UNIVERSIDAD POLITÉCNICA DE MADRID

ESCUELA TÉCNICA SUPERIOR
DE INGENIEROS DE TELECOMUNICACIÓN

Analysis of quality of experience in 3D video
systems

Ph. D. Thesis
Tesis Doctoral

Jesús Gutiérrez Sánchez
Ingeniero de Telecomunicación
2016

DEPARTAMENTO DE SEÑALES, SISTEMAS Y
RADIOCOMUNICACIONES

ESCUELA TÉCNICA SUPERIOR
DE INGENIEROS DE TELECOMUNICACIÓN

**Analysis of quality of experience in 3D video
systems**

Tesis Doctoral

Autor:

Jesús Gutiérrez Sánchez

Ingeniero de Telecomunicación

Universidad Politécnica de Madrid

Director:

Narciso García Santos

Doctor Ingeniero de Telecomunicación

Universidad Politécnica de Madrid

2016

TESIS DOCTORAL

Analysis of quality of experience in 3D video systems

Autor: Jesús Gutiérrez Sánchez

Director: Narciso García Santos

Tribunal nombrado por el Mgfc. y Excmo. Sr. Rector de la Universidad Politécnica de Madrid, el día de de 2015.

Presidente: D.

Vocal: D.

Vocal: D.

Vocal: D.

Secretario: D.

Realizado el acto de defensa y lectura de la Tesis el día de de 2016 en

Calificación:

EL PRESIDENTE

LOS VOCALES

EL SECRETARIO

Abstract

This thesis presents a comprehensive study of the evaluation of the Quality of Experience (QoE) perceived by the users of 3D video systems, analyzing the impact of effects introduced by all the elements of the 3D video processing chain. Therefore, various subjective assessment tests are presented, particularly designed to evaluate the systems under consideration, and taking into account all the perceptual factors related to the 3D visual experience, such as depth perception and visual discomfort.

In particular, a subjective test is presented, based on evaluating typical degradations that may appear during the content creation, for instance due to incorrect camera calibration or video processing algorithms (e.g., 2D to 3D conversion). Moreover, the process of generation of a high-quality dataset of 3D stereoscopic videos is described, which is freely available for the research community, and has been already widely used in different works related with 3D video. In addition, another inter-laboratory subjective study is presented analyzing the impact of coding impairments and representation formats of stereoscopic video. Also, three subjective tests are presented studying the effects of transmission events that take place in Internet Protocol Television (IPTV) networks and adaptive streaming scenarios for 3D video. For these cases, a novel subjective evaluation methodology, called *Content-Immersive Evaluation of Transmission Impairments* (CIETI), was proposed, which was especially designed to evaluate transmission events simulating realistic home-viewing conditions, to obtain more representative conclusions about the visual experience of the end users. Finally, two subjective experiments are exposed comparing various current 3D displays available in the consumer market, and evaluating perceptual factors of Super Multiview Video (SMV) systems, expected to be the future technology for consumer 3D displays thanks to a promising visualization of 3D content without specific glasses.

The work presented in this thesis has allowed to understand perceptual and technical factors related to the processing and visualization of 3D video content, which may be useful in the development of new technologies and approaches for QoE evaluation, both subjective methodologies and objective metrics.

Resumen

Esta tesis presenta un estudio exhaustivo sobre la evaluación de la calidad de experiencia (QoE, del inglés *Quality of Experience*) percibida por los usuarios de sistemas de vídeo 3D, analizando el impacto de los efectos introducidos por todos los elementos de la cadena de procesamiento de vídeo 3D. Por lo tanto, se presentan varias pruebas de evaluación subjetiva específicamente diseñadas para evaluar los sistemas considerados, teniendo en cuenta todos los factores perceptuales relacionados con la experiencia visual tridimensional, tales como la percepción de profundidad y la molestia visual.

Concretamente, se describe un test subjetivo basado en la evaluación de degradaciones típicas que pueden aparecer en el proceso de creación de contenidos de vídeo 3D, por ejemplo debidas a calibraciones incorrectas de las cámaras o a algoritmos de procesamiento de la señal de vídeo (p. ej., conversión de 2D a 3D). Además, se presenta el proceso de generación de una base de datos de vídeos estereoscópicos de alta calidad, disponible gratuitamente para la comunidad investigadora y que ha sido utilizada ampliamente en diferentes trabajos relacionados con vídeo 3D. Asimismo, se presenta otro estudio subjetivo, realizado entre varios laboratorios, con el que se analiza el impacto de degradaciones causadas por la codificación de vídeo, así como diversos formatos de representación de vídeo 3D. Igualmente, se describen tres pruebas subjetivas centradas en el estudio de posibles efectos causados por la transmisión de vídeo 3D a través de redes de televisión sobre IP (IPTV, del inglés *Internet Protocol Television*) y de sistemas de *streaming* adaptativo de vídeo. Para estos casos, se ha propuesto una innovadora metodología de evaluación subjetiva de calidad vídeo, denominada *Content-Immersive Evaluation of Transmission Impairments* (CIETI), diseñada específicamente para evaluar eventos de transmisión simulando condiciones realistas de visualización de vídeo en ámbitos domésticos, con el fin de obtener conclusiones más representativas sobre la experiencia visual de los usuarios finales. Finalmente, se exponen dos experimentos subjetivos comparando varias tecnologías actuales de televisores 3D disponibles en el mercado de consumo y evaluando factores perceptuales de sistemas *Super Multiview Video* (SMV), previstos a ser la tecnología futura de televisores 3D de consumo, gracias a una prometedora visualización de contenido 3D sin necesidad de gafas específicas.

El trabajo presentado en esta tesis ha permitido entender los factores perceptuales y técnicos relacionados con el procesamiento y visualización de contenidos de vídeo 3D, que pueden ser de utilidad en el desarrollo de nuevas tecnologías y técnicas de evaluación de la QoE, tanto metodologías subjetivas como métricas objetivas.

Agradecimientos

Ante todo, quiero agradecer especialmente a mi tutor Narciso García. En primer lugar, por darme la posibilidad de embarcarme en esta aventura formando parte de un grupo tan genial como el GTI y, en segundo lugar, por su apoyo, ayuda, guía y dedicación. Sin él esta tesis no hubiera sido posible, ni todas las vivencias que he tenido durante este tiempo trabajando en ella.

Muchas gracias también al resto de profesores del GTI: a Fernando, en especial, por su inestimable ayuda estando siempre dispuesto a echarme una mano con lo que hiciera falta (pruebas subjetivas, viajes a Montegancedo, papeleos, consejos, ...); a Julián y F por su atención y ayuda que ha ido más allá de los trabajos en los que hemos colaborado; a Luis por su ayuda y apoyo, especialmente en mi aterrizaje en el GTI; y a Nacho, por hacerme volver a sentirme estudiante de verdad durante los cursos de doctorado.

También quiero agradecer, muy especialmente a Pablo Pérez, por hacerme partícipe de sus ideas y permitirme colaborar con él en gran parte del trabajo presentado en esta tesis.

Muchísimas gracias al resto del GTI porque es imposible encontrar un grupo de trabajo tan fantástico, en el que no solo cuento con compañeros, sino también con amigos. Me extendería demasiado si detallara todos los motivos por los que les agradezco haber trabajado con ellos, así que los enumero sin orden ni concierto. A Esther por su inestimable ayuda en las pruebas subjetivas, por su apoyo y por estar ahí siempre. A Samira por su ayuda en los muchos trabajos en los que hemos colaborado. A Virginia, Ana y Susana, que junto con Tomás y Rafa dan al grupo ese aire juvenil tan necesario. A los veteranos Carlos Cuevas y Carlos Roberto por su ayuda y apoyo durante todo este tiempo. A todos los demás que siguen en el grupo o pasaron por él durante este tiempo: David, Marcos, Víctor, Guillermo, Jon, Sergio, etc. De entre éstos tengo que destacar al camarada Maykel, el “fratello” Gianlu y el sonriente Sasho, con los que he compartido (y espero seguir haciéndolo) grandes momentos dentro y fuera del curro; y, por supuesto, a Massi, que desde el primer día que llegamos juntos al GTI ha sido gran compañero y amigo, capaz de hacerme comer paella por la noche. También, aunque algunos me dejaré en el tintero, no se me pasa agradecer a Dani la ayuda, el apoyo y la calma que me transmite; a César por su amistad desde el primer día y porque siempre nos quedará Berlín; y a Pablo por estar ahí siempre dentro y fuera del “terreno de juego”. Finalmente, agradecer especialmente a mis compañeros, coinquilinos, amigos y hermanos (seguro que mi madre lo acepta) Raul y Fili, por todo, todo y todo.

También quiero agradecer su ayuda a todos aquellos que hicieron tan especiales, en lo profesional y en lo personal, mis estancias en Nantes y Berlín. Por un lado, a Patrick, Marcus, Jing, Matthieu, Romain, Romuald, Emilie, etc. Y por otro lado, a Alexander, Pierre, Savvas, Miguel, etc.

