 0% to 4%.

The test sample was mounted in a specially designed holder. During the measurement, the holder was rotated and tilted in both the horizontal and vertical directions from -30° to +30°. The crosstalk, luminance and colour were measured by means of spectro-radiometer. The result shows that the FPR type was more sensitive to the changes of vertical angle especially when there was an angle larger than 10°, where the crosstalk surged exponentially, see Fig. 4.6(a). Therefore a correct

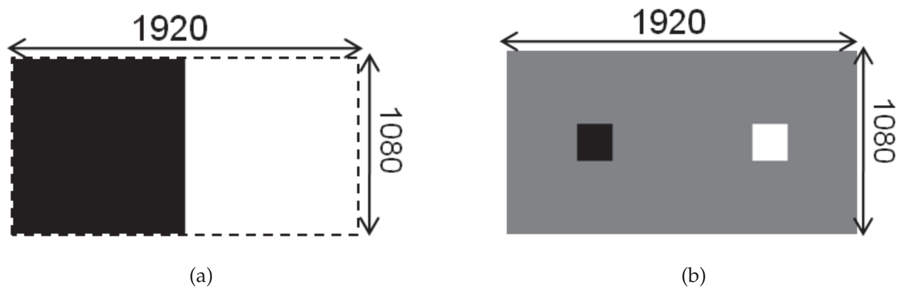


Figure 4.7: Crosstalk test images in side-by-side format. (a)full black/white image,(b)Black/white pattern only in center of the image, grey areas as a background fills the rest of the test image

placement is important for the FPR type 3D display. For the horizontally angular dependent crosstalk both types of TV had low crosstalk value, see Fig. 4.6(b), e.g. crosstalk less than 1% within +/- 10°.

It was also found that the measured crosstalk, especially for SG type display, can be affected by many factors for example test pattern and alignment of the eye-glasses. Fig. 4.6(b) shows the horizontal angular crosstalk to be below 1% across all the measured angles when the full black/white image was used in one of the measurements (measurement1 with 100% image loading in entire stereoscopic views as shown in Fig. 4.7(a). When another test pattern (Fig. 4.7(b)) was used in the other measurement (measurement2) the crosstalk increases with the larger horizontal angles. The crosstalk difference is mainly derived from the BW value (measured luminance of black channel, when the other channel is white) in equation 4.1. The measured BW value was higher in measurement1 than in measurement2, which is probably because certain TV functions (e.g.dynamic contrast adjustment, etc.) allow the display to be darker when showing one of the views fully black.

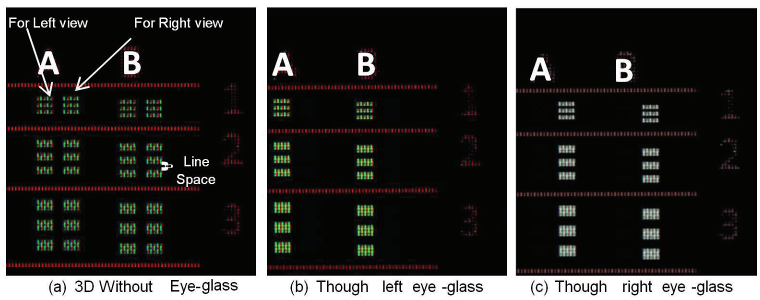
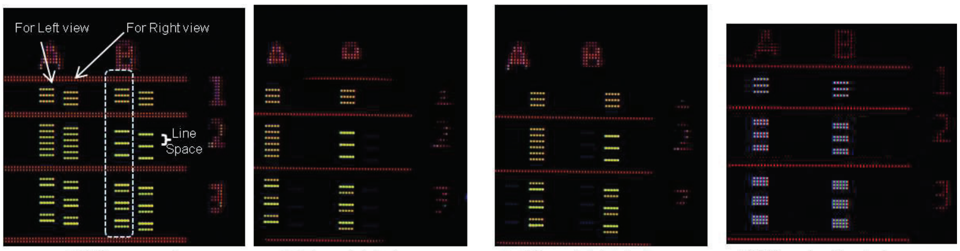


Figure 4.8: 3D resolution test pattern represented by SG type 3D display, (a) without eye-glasses, (b) through left glass, (c) through right glass. "A" and "B" group has one pixel line difference in vertical direction. There are two columns within each group "A" or "B", they are horizontally shifted with 8 pixels disparity , one for left view and one for right view.