Gracias a Amy Reibman y a los ya mencionados Patrick Le Callet y Fernando Jaureguizar, que aportaron valiosas sugerencias y comentarios con la revisión de la primera versión de la tesis. Así como a los miembros del tribunal, por su disposición.

Tampoco quiero olvidarme de todos aquellos que han participado en las pruebas subjetivas que se presentan en esta tesis por su esencial colaboración.

Asimismo, aunque no estén directamente ligados al trabajo aquí presentado, quiero expresar mi agradecimiento a otras personas sin las que no hubiese sido capaz de llevarlo a cabo. Por un lado, a los amigos que han hecho que mi estancia en Madrid durante estos años sea muy difícilmente mejorable: los “encomienderos” Ramon y José, Javi, Ali, Eva, los madrileños, los berlineses, los valencianos,... Por otro lado, a los *makis* y al resto de amigos de Orihuela, que, a pesar del tiempo y la distancia, siempre hacen que me sienta en casa. Y a Elena, por el apoyo, el ánimo, la paciencia, los motivos y por estar ahí en la última etapa de este trabajo, sin ella hubiese sido infinitamente más difícil acabarlo.

Finalmente, un agradecimiento inmenso a mi familia. Especialmente, a mi madre, mi padre y mi hermanica, sin los que no hubiese podido llegar hasta aquí ni a ningún lado. Y a Totón, al final no soy médico, pero, si todo va bien, sí doctor.

Contents

Abstract	I
Resumen	III
Agradecimientos	V
List of Figures	IX
List of Tables	XIII
List of Abbreviations	XV
1 Introduction	1
1.1 Motivation	1
1.2 Overview	2
2 Quality of Experience in 3D video	5
2.1 Introduction	5
2.2 Definition of QoE	5
2.3 Perceptual factors	7
2.4 System factors	14
2.4.1 Content production	14
2.4.2 Coding	19
2.4.3 Transmission	25
2.4.4 Display	29
2.5 Evaluation of 3D QoE	37
2.5.1 Subjective evaluation methodologies	37
2.5.2 Objective evaluation metrics	44
3 Content production	49
3.1 Introduction	49
3.1.1 SoA 3D databases	49
3.1.2 Quality evaluation	50
3.2 Generation of a freely available database of high quality stereoscopic 3D videos	51
3.2.1 Equipment	51
3.2.2 Depth map generation	53
3.2.3 Content diversity	54
3.2.4 Applicability of the dataset	55

3.3	Subjective evaluation of typical degradations	56
3.3.1	Subjective test setup	56
3.3.2	Experimental results	61
3.4	Conclusions	68
4	Coding	71
4.1	Introduction	71
4.1.1	Quality evaluation	71
4.2	Subjective evaluation of 3D coding artifacts	72
4.2.1	Subjective test setup	73
4.2.2	Experimental results	77
4.3	Conclusions	87
5	Transmission	89
5.1	Introduction	89
5.1.1	Quality evaluation	90
5.2	Proposed subjective methodology	92
5.3	3D video over IPTV	95
5.3.1	Subjective test setup	95
5.3.2	Experimental results	99
5.4	Adaptive streaming of 3D video	109
5.4.1	Subjective Experiment 1	110
5.4.2	Subjective Experiment 2	118
5.5	Validation of the methodology	130
5.5.1	Subjective test setup	130
5.5.2	Experimental results	132
5.6	Applications of the results from the subjective tests	136
5.7	Conclusions	138
6	Display	141
6.1	Introduction	141
6.1.1	Quality evaluation	141
6.2	Comparison of consumer television technologies for 3D display	143
6.2.1	Subjective test setup	144
6.2.2	Results	148
6.3	Evaluation of Super Multiview Video	155
6.3.1	Subjective test setup	156
6.3.2	Experimental results	159
6.4	Conclusions	165
7	Conclusions and future work	167
7.1	Conclusions	167
7.2	Future work	170
	Bibliography	171

List of Figures

2.1	3D QoE model	7
2.2	Example of perceptual masking	8
2.3	Diagram of depth of focus and depth of field	9
2.4	Monocular cues [112]	10
2.5	Basics of stereoscopic viewing	11
2.6	Stereoscopic comfort zone	13
2.7	Video processing chain	14
2.8	Examples of 3D cameras	15
2.9	Diagrams of camera configurations	15
2.10	Multiview camera systems [192]	15
2.11	Principles of depth cameras [38]	16
2.12	Incorrect parallaxes [12]	18
2.13	Cardboard effect in the right image in comparison to the original (left) [12] .	18
2.14	Frame-compatible formats	20
2.15	V+D representation	21
2.16	MVD representation of three views [68]	21
2.17	Typical MVC picture coding structure	22
2.18	Coding structure of 3D-HEVC [68]	24
2.19	Blockiness effect [12]	24
2.20	Diagram of HAS operation	26
2.21	3DTV signal formats [93]	27
2.22	Flowchart of processing FCC format	28
2.23	Effect of video losses on SbS video	29
2.24	Stereoscopes	31
2.25	Prototype of a HMD [132]	31
2.26	Anaglyph glasses	31
2.27	Schemes for time and polarization multiplexing displays [48]	32
2.28	Schemes of autostereoscopic displays	33
2.29	Super-multiview condition [187]	33
2.30	Scheme of the accommodation-vergence conflict [153]	36
2.31	Depth of field when viewing a conventional multiview display (above) and an SMV display (below) [187]	36
2.32	Diagrams of presentation of sequences for DSIS methodology	38
2.33	Diagrams of presentation of ACR methodology	39
2.34	Diagrams of presentations for pair comparison methodology	40
2.35	Position of two screens in time-parallel presentation	40
2.36	Rating scale used in SSMM	41
2.37	PSNR performance [219]	45

2.38	Generic diagram of an HVS model [219]	46
2.39	Diagrams of the metrics according to the required data from the reference	48
3.1	Sequences previews	53
3.2	Depth maps previews (frame 0)	54
3.3	Spatial, temporal, and coding characteristics	55
3.4	Test environment at UPM	57
3.5	Disparity offsets [209]	60
3.6	Processing chain for 2D to 3D conversion [209]	60
3.7	Comparison of the test conditions	62
3.8	Comparison of the test conditions for both sources	63
3.9	Comparison among the different laboratories	65
4.1	UPM test environment	73
4.2	Comparison of the different encoders and formats	78
4.3	Comparison among the 2D and 3D test conditions	79
4.4	Comparison among the different contents	80
4.5	IVC Test Environment	82
4.6	Comparison between results from IVC and UPM	83
4.7	Comparison results between IVC and UPM for different contents	85
4.8	Monte-Carlo simulation histogram	86
5.1	Diagram of the structure of the test sequences	93
5.2	Examples of frames of the first three segments of the test sequence	93
5.3	Example of questionnaire	95
5.4	SI vs TI	97
5.5	Impact of video losses	101
5.6	Impact of audio losses	103
5.7	Impact of bitrate and framerate reductions	104
5.8	Impact of outage events	105
5.9	Impact of video losses in 2D and 3D	106
5.10	Impact of audio losses in 2D and 3D	106
5.11	Impact of bitrate and framerate reductions in 2D and 3D	107
5.12	Impact of outages in 2D and 3D	107
5.13	Results for 3D performance	109
5.14	SI vs TI	111
5.15	Global results for the considered strategies	113
5.16	Results for video freeze events	115
5.17	Comparison of bitrate reductions per segment	116
5.18	Comparison between asymmetric coding and switching to 2D per segment	117
5.19	Comparison of freezing events per segment	117
5.20	Visual difficulties for each type of degradation	118
5.21	SI vs TI	121
5.22	Results for decreasing the quality	124
5.23	Results for increasing the quality	126
5.24	Results for quality oscillations	127

5.25	Results regarding visual discomfort	128
5.26	Observers' vote distribution	130
5.27	Comparison of the results with ACR and the proposed method	133
5.28	Scatter plot of the results with ACR and CIETI	135
5.29	Percentage of observers that felt visual discomfort	136
5.30	QuEM architecture design [140]	137
5.31	Diagram of the proposed rule-based system [62]	137
6.1	Test environment	146
6.2	Results for picture quality	149
6.3	Results for depth quality	151
6.4	Results for visual discomfort	152
6.5	Global evaluations	154
6.6	Displays ranked by the observers	155
6.7	SMV scenarios	156
6.8	Visual comfort results for different values of VSS	161
6.9	Smoothness with the stereoscopic display	162
6.10	Smoothness with the autostereoscopic display	163
6.11	3D quality for the maximum VD and for each sequence	164