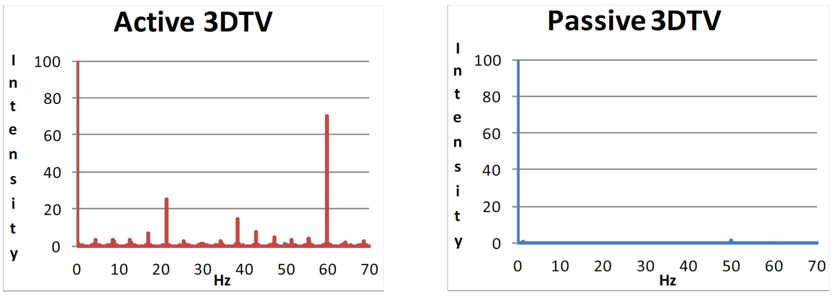
(a) Without eye-glasses (b) Through left glass (c) Through right glass (d) 2D mode without glasses

Figure 4.9: 3D resolution test pattern represented by FPR type 3D display, a) without eye-glasses, (b) through left glass, (c) through right glass, (d) displayed in 2D model.

The resolution of the displays was measured according to the ICDM standard [ICDM, 2012]. The active type 3D TV can display the full HD resolution for each of the stereoscopic views, while the FPR type display delivers half of the full HD resolution to each view. Designed bar test patterns were also used to examine the display resolution in the 3D mode. For the SG display, all bar patterns (Fig. 4.8) were correctly represented to each of the stereoscopic views. For the FPR type display (Fig. 4.9) only the one-pixel height bar pattern was correctly represented and for the two and more pixel height bars, it cannot reproduce the same pattern as the original due to the film-type patterned retarder technology. In 2D displaying mode, both types of TV represent the original bar patterns correctly.

The transmittance of the eye-glasses makes the active type darker than for the passive type. The low transmittance of the eye-glasses will directly affect the perceived brightness and contrast of the videos, therefore it will affect the video quality attributes, e.g. perceived 3D resolution and detailedness.

Critical flicker frequency was detected in the active type 3D TV, as shown in Fig. 4.10(a). The 60Hz spectrum (corresponding to an S3D individual view frame refresh rate at 60Hz) had a highest magnitude of about 70% of the fundamental frequency of the display luminance modulation (0Hz spectrum), the second highest spike occurred at 22Hz where the magnitude was about 25% of the magnitude at 0Hz. However, the spike at 22Hz will produce a strong visible flicker, while the 39Hz and 60Hz spikes will produce negligible visible flicker due to the human temporal contrast sensitivity i.e. TCSF. The perceivable flicker is largely related to the contrast of the video content, lighting conditions and surrounding environment. Hence, the active SG type 3D TV requires carefully adjustment in relation to the view conditions in order to minimize the perceived flicker. For the FPR type display, no critical flicker frequency was detected.



(a) (b)

Figure 4.10: Flicker measurement results, (a) SG type 3D display,(b) FPR type 3D display. X-axis is the frequency response in Hz, Y-axis is the intensity of the spectrum normalized from 0 to 100.