List of Tables

2.1	Five-grade quality, degradation, and comfort scales	38
2.2	Comparison scale [89]	40
3.1	Characteristics of the sequences of NAMA3DS1 dataset	52
3.2	Test equipment at UPM	57
3.3	Source sequences	59
3.4	Test conditions	60
3.5	R_{ORD} matrix	61
3.6	Settings of the other test laboratories	64
3.7	Comparisons between HRC11 and HRC12	66
3.8	Comparisons between HRC11 and HRC6	66
3.9	Comparisons between HRC12 and HRC6	66
3.10	Comparisons between HRC11 and HRC8	67
3.11	Comparisons between HRC10 and HRC1	67
3.12	Comparisons between HRC10 and HRC4	68
3.13	Comparisons between HRC10 and HRC5	68
4.1	Source sequences	74
4.2	Test conditions	75
4.3	R_{ORD} matrix for Experiment 1	76
4.4	Observers at UPM	77
4.4	R_{ORD} matrix for Experiment 2	77
4.6	Test equipment at IVC	82
4.7	Observers at IVC	82
5.1	Test Sequences	97
5.2	Considered transmission errors	98
5.3	Specific error patterns considered in the tests	98
5.4	Results of Wilcoxon signed-rank test	108
5.5	Results for preference between 2D and 3D presentations	109
5.6	Test sequences	111
5.7	Considered strategy patterns	112
5.8	Evaluation scale	119
5.9	Test sequences	120
5.10	Quality levels	121
5.11	Test conditions	122
5.12	Comparison between MOS of subjects with and without visual discomfort	129
5.13	Considered strategy patterns	131

6.1	Characteristics of the displays	146
6.2	Sequences used in the tests	147
6.3	Characteristics of the SMV sequences	158
6.4	Combinations of VSS and VD tested	159
6.5	Comfortable values of VSS	160
6.6	Minimum value of VD per value of VSS	161

List of Abbreviations

3DTV	3D Television
ACR	Absolute Category Rating
ANOVA	Analysis of Variance
AVC	Advanced Video Coding
BT	Bradley-Terry
CI	Confidence Interval
CIETI	Content-Immersive Evaluation of Transmission Impairments
DCR	Degradation Category Rating
DCT	Discrete Cosine Transform
DIBR	Depth-Image Based Rendering
DSCQS	Double Stimulus Continuous Quality Scale
DSI	Depth Spatial Information
DSIS	Double Stimulus Impairment Scale
DTI	Depth Temporal Information
DVB	Digital Video Broadcasting
FCC	Frame-Compatible Compatible
FR	Full-Reference
FTV	Free-Viewpoint Television
Full-HD	Full High Definition
HAS	HTTP Adaptive Streaming
HEVC	High Efficiency Video Coding
HMD	Head-Mounted Displays

HRC	Hypothetical Reference Circuit
HVS	Human Visual System
IDR	Instantaneous Decoder Refresh
IF	Influence Factor
IPTV	Internet Protocol Television
ITU	International Telecommunication Union
IVC	Images et Video Communications
JND	Just Noticeable Difference
JVT	Joint Video Team
LDI	Layered Depth Image
MOS	Mean Opinion Score
MPEG	Moving Pictures Experts Group
MSE	Mean Square Error
3DV	3D Video Coding
MVC	Multiview Video Coding
MVD	Multi-view plus Depth
NAMA3DS1	Nantes-Madrid-3D-Stereoscopic-V1
NR	No-Reference
NR-B	NR-Bitstream-based
NR-H	NR-Hybrid
NR-P	NR-Pixel-based
ORD	Optimized Rectangular Design
OTT	Over-the-Top
P2P	Peer-to-Peer
PC	Pair Comparison
PoE	Preference of Experience

PSNR	Peak Signal-to-Noise Ratio
QoE	Quality of Experience
QoS	Quality of Service
RMSE	Root Mean Squared Error
RR	Reduced-Reference
S3D	Stereoscopic 3D
SAMVIQ	Subjective Assessment Methodology for Video Quality
SbS	Side-by-Side
SEI	Supplemental Enhancement Information
SI	Spatial Perceptual Information
SMV	Super Multiview
SSCQE	Single Stimulus Continuous Quality Evaluation
SSIM	Structural Similarity
SSMM	Single Stimulus Multimedia
STB	Set-top Box
SVC	Scalable Video Coding
TI	Temporal Perceptual Information
ToF	Time-of-Flight
UHD	Ultra-High-Definition
UPM	Universidad Politécnica de Madrid
V+D	Video plus Depth
VD	View Density
VQEG	Video Quality Experts Group
VQM	Video Quality Model
VR	View Range

1 Introduction

1.1. Motivation

Although previous efforts were made to introduce 3D video technology in entertainment media in the 1950s and 1980s, the last attempt to establish 3D video technologies in the entertainment market has taken place recently (probably ignited by the appearance of novel 3D movies, such as *The Polar Express* [130]). This time, the technological advances and the motivations of the industry and research community have allowed that 3D video technology is nowadays widely spread across the consumer market, allowing the users the possibility of watching 3D content, not only at cinemas, but also at households with 3D television sets, and even in mobile devices, such as smartphones. Actually, the current renaissance of 3D video has been compared to previous revolutions in the audiovisual media, such as the introduction of audio in movies, or the color television [3].

As with any other multimedia system, service or application, essentially addressed to people, to know whether the end users' expectations are satisfied is crucial for the success and definite implantation of 3D video technologies in the consumer market. Therefore, an extensive research work is being carried out during the last years dealing with the evaluation of the experiences of end users of multimedia services, to obtain representative conclusions of how these services fulfill the users' demands. In fact, to extend and actualize the traditional approaches for evaluating the quality of audiovisual systems and applications, a theoretical framework has been recently issued, defining the term Quality of Experience (QoE) and formulating all the aspects in relation with the visual experience of end users of multimedia applications, such as the influence factors, application fields, and evaluation methodologies [111].

Consequently, the research work carried out during the last years with relation to 3D video, has led to the identification of various factors that have hindered, for now, the absolute establishment of 3D video technologies in the consumer market. Mainly, these factors are: the lack of high quality 3D video content, the need of specific glasses to watch 3D videos with some visualization systems, and certain annoying symptoms that users feel when watching 3D video, referred as visual discomfort.

Therefore, there is still the need to research and develop 3D video systems to achieve the entire acceptance of the consumers and provide them a totally satisfactory QoE. In this sense, on one side, the investigation on technical advances of 3D video technologies is crucial to improve all the elements of the whole processing chain of 3D video. Actually, all the processes involved in this chain may introduce effects that influence the QoE of the end users, such as calibration issues related to capture, impairments introduced

by encoding the video, errors in the transmission schemes, and visualization effects regarding 3D displays.

On the other side, it is essential to reliably analyze how these technologies impact the viewers' QoE, as well as continuing the studies concerning the perceptual factors involved in the 3D visual experience, such as depth perception and visual discomfort. Therefore, the research on methodologies to evaluate the QoE of the end users of 3D video systems is required, especially regarding subjective assessment tests. These experiments are based on the evaluation of the test conditions under study by a certain number of observers. Thus, since the evaluation is directly carried out by people, these tests are the most reliable way to analyze the visual experience of the users. Nevertheless, the design of subjective test methodologies to obtain robust and representative conclusions of what end users perceive is not trivial, since it should be carried out taking into account the application under study. For instance, regarding 3D video, several factors are involved in the QoE, not only perceptual, but also related to the viewing environment. For example, how the end users would use 3D video technologies in real life should be considered. Furthermore, the results from subjective experiments will help in the understanding of the Human Visual System (HVS), the improvement of 3D video technologies, and the development of algorithms to automatically estimate the end users' QoE (i.e., objective metrics) with the main objective, among others, of knowing almost in real time the QoE of the end users of 3D video delivery systems.

These activities and developments may result in the definite establishment of 3D video technologies in the consumer market, fulfilling the demands and expectations of end users and providing a comfortable and satisfactory visual experience, more immersive and complete than with conventional video.

Taking this into account, the work presented in this thesis was carried out with the objective of understanding how the different elements of the 3D video processing chain impact the end users' QoE and obtaining reliable methodologies for the evaluation of their visual experience.

1.2. Overview

The aim of this thesis was to evaluate the QoE of the users of 3D video, focusing on the impact of the typical effects introduced by all the components of the video processing chain, and on the development of reliable methodologies to obtain representative conclusions of the visual experience perceived by the end users in real usage scenarios.

Thus, one of the main contributions of this thesis is the analysis, by means of subjective experiments, of the impact on the QoE of: typical degradations introduced by the creation of stereoscopic 3D (S3D) video (e.g., caused by an incorrect calibration of the cameras), impairments produced by different encoder settings and representation formats for 3D video (e.g., Side-by-Side (SbS), Full High Definition (Full-HD)), effects caused by transmission events or errors in Internet Protocol Television (IPTV) networks and adaptive streaming scenarios, and factors in relation to current consumer

3D displays (e.g., autostereoscopic displays, stereoscopic displays with active and passive glasses, etc.) and to future Super Multiview Video (SMV) displays.

In addition, a database of stereoscopic videos freely-available for the research community to carry out activities in relation to 3D video (e.g., subjective tests, algorithms validation, etc.) was created and used in various tests, even under the scope of Video Quality Experts Group (VQEG) projects.

Finally, a novel methodology for subjective evaluation of typical events occurring in 3D video delivery scenarios has been proposed, aiming at obtaining representative results of the QoE perceived by the end users at their homes, thus mimicking real-viewing conditions.

Taking this into account, this thesis has been structured dedicating one chapter to each element of the 3D video processing chain, after an introductory chapter to 3D QoE. In particular, the thesis is organized as follows:

- Chapter 2 presents general aspects in relation with QoE evaluation of 3D video technologies. Firstly, the definition of the term QoE is presented. Then, the perceptual factors that are involved in the 3D visual experience of the viewers are described, in addition to those effects that may appear through the 3D video processing chain and may impact the QoE of the end users of 3D video systems. Finally, general aspects are exposed regarding the evaluation of QoE via subjective tests and objective metrics.
- Chapter 3 addresses the study of the aspects related to the production of 3D video content. In particular, the process of generation of a high-quality freely available dataset of S3D videos is presented, which was created in collaboration with the *Images et Video Communications* (IVC) laboratory of the *Université de Nantes*. In addition, a subjective test carried out to evaluate typical degradations that may appear in the content production process is presented. This test was carried out under the scope of the 3D Television (3DTV) project by VQEG.
- Chapter 4 deals with the evaluation of the impact of coding effects on the viewers' QoE. Concretely, an inter-laboratory subjective test carried out in collaboration with the IVC laboratory of the *Université de Nantes* is presented, which was focused on the analysis of different coding degradations and representation formats of 3D video.
- Chapter 5 addresses the problematic of evaluating transmission events in 3D video delivery systems. In particular, a novel methodology for subjective testing is proposed, which is based on the simulation of real home-viewing conditions with the aim of obtaining more representative results of what end users perceive at their homes. Moreover, three subjective studies regarding S3D video transmission through IPTV networks and adaptive streaming scenarios are presented.
- Chapter 6 deals with the quality evaluation of 3D displays, with the description of two subjective experiments. The first one, was carried out to compare the performance of different consumer 3D displays. The second one was performed to investigate a methodology to evaluate different factors in relation with SMV displays, which seem to be the most promising glasses-free 3D display technology.

- Chapter 7 presents the general conclusions of the thesis, and provides some proposals for future research continuing the work presented here.

2 Quality of Experience in 3D video

2.1. Introduction

This chapter introduces the main aspects related to the evaluation of the QoE with 3D video, establishing a conceptual basis for the rest of the thesis. So, initially, the definition of the term QoE is addressed, as well as the description of the main factors influencing it.

Then, among those factors, the most interesting within the scope of this thesis are related to perceptual aspects of the HVS, and to effects introduced by the systems involved in the 3D video processing chain. In fact, the rest of the chapters of this thesis deal with evaluations of perceptual factors related to each element of the processing chain. Therefore, they are described in detail in sections 2.3 and 2.4.

Finally, section 2.5, presents the general aspects concerning the evaluation of 3D QoE by means of subjective assessment tests and objective metrics.

2.2. Definition of QoE

In multimedia communications, the evaluation of the quality of a system, service, or application is a key factor, either in the design process or during operation, to understand its performance. Moreover, since generally these systems are addressed to be used by people, in the consumer or entertainment market, it is crucial to know the quality provided by the services or applications to assess their possibilities of success.

Traditionally, quality in multimedia communications has been referred from an engineering perspective as reflected by the widespread use of the term Quality of Service (QoS), which is defined by the International Telecommunication Union (ITU) as the “totality of characteristics of a telecommunications service that bear on its ability to satisfy stated and implied needs of the user of the service” [85]. This definition suggests a formulation from a perspective explicitly centered in a service provider point of view, and does not cover many factors involved in communications systems. Therefore, a thorough work has been carried out during the last years by the research community to formulate a better definition of such an important concept. For instance, the ITU extended this definition including subjective effects (e.g., user expectations and context) to the system factors, resulting in the statement of the term QoE as “the overall acceptability of an application or service, as perceived subjectively by the end user”. In essence, this definition shifts the perspective of the formulation of quality to the end users, who are the real judges of the performance of a communication service.

Furthermore, lately, an international consortium of research institutions within the European Network of Excellence “Qualinet” (COST IC2003) have developed a complete theoretical framework in relation with quality evaluation of multimedia services. In fact, the work carried out by this group resulted in the most precise definition of QoE as “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” [111].

In addition, they exhaustively described the application areas of QoE (covering multimedia content distribution, web and cloud applications, multimedia learning, sensory experiences, or haptic communications), and each Influence Factor (IF) that affect the QoE. Since several features may impact the users’ QoE, the IFs have been grouped into three categories [154]:

- Human IFs: Are those properties that may influence the QoE that are subjective and intrinsic of the human users, which make them highly complex. In essence, they could be classified into low-level (early sensory processing) and high-level (cognitive processing) factors. On one side, low-level factors are those related to the physical, emotional, and mental constitution of the user, going from visual and auditory acuity, human perception mechanisms, gender, or age, to users’ mood, motivation, or attention. On the other side, high-level factors are those involved in the understanding and interpretation of stimuli, such as socio-cultural and educational background, socio-economic situations, consuming behaviors, preferences, etc.
- System IFs: Are those technical properties that impact the quality of an application or service, which could be related to the content, the media, the network, or the device. Firstly, the content itself could notably influence the visual experience of the end users, as well as properties characterizing the content, such as bandwidth and dynamic range for audio content, and amount of detail and motion activity for video content, as well as amount of depth when 3D video is considered. Secondly, examples of media-related factors are those linked with the media configuration, such as encoding, resolution, framerate, etc. Also, network-related factors include those properties of the transmission system, like the bandwidth, delay, jitter, loss rates, etc. Finally, the device-related factors are those associated to the systems of the audiovisual signal processing chain, for instance, the technical features of the displays or loudspeakers.
- Context IFs: Are those properties that describe the users’ environment, in terms of physical (e.g., location), temporal (e.g., experience), social (e.g., inter-personal relations during the usage of the application), economic (e.g., costs of the service), and technical (e.g., interaction between system of interest and other systems) characteristics. Thus, some of these factors are strongly linked to human and systems IFs, and are very difficult to cover in QoE evaluation.

Moreover, in contrast to images and 2D video that have been extensively investigated in terms of quality evaluation [219], the new factors related to 3D watching experiences [19] are still under research. In fact, one of the main issues regarding 3D QoE

Figure 2.1: 3D QoE model

evaluation is its multi-dimensionality, since apart from picture quality (that allow the evaluation of 2D video quality), other agents take place in the overall 3D visual experience. In essence, as presented by Seuntjens [165], the 3D visual experience is mainly determined by: image quality, depth perception, and visual comfort. In addition, some secondary factors should be considered, such as naturalness and sense of presence, as depicted in Figure 2.1. This model has been considered by the ITU in the standardization of 3D QoE evaluation [92], formulating the primary (image quality, depth perception, and visual comfort) and secondary (naturalness and sense of presence) perceptual dimensions of the 3d visual experience. In addition, it is worth noting that depth perception is composed by two different factors: depth quantity (how much depth is perceived) and depth quality (how plausible is the depth rendering) [20].

Taking this into account, since context effects are related to personal aspects out of the scope of this work, the following sections present the main human and system factors influencing the 3D QoE of end users, and the paradigms for evaluating the visual experience of the users of 3D video services.

2.3. Perceptual factors

Among the human IFs described before, those related to the perception of the visual stimuli by the HVS are of topmost importance, since they allow to understand how people perceive visual content. In addition, this helps to investigate on processing techniques to apply to visual signals in order to provide the highest possible QoE. Thus, taking into account the model presented in Figure 2.1, the main factors influencing each element of the model are described in the following.