Chapter 5

Conclusions

This thesis investigated stereoscopic 3D video quality that can be influenced during the 3D video delivery chain, especially in the coding, transmission and display stages. The study provides the following major outcomes:

- The thesis suggested a bitrate efficient means of encoding and transmission stereoscopic 3D videos with an acceptable users' quality of experience.
- The thesis suggested an error concealment method for 3D video transmitted over error-prone network.
- The thesis measured and compared two current types of stereoscopic 3D display systems with major visual ergonomic parameters.
- The thesis described and compared two 3D subjective measurement models for assessing end-users' 3D video QoE, as well as the measurement methods for characterizing 3D display system.

The outcome of the thesis could benefit 3D video industries in relation to improving their technologies in order to delivery better 3D quality of experience to customers. The following section will discuss the achievements of this thesis in greater detail.

5.1 Outcome

5.1.1 Stereoscopic 3D coding and transmission quality

The first goal addressed in this thesis was to investigate the impact of coding artefacts on users' 3D quality of experience. The study showed that a coding compression rate of QP32 using a H.264 codec was the minimum threshold with regards to maintaining the watching experience of the users at a "good" level. A transmission bitrate of 10 Mbps was necessary in order to achieve a quality mark of "fair" or higher. A resolution reduction of four, prior to the coding process, not only assist the service provider in improving transmission efficiency but also saves some hardware processing which would be required for encoding and decoding. The study results are useful for 3D video service providers to determine an appropriate compression degree and bandwidth for S3D video transmissions.

The second goal addressed in this thesis was to investigate the impact of transmission errors on users' 3D quality of experience. The thesis showed that the influence of packet loss in 3D videos will not only affect the video quality but also make users feel a visual discomfort. The discomfort becomes more severe when the packet loss rate is higher. Four error concealment methods for S3D video were proposed and compared. The "switch to 2D presentation" method is the best method.

The third goal of the thesis was to investigate the subjective methods for measuring users' 3D QoE and to work towards establishing a reliable subjective test method for 3D videos. The thesis described two quality models for 3D video subjective measurement: one scale model (all perceptual attributes of the video were integrated as one scale rated by users) and multi-scales model (major attributes listed as separated rating scales presented to the viewer). In the subjective tests included in this thesis, the plan was to evaluate the general 3D visual QoE compared to viewing experience on a conventional 2DTV, therefore, the one scale model with an additional scale of visual comfort were attempted. The thesis also proposed a method of cross-lab and cross-experiment data analysis.

5.1.2 Stereoscopic 3D Display Quality

The fourth goal of the thesis addressed the quality characterization of the S3D display system. Three important visual ergonomic parameters of the display system such as crosstalk, flicker and resolution were measured. Two major types of 3D TVs from the current consumer market, displays based on active shutter glasses (SG) and display based on passive polarized glasses (FPR), were examined and compared. The results showed the 3D crosstalk increases when the viewing angle increases. The FPR type display is more sensitive to the angular changes in the vertical direction than in horizontal direction. The SG type TV can display a full HD resolution for each view of the stereoscopic videos, but for the passive FPR display only half of the vertical resolution can be displayed to each eye due to the film-type patterned retarder technology.

The measured transmittance of the polarized glasses was higher than that for the active shutter glasses, which makes the FPR type display appear to be brighter through the eye-glasses. The perceptible flicker frequency was measured and detected in the SG type display system but not in the FPR display system. The study results are useful for 3D display manufactures to improve the display technologies so that end-users can obtain a better 3D video QoE.

5.2 Limitations of the work

The research work included in this thesis has the following major limitations:

- The subjective experiments conducted in this thesis are mainly using the one scale QoE evaluation model. The individual depth features of 3D video are not clearly evaluated.
- The 3D video quality of experience studies on 3D coding and transmission were based on a temporal multiplexed 3D display system with active eye-glasses, therefore the results may not apply to other types of 3D display technologies.
- The subjective tests setup and video content selection were based on the available technologies and resources at that time, with newly improved 3D displays and 3D content capturing technologies, the results may differ.
- The measurement methods for the 3D display system used in this thesis are not mature, the human 3D perception aspects were missing in the development of the measurement methods. For example, the resolution measurement were based on measuring the 2D display resolution, however, it did not really measure the 3D resolution when binocular summation processes etc. were involved.

5.3 Future works

The continue work of this thesis could include:

- Development of reliable subjective methods for 3D videos: Both one scale quality evaluation model and multi-scales model have their pros and cons. A comparison study of these two quality models is necessary to determine the relationship between the overall QoE and individual quality perceptual attributes, and hence to obtain a more accurate measure in relation to 3D video QoE.
- Development of objective metrics: Objective metrics is a fast and automatic means of measuring or predicting 3D video QoE. For 2D videos significant efforts have been involved in relation to developing objective models; however, for 3D videos, the objective quality metrics have still not been widely studied. A continuation work of this thesis is to develop 3D objective metrics based on the subjective data obtained in this thesis.
- Development of measurement methods for 3D display system: although this thesis made an experimental study with regards to measuring visual ergonomic parameters of 3D display system, the method for measuring 3D displays are far from being robust. Based on the involvement of 3D binocular vision, many methods such as that for measuring 3D resolution, flicker etc must be further developed.

5.4 Ethical considerations

Ethical aspects have been considered in both the author's research work and with regards to the presentation of the results.

All subjective testing procedures have been pursued with the well-being of the test subjects being taken into consideration. For examples: a person can stop the test at anytime without offering any reason; the recruiting procedures were independent with the subjective experiment and any personal information being handled anonymously; the test were designed with multiple sessions with short breaks in between; all subjects' contributions were acknowledged.

The subjects were screened prior to the subjective test according to their visual acuity, color blindness and stereoscopic acuity. After the subjective experiment, all data were screened according to standard procedures [Recommendation, 2012b] and [VQEG *et al.*, 2009]

Bibliography

Alatan, A Aydın, Yemez, Yücel, Gudukbay, Uur, Zabulis, Xenophon, Muller, Karsten, Erdem, Çigdem Eroglu, Weigel, Christian, & Smolic, Aljoscha. 2007. Scene representation technologies for 3DTV–a survey. *Circuits and Systems for Video Technology, IEEE Transactions on*, **17**(11), 1587–1605.

Barten, Peter GJ. 1999. *Contrast sensitivity of the human eye and its effects on image quality*. SPIE-International Society for Optical Engineering.

Boff, Kenneth R, Kaufman, Lloyd, & Thomas, James P. 1986. *Handbook of Perception and Human Performance*. New York, USA: Wiley.

Bruls, WHA, Varekamp, C, Gunnewiek, R Klein, Barenbrug, B, & Bourge, A. 2007. Enabling introduction of stereoscopic (3d) video: Formats and compression standards. *Pages I–89 of: Image Processing, 2007. ICIP 2007. IEEE International Conference on*, vol. 1. IEEE.

Brundrett, GW. 1974. Human sensitivity to flicker. *Lighting Research and Technology*, **6**(3), 127–143.

Campbell, FW, & Green, DG. 1965. Monocular versus binocular visual acuity. *Nature*.

Dictionary, Oxford English. 1989. *Oxford: Oxford University Press*.

Docherty, A Woods T, & Koch, R. 1993. Image distortion in stereoscopic video systems. *Proc. SPIE-Int'l Soc. Optical Eng., Stereoscopic Displays and Applications IV*, **1915**, 36–48.

Dodgson, Neil A. 2005. Autostereoscopic 3D displays. *Computer*, **38**(8), 31–36.

Engeldrum, Peter G. 2000. *Psychometric scaling: a toolkit for imaging systems development*. Imcotek Press.

Factors, Human, & Society, Engineering. 2007. *ANSI/HFES 100-2007,Human Factors and Engineering of Computer Workstations*.

Farrell, JE, Benson, Brian L, & Haynie, Carl R. 1987. Predicting flicker thresholds for video display terminals. *Pages 449–453 of: Proceedings of the SID*, vol. 28.