Image quality Firstly, the perception of image quality, which has been extensively studied in the past years [219], does not significantly differ from the images and 2D video cases [165]. Therefore, the main factors of the HVS related to the perception of pictorial quality are:

Figure 2.2: Example of perceptual masking

- **Visual sensitivity:** It determines how the HVS respond to light and visual stimuli. In essence, the response of the HVS depends on the spatial and temporal frequencies of the stimuli, as well as on the luminance contrasts around the fixation points (more than on the absolute luminance, as established by the Weber-Fechner law). In addition, visual sensitivity is influenced by the non-uniform distribution of the photo receptors on the retina, so it decreases away from the fovea [154]. Thus, contrast sensitivity functions have been widely investigated to quantify these effects and they have been also used in visual models applied to estimate the QoE [219].
- **Color perception:** It is one of the most difficult factors of the HVS to model, due to its subjectivity. In addition, although it is a very important aspect regarding human visual perception, it is not considered in many visual quality models, since the HVS is less sensitive to color than to luminance [219].
- **Spatial and temporal masking:** It is based on the impossibility to perceive a certain stimulus in the presence of another one. The masking mechanism of the HVS is composed of a spatial and a temporal component. The spatial masking results from the lower sensitivity of the HVS to perceive artifacts in high-frequency regions, so highly textured parts of the image could hide certain degradations (e.g., noise) that are much more perceptible in uniform regions. This effect is depicted in Figure 2.2, where a patch with the same amount of noise has been inserted in the original image (left), but in the higher part (uniform) in center image and in the right part (textured) in the right image, being much more noticeable in the center image. On the other side, temporal masking is based on the elevation of visibility thresholds with temporal intensity variations, so in video sequences it could be harder to notice certain stimuli in scenes with high motion activity than in static scenes [219].
- **Attention:** The perceived quality also depends on the part of the image where the degradations appear, since they are more noticeable when they are located around elements or parts of the image that center the attention of the viewers (e.g., the artifacts are generally more perceptible in the foreground elements than in the background). This fact has encouraged the research on identifying regions of interest (e.g., eye-tracking) and attention models for quality evaluation applications [71].

Figure 2.3: Diagram of depth of focus and depth of field

Depth perception Secondly, the HVS extracts information from several sources to make possible the perception of the world in three dimensions. These sources are called depth cues and can be classified into oculomotor cues, which are related to the position and tension of the muscles of the eye, and visual cues, which in turn, could be divided into monocular (work with one eye, so in two dimensions) and binocular (depend on both eyes, thus work in three dimensions) cues [51, 153].

- Oculomotor cues: The responses of the muscles of the eyes when they are fixed in a target are accommodation and convergence. These two mechanisms interact (one change in convergence entails a change in accommodation, and vice versa) to provide a sharp and comfortable binocular vision, together with the pupillary constrictions.
 - Convergence: It refers to the movement of the eyes to localize an observed object and project it into the fovea. Also, the convergence point of both eyes is projected into corresponding points of the retinas, so there is no disparity between both points in the images captured by both eyes. The region of the space that contains this type of points is called horopter.
 - Accommodation: When the eyes are fixated into an object converging into the point of interest, the accommodation allows that object to be in focus in the retina and to obtain a sharp image of the object, by changing the eyes' optical power. Nevertheless, the HVS has the ability to allow a certain retinal defocus without readjusting the accommodation, so there is a region around the focused point where the objects could be perceived with sufficient sharpness, called depth of field. As depicted in Figure 2.3, the corresponding optical power difference (i.e., in the image space) to the depth of field, is called depth of focus, and it is estimated to range between ± 0.2 and ± 0.5 diopters [110]. Thus, those objects out of the depth of field are perceived blurred, which is an effect that has been used in photography and video processing to provide depth information of the scene to the observers. In addition, it is worth noting that changes in the pupil size provoke changes in the focus of the retinal image, affecting the accommodation mechanism, and thus, the convergence.
- Monocular cues: They could be classified into pictorial cues and motion-based cues [51]. Some of the following cues are represented in Figure 2.4.
 - Pictorial cues: Are those sources of depth information that can be extracted from a static picture, such as:

Figure 2.4: Monocular cues [112]

- Occlusion: When one object is occluded or partially occluded by another object, the former is perceived to be further away than the latter. Thus, this cue provides relative depth information.
- Relative height: When objects could be seen in relation to the horizon, those elements nearer to it are perceived to be further away than those which are more distant to the horizon.
- Relative size: When the relation between the sizes of two objects is known to some extent, the one that occupies less field of view is perceived to be further away than the other.
- Familiar size: When the size of an object is known, it is possible to extract depth information about this object from our prior knowledge in combination with the field of view that it takes up.
- Perspective convergence: Since parallel lines are perceived to converge in the distance, information of the relative distance of two objects can be extracted from this effect when it appears in the field of view.
- Atmospheric convergence: Due to atmospheric components (e.g., air, dust, water droplets, pollution, etc.), objects that are more distant are perceived less sharp and with a slight blue tint.
- Texture gradient: The features of texture elements (e.g., shape, size, etc.) are better differentiated when they are nearer to the observer, since when they are further away they seem to be more bunched together.
- Shadows: The shadow associated to an object provides information about its shape and the position in space.
- Motion-based cues: Are those effects that provide depth information from motion, such as:
 - Motion parallax: When the observer moves, the velocity of the objects within the field of view provide information about their distance, since nearer objects are perceived to move faster than distant objects.

Figure 2.5: Basics of stereoscopic viewing

- Deletion and accretion: When an observer moves (e.g., moving laterally the head), some objects or parts of objects appear from occluded areas while others become covered. This effect provides information about the relative distance of the objects.
- Binocular disparity: Due to the separation between the two eyes (approximately 6 cm), there is a difference in the viewpoint of the scene for each eye, so each one captures slightly different retinal images. The brain is able to fuse both images into one 3D image, called cyclopean image, extracting the depth information of the scene. As depicted in Figure 2.5, the fixation point, set by the convergence mechanism and represented as F , determines the horofter and is projected into corresponding points in the retinas of each eye. On the other side, points x , y , and z stimulate points in the retinas of each eye that are not correspondent, so they present a certain disparity. In fact, those objects in front of the horofter have crossed disparity (e.g., point x and y), while the objects behind it have uncrossed disparity (e.g., point z). In addition, around the horofter there is a region, called Panum's area, containing the points that are projected into the retinas with an acceptable disparity (e.g., point x), so the brain is able to fuse the images from each eye. On the other hand, points outside the Panum's fusional area (e.g., point y) present an excessive disparity, so the brain is not able to fuse both images coming from each eye. This is the operational principle of most of the 3D displays currently available, as detailed described in the following subsection 2.4.4. Thus, one view is showed to the left eye and another view to the right eye with certain disparity between them. These two views are fused by the HVS to obtain a stereoscopic perception of the objects projected in front or behind the screen (i.e., with crossed or uncrossed disparity).

It is also worth noting that the depth perception can considerably change among individuals, so the QoE of each viewer could be different in this sense. One of the causes of this fact is the inter-ocular distance, which could range between

4 cm and 8 cm, considering children and adults [110]. In essence, a smaller interocular distance entails a smaller Panum's area, so less objects would present an acceptable disparity to be correctly perceived. Furthermore, about 10%-15% of people do not perceive appropriately the binocular cues [19], and certain visual abilities (e.g., accommodation) are deteriorated with age.

Visual comfort Furthermore, visual comfort is one of the most crucial factors regarding 3D video visualization, and probably the most decisive issue concerning the acceptance and success of 3D video technologies by the consumers. In fact, an important research work has been done analyzing the possible harmful effects of 3D video, since several viewers have experienced dizziness, double vision, discomfort, eye strain, headache, or nausea during the visualization of 3D content, which are related to the ophthalmologic term of asthenopia [110, 226]. Also, these effects may also affect other high-level perceptual dimensions, such as viewers' emotions [8].

In the literature, two concepts have been mainly used to refer the aforementioned symptoms: visual discomfort and visual fatigue. Although in some cases both terms have been used indistinctly, a different interpretation could be assigned to each one. Firstly, visual fatigue is usually caused by the extended visualization of 3D content during a long time, while visual comfort may refer to instantaneous effects that provoke visual difficulties to obtain a pleasant 3D viewing experience, such as excessive disparity. Secondly, other distinction between both terms could be that the visual fatigue entails a reduction of the HVS performance caused by physiologic alterations and could be objectively measured (e.g., changes in the pupil size or eye movements), while visual comfort refers to the subjective sensation caused by the physiologic change and, thus, only asking to the observer could be known [110]. Therefore, taking into account this differentiation, several studies have been carried out to identify the effects causing visual comfort. In fact, although it is still an active focus of research requiring more comprehensive studies, some conclusions have been already found out. For instance, the main causes of visual comfort already identified are: excessive disparity, stereoscopic degradations that may appear through the processing chain of 3D video, and the accommodation-vergence conflict [190].

Firstly, an excessive disparity between the two views captured by each eye hinders its fusion by the HVS, causing double vision and visual discomfort. This effect appears when the observed object is out of the Panum's area, as happens with points y and z in Figure 2.5. On the contrary, the point x is comfortably observed since it is within the Panum's fusional area. Thus, concerning 3D video visualization, a comfortable zone is usually established (equivalent to the Panum's area) [110], which is obtained from the depth of focus of the HVS, comprising a range of distances in front and behind the screen where the projected objects are correctly visualized by the viewers without causing discomfort. This comfortable region, depicted in Figure 2.6, is taken into account in the processes of capturing and visualizing 3D content, since both the disparity between the acquired stereo views and the viewing distance of the observers are involved.