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

FPDM, VESA, Association, Video Electronics Standards, *et al.* 2005. *Flat Panel Display Measurements Standard*.

Graham, Clarence Henry. 1965. *Vision and visual perception*. John Wiley & Sons.

Grau, Oliver. 2004. 3d sequence generation from multiple cameras. *Pages 391–394 of: Multimedia Signal Processing, 2004 IEEE 6th Workshop on*. IEEE.

Häkkinen, Jukka, Kawai, Takashi, Takatalo, Jari, Leisti, Tuomas, Radun, Jenni, Hirsaho, Anni, & Nyman, Göte. 2008. Measuring stereoscopic image quality experience with interpretation based quality methodology. *Page 68081B of: Proc. SPIE*, vol. 6808.

Heynderickx, Ingrid, & Kaptein, Ronald. 2009. Perception of detail in 3D images. *Page 72420W of: Proc. SPIE*, vol. 7242.

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

Hori, Hiroki, Shiomi, Tomoki, Kanda, Tetsuya, Hasegawa, Akira, Ishio, Hiromu, Matsuura, Yasuyuki, Omori, Masako, Takada, Hiroki, Hasegawa, Satoshi, & Miyao, Masaru. 2011. Comparison of accommodation and convergence by simultaneous measurements during 2D and 3D vision gaze. *Virtual and Mixed Reality-New Trends*, 306–314.

ICDM, International Committee for Display Metrology. 2012. *Information Display Measurement Standard*.

IJsselsteijn, Wijnand A, de Ridder, Huib, Freeman, Jonathan, & Avons, Steve E. 2000. Presence: concept, determinants, and measurement. *Pages 520–529 of: Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, vol. 3959.

ITU-T, Recommandation. 2008. P910. *Subjective video quality assessment methods for multimedia applications*.

Kaptein, Ronald G, Kuijsters, André, Lambooij, Marc TM, IJsselsteijn, Wijnand A, & Heynderickx, Ingrid. 2008. Performance evaluation of 3D-TV systems. *Pages 680819–680819 of: Electronic Imaging 2008*. International Society for Optics and Photonics.

Kim, Joohwan S, & Banks, Martin S. 2012. 65.1: Effective Spatial Resolution of Temporally and Spatially Interlaced Stereo 3D Televisions. *Pages 879–882 of: SID Symposium Digest of Technical Papers*, vol. 43. Wiley Online Library.

Kim, Sang Soo, You, Bong Hyun, Choi, Heejin, Berkeley, Brian H, Kim, Dong Gyu, & Kim, Nam Deog. 2012. 31.1: Invited Paper: World's First 240Hz TFT-LCD Technology for Full-HD LCD-TV and Its Application to 3D Display. *Pages 424–427 of: SID Symposium Digest of Technical Papers*, vol. 40. Wiley Online Library.

King, Laura A. 2009. *Experience Psychology*. 1st edn. New York, USA: McGraw-Hill Inc.

Le Callet, Patrick, Möller, Sebastian, Perkis, Andrew, & eds. 2012. *Qualinet White Paper on Definitions of Quality of Experience (QoE) and Related Concepts*.

Lee, Sung Jung, Kim, Min Jung, Lee, Kyo Hyeon, & Park, Kwang Hoon. 2009. Review of wire grid polarizer and retarder for stereoscopic display. *Proc. SPIEVInt. Soc. Opt. Eng*, **7237**(1).

Liou, Jian-Chiun, Lee, Kuen, Tseng, Fan-Gang, Huang, Jui-Feng, Yen, Wei-Ting, & Hsu, Wei-Liang. 2009. Shutter glasses stereo LCD with a dynamic backlight. *Pages 72370X–72370X of: IS&T/SPIE Electronic Imaging*. International Society for Optics and Photonics.

Meenowa, Joshan, Hands, David S, Young, Rhea, & Bayart, Damien. 2010. Subjective assessment of HDTV content: comparison of quality across HDTV formats. *Page 75270I of: Proc. SPIE*, vol. 7527.

Meesters, Lydia MJ, IJsselsteijn, Wijnand A, & Seuntiëns, Pieter JH. 2004. A survey of perceptual evaluations and requirements of three-dimensional TV. *Circuits and Systems for Video Technology, IEEE Transactions on*, **14**(3), 381–391.

Merkle, Philipp, Smolic, Aljoscha, Muller, Karsten, & Wiegand, Thomas. 2007. Multi-view video plus depth representation and coding. *Pages I–201 of: Image Processing, 2007. ICIP 2007. IEEE International Conference on*, vol. 1. IEEE.

Mu, Mu, Gostner, Roswitha, Mauthe, Andreas, Tyson, Gareth, & Garcia, Francisco. 2009. Visibility of individual packet loss on H. 264 encoded video stream: a user study on the impact of packet loss on perceived video quality. *Page 725302 of: Proceedings of SPIE*, vol. 7253.

Ozaktas, Haldun M, & Onural, Levent. 2007. *Three-dimensional television: capture, transmission, display*. Springer Publishing Company, Incorporated.

Pinson, Margaret H, Wolf, Stephen, & Cermak, Gregory. 2010. HDTV subjective quality of H. 264 vs. MPEG-2, with and without packet loss. *Broadcasting, IEEE Transactions on*, **56**(1), 86–91.

Punchihewa, Amal, & Bailey, Donald G. 2002. Artefacts in image and video systems; classification and mitigation. *Pages 197–202 of: Proceedings of image and vision computing New Zealand*.

Recommendation, ITU-RBT. 2000. 1438, Subjective Assessment of Stereoscope Television Pictures. *International Telecommunication Union, Geneva, Switzerland*.

Recommendation, ITURBT. 2012a. 2160-3, Features of three-dimensional television video systems for broadcasting. *International Telecommunication Union, Geneva, Switzerland*.

Recommendation, ITURBT. 2012b. 500-13, Methodology for the subjective assessment of the quality of television pictures. *International Telecommunication Union, Geneva, Switzerland*.

Recommendation, ITUT. 2007. H.264 Advanced Video Coding for Generic Audiovisual Services. *ITU-T Rec. H*, **264**, 14496–10.

Saygili, Görkem, Gurler, CG, & Tekalp, A Murat. 2009. 3D display dependent quality evaluation and rate allocation using scalable video coding. *Pages 717–720 of: Image Processing (ICIP), 2009 16th IEEE International Conference on*. IEEE.

Saygılı, Görkem, Gürler, C Göktuğ, & Tekalp, A Murat. 2010. Quality assessment of asymmetric stereo video coding. *In: IEEE Int. Conf. on Image Processing (ICIP)*.

Schertz, A. 1992. Source coding of stereoscopic television pictures. *Pages 462–464 of: Image Processing and its Applications, 1992., International Conference on*. IET.

Schreer, Oliver, Kauff, Peter, & Sikora, Thomas. 2005. *3D videocommunication*. Wiley Online Library.

Schwarz, Heiko, Marpe, Detlev, & Wiegand, Thomas. 2007. Overview of the scalable video coding extension of the H. 264/AVC standard. *Circuits and Systems for Video Technology, IEEE Transactions on*, **17**(9), 1103–1120.

Seuntiëns, P, Heynderickx, I, IJsselsteijn, Wijnand A, van den Avoort, Paul MJ, Berentsen, Jelle, Dalm, Iwan J, Lambooij, Marc T, & Oosting, Willem. 2005. Viewing experience and naturalness of 3D images. *Pages 43–49 of: Proceedings of the SPIE*, vol. 6016.