In addition, as described in the following section, the elements of the 3D video pro-

Figure 2.6: Stereoscopic comfort zone

cessing chain may introduce artifacts that could cause binocular difficulties, as deeply described in the following section 2.4. Especially important are the effects that cause severe differences between the stereo views, such as geometric distortions in the capture process, or asymmetric effects produced in the coding or transmission processes. Moreover, other factors have been identified to provide a more comfortable 3D visual experience to the viewers, such as projecting the objects behind the screen (i.e., with uncrossed disparity) that use to be more comfortable than in front of it (i.e., with crossed disparity). Finally, it is also worth mentioning that visual discomfort may also be caused by the fact that the HVS is not used to the visualization of 3D content with the recent technologies, so a period to familiarize with this systems may be necessary, which is not noticeable with conventional television, because it is used since the childhood.

Moreover, the accommodation-vergence conflict, which will be described in detail in subsection 2.4.4, is caused by some 3D visualization technologies that evoke the depth perception showing one image to each eye, simulating the binocular disparity cue. However, in the natural stereoscopic perception of the real world, the convergence and the accommodation distance coincide in the observed object, while watching 3D content in these displays, the convergence point is located where the objects are virtually projected, while the accommodation mechanism establishes the focal distance in the screen (where the objects are really displayed). Thus, the unnatural difference between both mechanisms when using 3D visualization systems entails internal processes to compensate this effect that may cause visual discomfort.

Naturalness and sense of presence Finally, as depicted in Figure 2.1, other higher level factors are involved in the overall 3D visual experience, which are:

- Immersiveness or sense of presence: Is the subjective experience of being in the place or environment represented by the 3D video [92]. In fact, depth perception could help the viewers to be more engaged with the virtual world represented by the content, since it seems to be more realistic and interactive than conventional video [137].
- Naturalness: It refers to the fidelity of the representation of the environment with comparison with the reality, also known as perceptual realism. When a stereoscopic image faithfully reproduce the captured scene, it is called orthoscopic

image. Actually, some effects may cause that a 3D image does not look realistic (e.g., excessive depth, size of the objects, etc.), which can be provoked by the processes of content creation, coding and visualization [165].

2.4. System factors

All the elements composing the 3D video processing chain, depicted in Figure 2.7, can introduce effects that influence the QoE of the viewers. Therefore, it is important to understand the technological principles of each process and the main perceptual effects that may appear when creating, encoding, transmitting, and visualizing 3D video content. Thus, the following subsections present these aspects for each one of the elements of the video processing chain, setting a basis for the studies presented in the following chapters of this thesis.

2.4.1. Content production

2.4.1.1. Content production techniques

Content production includes the acquisition of 3D video sequences and processing mechanisms that could be applied to the video signal prior to the subsequent steps in the processing chain (e.g., coding, storage, visualization, etc.). Thus, the main approaches for capturing 3D video content are described in the following, together with the main processing techniques that could be applied to the 3D video signal.

Capture The most common alternative to capture 3D video is using stereoscopic cameras. This type of cameras can be composed by two lenses integrated in a single device or by two cameras installed in a platform with a certain horizontal separation or in a mirror setting, as depicted in Figure 2.8. This way it is possible to directly obtain two video streams corresponding to each stereo view. These camera systems could be configured depending on the field of view of the cameras, the distance between the cameras, and the convergence of the optical axes [223]. In fact, convergent (toed-in) or parallel configurations of cameras could be used, as depicted in Figure 2.9. It is worth noting that, since using convergent camera configurations may introduce geometric distortions (as described in the following subsection), the possibility of adjusting the convergence distance in parallel camera configurations (so objects could be visualized

Figure 2.7: Video processing chain

Figure 2.8: Examples of 3D cameras **Figure 2.9:** Diagrams of camera configurations

in front of the screen, with positive disparities) has been proposed by Fehn and Pastoor [44] with the horizontal translation of the sensors of the stereoscopic cameras, simulating a toed-in setting.

Furthermore, it is critically important to not have mismatches in the settings of the cameras composing the capture system, so precise color and geometry calibrations, and accurate synchronization are crucial issues [50].

An extension of the aforementioned systems is based on employing multiple cameras to capture the scene from various viewpoints, as depicted in Figure 2.10. This way it is possible to obtain multiview sequences (described in detail in subsection 2.4.2),

Figure 2.10: Multiview camera systems [192]

Figure 2.11: Principles of depth cameras [38]

which may provide a more complete visual experience to viewers. Nevertheless, the calibration, rectification, and synchronization of the cameras turns very complicated the capture process.

Another possibility to obtain 3D video is by means of a conventional camera together with a device to capture depth information of the scene, creating content for depth-based representations (described in detail in subsection 2.4.2), such as depth maps. Mainly, there are two types of systems able to capture depth information from the scene: Time-of-Flight (ToF) cameras and structured-light sensors, depicted in Figure 2.11. The former type of systems, estimates the distance between the sensor and the objects in the scene using light pulses and analyzing their phase information, while the latter type employs a projector to illuminate objects with special light patterns that allow the reconstruction of 3D shapes [221]. Furthermore, this type of cameras could be used with multiview capture systems, acquiring a more complete 3D representation of the scene, (generating Multi-view plus Depth (MVD) content, as described in subsection 2.4.2).

Finally, it is also worth mentioning the possibility of obtaining 3D content using computer graphics. Although the benefits of having natural content could not be provided, this alternative allows the generation of 3D models of the scene that could be visualized from as many viewpoints as the computational capacity of the computer permits.

Processing Various techniques has been proposed to process the 3D video signal after the acquisition or even previously to the visualization. Apart from the techniques proposed for 2D video (e.g., edge enhancement, noise filtering, etc.), some other methods have been specially developed for improving the performance of 3D video content, such as color adjustment to correct mismatches, or controlled blurring of background elements to enhance the depth perception. Moreover, together with these methods, the main processing techniques are:

- 2D-to-3D conversion: Various alternatives have been developed for converting 2D video to 3D, from simple approaches based on duplicating the frames and cropping and displacing them to generate two stereo views [230], to more complex proposals exploiting the information from monocular depth cues, such as motion parallax or vanishing points [221]. For example, motion parallax could be used

to generate depth maps considering that elements of the scene with different motions could be in different depth locations (e.g., foreground objects move faster than background elements) [148]. However, in general, these algorithms present limited performance, so semi-automatic algorithms have been proposed [123]. In addition, the best performance of these approaches are shown when, although in 2D, the content is captured following some rules with the idea of creating at the end 3D content.

- Depth adjustments: To obtain a comfortable and high-quality 3D perception, it may be necessary to apply depth modifications, for example to correct excessive disparities or to adapt the shooting configuration to the visualization conditions. For example, when 3D movies are created for cinemas, depth adaptations may be necessary to be correctly viewed in TV consumer devices. These treatments are based on rescaling the video, taking into account depth range, display size, and viewing distance. Therefore, since a simple linear scaling may introduce distortions, other methods are required, such as stereo retargeting. Other techniques allow the possibility to modify the convergence distance in parallel camera configurations (originally at infinite), thus changing the disparity of the scene, using horizontal image translation techniques [221].

2.4.1.2. Main perceptual effects

While the processing techniques aim at improving the performance of 3D video content, several degradations could appear caused by the capture process. In fact, acquiring 3D video is an extremely complex process, since, apart from the typical degradations that may appear due to the capture devices (e.g., geometrical distortions caused by optics issues, blur, noise, chromatic aberrations, etc.), other factors of the HVS related to stereoscopic perception should be taken into account to obtain high-quality and comfortable 3D content. For instance, issues like excessive disparities should be avoided and the comfortable viewing region (described in section 2.3) must be respected, taking into account the distance between the cameras. Therefore, specific perceptual effects may appear when capturing 3D video, such as [12]:

- Incorrect parallaxes: These effects appear when convergent camera configurations are used in the capture process and the convergence distances are different for each camera. This fact causes an incorrect parallax, both horizontal (depth plane curvature) and vertical (keystone distortion), which entail the appearance of artifacts specially perceptible in the corners of the elements of the image.
- Cardboard effect: One of the most common degradations related to the capture of depth maps is due to a limited acquisition of the depth information without enough continuity, so the scene seems to be divided in discrete depth planes. This fact causes that the objects in the scene are perceived unnaturally flat. Also, this effect may appear when using stereoscopic cameras with incorrect configurations of the acquisition parameters, such as convergence distance and focal length. In addition, it could also be caused by a coarse quantization of depth values, as mentioned in the following subsection 2.4.2.

Figure 2.12: Incorrect parallaxes [12]

Figure 2.13: Cardboard effect in the right image in comparison to the original (left) [12]

- **Color mismatch:** It is generally caused by a discordance in the color acquisition of the cameras used in the capture system (e.g., due to a different white balance), so each of the views present distinct color tonalities, which may cause severe degradations of the viewers' QoE.
- **Puppet-theater effect:** Employing convergent camera configurations may entail that the size of the objects in the scene and the distance to them do not correspond with the normal perception in the real world. Thus, for example, when people appear in the scene, they seem to be puppets.
- **Depth mismatch:** Apart from possible cue conflicts that may provoke depth mismatch, inserting added contents in post-production, such as subtitles, may cause depth conflicts [221]. In this sense, some factors, like the depth range or shooting settings, should be taken into account when adding these insertions in order to respect the depth perception of the original content. If not, incoherent issues may appear, like locating subtitles in a depth plane that should be occluded by other objects of the scene [109].

2.4.2. Coding

2.4.2.1. Representation and coding of 3D video

The encoding process of 3D video depends on the way it has been generated (e.g., by stereoscopic cameras, acquiring depth maps, etc.). Therefore, different representation formats for 3D video have been defined, and various alternatives to encode them have been proposed. Thus, the main options for 3D video representation and coding are described in the following.

Representation formats The main representation formats for 3D video are:

- Full-resolution stereo format: In this format, both stereo views have Full-HD resolution, so no spatial or temporal resolution reductions are applied, thus, the raw data is the double in comparison with a conventional monoscopic video [203].
- Frame-compatible representations: These stereo formats were defined in order to take advantage of the existing infrastructure and equipments for conventional video, with the aim of ease the introduction of 3D video technologies in the consumer market. Thus, in a frame-compatible scenario (as described by Digital Video Broadcasting (DVB) [40]), the stereo views are multiplexed into a single conventional frame or sequence of frames. Although the most common frame-compatible representation is SbS, there are various alternatives to achieve this type of representation [32] (see Figure 2.14):
 - Temporal multiplexing: The 3D video is composed of a sequence of alternate frames from each view, so the resulting framerate is doubled, and it should be reduced to equalize the amount of data of 2D video. In contrast, the spatial resolution is maintained. This alternative is called frame sequential representation (see Figure 2.14a), but field sequential formats could also be similarly obtained [203].
 - Spatial multiplexing. The frames from each camera are sub-sampled and combined into a single conventional frame, using different techniques:
 - Side-by-side (SbS): Both input frames are downsized to half their original width, and joined one next to the other (see Figure 2.14b).
 - Top-and-bottom (TaB): The vertical resolution of both input frames is halved, and joined one above the other (see Figure 2.14c).
 - Line-by-line: The output frame is composed interlacing both input frames, and half of the lines of each view are discarded to respect the size of a conventional frame (see Figure 2.14d).
 - Checkerboard: The stereo views are combined following a checkerboard pattern to compose the output frame. To obtain the size of a conventional frame, in this case, a quincunx sampling is applied (see Figure 2.14e).

Figure 2.14: Frame-compatible formats

- **Multiview representation:** It is similar to the aforementioned full-resolution stereo format, but in this case more views are captured from different viewpoints. Therefore, having a 3D video with N views, there is an N -fold increase in the raw data in comparison with 2D video [203].
- **Depth-based representations:** In these formats the 3D video is composed, not only by the views (texture), but also by information about the distance of the objects in the scene. This information could be achieved from algorithms that analyze stereo correspondences [160] or directly from special cameras (e.g., infrared cameras, time-of-flight cameras, etc.) [221], as explained in subsection 2.4.1. Generally, it is represented in depth maps, which are frames or sequences of frames with a depth value per pixel, that, usually, are quantized to form monochromatic images. In addition, these depth-based formats allow the generation of virtual views, using Depth-Image Based Rendering (DBIR) techniques. Furthermore, these formats present the advantage to be applicable to both stereo and multiview displays, and allow to easily perform depth adjustments according to the viewing conditions [203]. The typical representations are referred as:
 - **Video plus Depth (V+D):** It is composed of a single texture video and a single depth map, as depicted in Figure 2.15.
 - **MVD:** As depicted in Figure 2.16, it is composed of a number of views of texture and depth data [68], which can be represented by depth maps or Layered Depth Images (LDI) [167]. This latter technique joins depth information from different cameras into one single multi-layered representation, so each pixel of the LDI contains information about the visible and hidden pixels, which facilitate the view synthesis of algorithms handling occlusions.

Figure 2.15: V+D representation

Figure 2.16: MVD representation of three views [68]

Coding techniques The main approaches for 3D video coding are:

- 2D video codecs with signaling:
 - Simulcast: It is the instinctive way to compress 3D video, consisting of encoding each component (texture views or depth maps) independently using a conventional encoder, such as H.264/AVC. Although it presents the advantages of low complexity and latency (since no redundancies are exploited) and the compatibility with existing schemes (with simple signaling mechanisms to properly render 3D content), the main drawback is the poor coding efficiency that is achieved, entailing high bandwidth requirements [203].
 - Frame-compatible: This approach presents the advantage, like simulcast, of being compatible with existing video infrastructures, but presents less bandwidth requirements since frame-compatible video formats are used (i.e., with temporal or spatial resolution reductions). In order to decode and render correctly the 3D content, additional signaling is needed to identify the views and the interleaving scheme that has been applied. In the H.264/AVC standard [87], Supplemental Enhancement Information (SEI) has been added for this purpose, so it is required that the decoder supports the use of SEIs in order to undo the multiplexing of the views [32]. This approach has been standardized by DVB [40].
 - Scalable Video Coding (SVC): This is an extension of the H.264/AVC standard [87] that enables the encoding of a high-quality video bitstream that also contains a number of sub-streams. So, the typical use of this approach is to have a base stream with a version of the video that could be easily transmitted over limited network conditions, and a number of enhancement layers that are decoded when the network allows it to complement the base layer, obtaining a higher global quality, using temporal, resolution or coding quality scalability [161]. However, it could be also used for transmitting 3D

Figure 2.17: Typical MVC picture coding structure

video, for instance using a frame-compatible format as base layer and scale it to full resolution when possible, using spatial or temporal scalability in the enhancement layer [203].

- Free-Viewpoint Television (FTV) is a visual media that enables the observers to view a 3D scene by freely changing the viewpoint, offering the viewers the closest experience to be in the real scene [192]. Thus, ideally, FTV provides an infinite number of views of the scene, since the viewpoint can be placed anywhere. With this aim, Moving Pictures Experts Group (MPEG) started the research and standardization of FTV technologies in 2001, when the 3D Audio Visual (3DAV) activity began with the exploration experiments related to omni-directional video, FTV, S3D video, and depth coding [68]. Until now, three phases constitute the FTV activity:

1. Multiview Video Coding (MVC): The standardization activities in relation to MVC started in 2004 and finished in 2009 with an amendment to the H.264/AVC standard [204]. The main development of MVC was to add to the temporal redundancy, the possibility to exploit redundancy between views (i.e., inter-view redundancy), as depicted in Figure 2.17, to improve the coding efficiency of the simulcast approach providing a higher quality than the frame-compatible alternatives. Another important property of MVC to allow 3D viewing, was the mandatory inclusion in the stream of a base view compatible with H.264/AVC, which has the drawback of limiting the compression efficiency, since the inter-view compression tools were restricted. However, a latter revision included the stereo high profile (adopted by the Blu-ray 3D standard), in an attempt to customize MVC tools targeting Full-HD stereo video.
2. 3D Video Coding (3DV): The second phase of the FTV activity started in 2007, with the main feature of adding view synthesis to the framework, so, unlike in MVC, the number of output views could be larger than the number of input views. Although this feature was very limited in this phase, since only view synthesis in the horizontal direction was considered [193], 3DV provided a new data format for both stereoscopic and multiview systems: MVD. Four coding standards are included under this activity:
 - a) MVC plus Depth (MVC+D): It denotes depth-extension of MVC to encode MVD, so texture and depth videos are encoded into a single bitstream. Concretely, depth view components are coded and decoded similarly to texture view components and they are independent (i.e.,

using the MVC process), so it is possible to decode the texture video without having the depth components, and vice versa. However, this also means that possible redundancies between texture and depth components are not exploited. It is worth noting that this standard includes reduced-resolution depth coding (i.e., the depth video is downsampled to reduce the amount of data to encode due to the less importance of depth maps against texture). The standardization was technically finished in 2013 [22].

- b) 3D Advanced Video Coding (3D-AVC): The call for proposals on AVC-based coding of MVD data was released in 2011 [77] and posterior subjective evaluation campaigns were organized to find the best performing algorithm [78]. One of the main properties of the resulting standard is that, unlike MVC+D, the encoding of texture and depth is closely related, since it exploits redundancy between both types of videos. In addition, it included tools such as non-linear depth representation to improve the accuracy of depth maps and view synthesis prediction to enhance texture coding [68].
 - c) Multiview High Efficiency Video Coding (MV-HEVC): It is an MVC extension of the new HEVC standard [183], for which the same design principles than for the AVC-based MVC were followed. Thus, it allows encoding multiple views with inter-layer prediction [183]. It was included in the second version of HEVC, which was finalized in October 2014.
 - d) 3D-HEVC: It consists in encoding MVD based on the standard HEVC, so it outperforms the 3D-AVC. As shown in the coding structure depicted in Figure 2.18, the texture and the depth map are encoded for each view, including tools like disparity-compensated prediction and depth-based block partitioning [68].
3. Third phase: It started in August 2013 and looks for the introduction of SMV (i.e., more than 80 views) and free navigation applications, improving the limited view synthesis process, and including arbitrary camera arrangements (e.g., arc and linear configurations, see Figure 2.10) [68].