Seuntiens, Pieter, Meesters, Lydia, & IJsselsteijn, Wijnand. 2003. Perceptual evaluation of JPEG coded stereoscopic images. *Pages 215–226 of: Proc. SPIE*, vol. 5006.

Seuntiens, Pieter, Meesters, Lydia, & Ijsselsteijn, Wijnand. 2006. 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.

Seuntiens, Pieter, Vogels, Ingrid, & van Keersop, Arnold. 2007. Visual experience of 3D-TV with pixelated ambilight. *Pages 339–344 of: Proc. 10th Ann. Int. Workshop Presence*.

Sullivan, Gary J, Ohm, Jens-Rainer, Han, Woo-Jin, & Wiegand, Thomas. 2012. Overview of the High Efficiency Video Coding (HEVC) Standard. *IEEE Trans. Circuits and Systems for Video Tech*.

Suryakumar, Rajaraman. 2005. *Study of the dynamic interactions between vergence and accommodation*. Ph.D. thesis, University of Waterloo.

Tam, Wa James, & Zhang, Liang. 2006. 3D-TV content generation: 2D-to-3D conversion. *Pages 1869–1872 of: Multimedia and Expo, 2006 IEEE International Conference on*. IEEE.

Tourancheau, Sylvain, Wang, Kun, Bułat, Jarosław, Cousseau, Romain, Janowski, Lucjan, Brunnström, Kjell, & Barkowsky, Marcus. 2012. Reproducibility of crosstalk measurements on active glasses 3D LCD displays based on temporal characterization. *Pages 82880Y–82880Y of: IS&T/SPIE Electronic Imaging*. International Society for Optics and Photonics.

Ukai, Kazuhiko, & Howarth, Peter A. 2008. Visual fatigue caused by viewing stereoscopic motion images: Background, theories, and observations. *Displays*, **29**(2), 106–116.

VQEG, Video Quality Experts Group, *et al.* 2009. *VQEG HDTV test plan*.

Wang, Yao, & Zhu, Qin-Fan. 1998. Error control and concealment for video communication: A review. *Proceedings of the IEEE*, **86**(5), 974–997.

Watson, Andrew B, & Ahumada, Albert J. 2012. 64.3: Flicker Visibility: A Perceptual Metric for Display Flicker. *Pages 957–959 of: SID Symposium Digest of Technical Papers*, vol. 42. Wiley Online Library.

Wheatstone, Charles. 1838. Contributions to the physiology of vision.–Part the first. On some remarkable, and hitherto unobserved, phenomena of binocular vision. *Philosophical transactions of the Royal Society of London*, **128**, 371–394.

Winkler, Stefan. 2005. *Digital video quality*. Wiley.

Woods, Andrew J. 2011. How are crosstalk and ghosting defined in the stereoscopic literature? *Page 78630Z of: Proc. SPIE*, vol. 7863.

Woods, Andrew J. 2012. Crosstalk in stereoscopic displays: a review. *Journal of Electronic Imaging*, **21**(4), 040902–040902.

Yoshikawa, Hiroshi, & Yamaguchi, Takeshi. 2012. Recent progress on digital holography for 3D display. *Pages 85570C–85570C of: Photonics Asia*. International Society for Optics and Photonics.

Biography

Kun Wang was born on the 1st of October 1979 in Nanjing, China. He received the Master of Science in Electrical Engineering from the Royal Institute of Technology, Sweden in March 2007. After his graduation he started as a research fellow in Eindhoven University of Technology and later worked at Belden Europe and Alcatel-Lucent, in the Netherlands. In 2009 he started his PhD studies at Mid Sweden University and at Acreo Swedish ICT AB as an industrial PhD student with a research focus on the 3D video quality of experience. Apart from PhD studies, he also actively contributes to other research fields. He has been working in many EU research projects e.g. FP6-MUSE, ISIS and FP7-ALPHA, OASE in the field of broadband access network. He has authored and co-authored 25 international publications.