Apart from these approaches, an important technique that could be employed (especially in simulcast, MVC, and SVC scenarios) to reduce the global bitrate required to transmit 3D video is to use asymmetric coding, in which one of the views is encoded with less bitrate than the other [203]. Different techniques could be used to save bitrate in one of the views. For example, using a coarse quantization in one of the views, which usually leads to perceive the average image quality of both stereo views [165]. Another option is to downscale one of the views prior to the encoding process, since in case of blurring, the global 3D quality perceived by the viewers use to be the maximum between the quality of both stereo views [176]. Nevertheless, this mechanism also presents some limitations, since it may provoke large differences between both stereo views, hindering the 3D perception by the HVS [112].

Figure 2.18: Coding structure of 3D-HEVC [68]

Figure 2.19: Blockiness effect [12]

2.4.2.2. Main perceptual effects

The compression of the data when video content is encoded could cause artifacts that may influence the viewers' QoE. Among other distortions, such as color bleeding, flickering, or ringing [12], the main degradations that typical encoders for monoscopic video introduce are blurring and blocking effects.

In particular, blurring, or edge-shimmer, has been reported by consumers of video broadcasting services as one of the most annoying artifacts related to video coding [17]. Specifically, it is caused by the suppression of high-frequency Discrete Cosine Transform (DCT) coefficients that could be used to reduce the bitrate. In essence, it is perceived as a loss of the spatial details and sharpness of the edges of the elements in the images.

On the other hand, blocking effects are also due to the quantization process, since the DCT of the macroblocks that compose the frames are independently quantized. Thus, when a coarse quantization is applied, blocking patterns appear in the images introducing perceptible discontinuities on the boundaries between the macroblocks, as depicted in Figure 2.19. It is worth noting that some techniques have been developed and implemented in standard encoders, as the deblocking filter of the H.264/AVC encoder, to reduce the impact of these effects.

Generally, since the techniques proposed for 3D video coding, described before, are

based on previous coding standards for 2D video (e.g., H.264/AVC) and, thus, on the use of the DCT, the aforementioned degradations also appear in encoded 3D video. Nevertheless, their impact on the viewers' QoE may be different, since not only the image quality of the sequence could be distorted, but also they could affect the observers' 3D perception.

For example, the appearance of severe blocking or blurring effects in the stereo views of 3D video caused by a coarse coding, may lead to differences between views (i.e., mismatches) that could entail binocular rivalry [221]. However, the HVS could tolerate differences between both stereo views to a certain degree, based on the binocular suppression mechanism [117]. In fact, the aforementioned asymmetric coding techniques for 3D video coding are based on this theory. Thus, there are various studies analyzing the perceptual limits of the HVS to fuse mismatched stereo views, which seem to be related to the correspondence between low and high frequency components of both views [141]. However, some studies have shown that using a coarse asymmetric coding could cause distorted depth perceptions, as well as watching this type of content for long periods may cause visual discomfort [165].

Furthermore, apart from these artifacts related to the encoding of texture views, there are some distortions that appear when encoding the depth maps of 3D video sequences. For example, cardboard effects (that also could appear in the capture process as seen in subsection 2.4.1.2) may appear when a coarse quantization of the depth values is used [12]. This effect may affect the perception of depth, since the continuity of the depth range is reduced into a discrete set of depth planes, resulting in a unnaturally flat stereoscopic content.

Another distortion that may appear due to a harsh quantization of the depth maps is called “depth bleeding” or “depth ringing” (since it is similar to the ringing effects that could appear in monoscopic video) [12]. It consists of the appearance of artifacts in the borders of the elements of the scene due to misalignments between the contours of the texture video and the depth maps.

2.4.3. Transmission

2.4.3.1. 3DTV transmission scenarios

The delivery of 3D content over traditional video distribution systems (e.g., satellite, terrestrial, or cable television) to households should face some constraints, such as bandwidth limitations, backward compatibility, and receiver requirements [203]. Furthermore, other transmission scenarios over packet-based networks that are gaining popularity in the last years are being considered, such as IPTV [32], Peer-to-Peer (P2P) [54] applications, or Over-the-Top (OTT) services. One of the main advantages of these approaches is the facility to access those services from many different scenarios than the traditional domestic environment, thanks to the possibility of using other devices, in addition to typical television sets, such as personal computers, laptops, tablets, or mobile phones.

Figure 2.20: Diagram of HAS operation

In this type of networks the development of traffic management tools has been a crucial aspect due to the varying transmission conditions and scenarios of the distribution systems. These techniques are aimed at obtaining a flexible and efficient data transmission to cope with adverse situations in the network. In fact, when video services are delivered, this is especially a critical aspect to guarantee a satisfactory visual experience to the end users, whose quality expectations are more and more demanding. Therefore, some techniques have been proposed to face those events that reduce the network performances, trying to keep the quality level as high as possible. For instance, packet prioritization models have been developed to cope with network losses, taking into account that video packets have information that could influence differently the users' QoE [139].

Other examples are the techniques applied when the QoS of the network drops, and the amount of data to be transmitted should be reduced. For instance, the strategy called *denting*, which is based on dropping some frames of the video sequence, has been already applied in IPTV scenarios [36].

Another example is transmitting lower-quality versions of the video content [178], which is also applicable in cases in which the same content should be delivered to devices with different technologies (e.g., mobile devices that do not provide Full-HD resolution as TV sets do, so it is not necessary to transmit the same information to both systems). This latter approach is the basis of HTTP Adaptive Streaming (HAS), which is currently becoming a very popular alternative for video distribution to households and has been already standardized by MPEG [131]. The operation of HAS is depicted in Figure 2.20, and is based on the generation of various segmented versions of the same content with different bitrates. So, according to the network and processing conditions, the client asks the server for the appropriate version that could be correctly received. The switching from one version to another should be done in pre-fixed points of the video in order to maintain the continuity of the sequence, and in a transparent way to the user. Therefore, the copies of the video are divided in *chunks*, which can be independently decoded.

Although the research on adaptive streaming mechanisms for conventional video is still an open issue, the arrival of 3DTV services to households has motivated the investigation on similar methods to control the efficient transmission of 3D content. For example, when the network could not withstand the data load needed for 3DTV, it is

Compatibility level ↓	Conventional HD Service Compatible (CSC) Level 4	2D HD + MVC⁽¹⁾ (i.e. MVC) (L, R formed by matrixing)	2D HD + MVC⁽²⁾ (i.e. MVC) (Depth, occlusion, transparency data)	
	HD Frame-Compatible Compatible (FCC) Level 3	Frame compatible plus MPEG resolution extension, for example SVC ⁽³⁾		
	Conventional HD Frame Compatible (CFC) Level 2	Frame compatible (L, R in same HD frame)		
	Conventional HD Display Compatible (CDC) Level 1	Optimized colour anaglyph		
		Plano-stereoscopic profile 1st generation 3DTV	Multiview profile 2nd generation 3DTV	Object wave profile 3rd generation 3DTV
		Generation profile →		

Figure 2.21: 3DTV signal formats [93]

possible to switch to the 2D version of the video, reducing the bitrate to transmit [6]. Also, the concept of asymmetric coding could be exploited [55].

To overcome the aforementioned constraints and bring 3D content to households, the ongoing research on the new formats and coding technologies previously mentioned (see subsection 2.4.2) should be followed by a technological development regarding the transmission systems and the reception equipment.

In this sense, standardization efforts have been already carried out in order to guide this process avoiding a battle of formats, to ensure the development of valuable 3D video solutions for home entertainment. In this regard, ITU [93] and DVB [40] research groups have been the main actors, defining recommendations and specifications for a progressive establishment of 3D video in the consumer market, as depicted in Figure 2.21. Nowadays, according to the figure, 3DTV services are mainly in the first generation and second level, since 3DTV-capable displays are already available on the market, and the majority of 3D video channels offered to consumers are based on frame-compatible approaches. This means that service providers use the existing HDTV infrastructures to deliver S3D content, with minor modifications like, as aforementioned, additional signaling (e.g., SEI).

A further step would be the introduction of Frame-Compatible Compatible (FCC) services, which provide full resolution for both views, while keeping backward compatibility with current generation of decoding devices limited to frame compatible formats. A possible implementation is illustrated in the processing flowchart shown in Figure 2.22. For each frame of the stereo views, the even columns are taken to form one SbS frame

Figure 2.22: Flowchart of processing FCC format

(with half the original columns), and the odd columns are taken to form another SbS frame. Then two individual SbS frames (with half horizontal size) are obtained, so in the coding process (e.g., SVC), one is used in the base layer, and the other one in the enhancement layer. This way the frame compatible systems are able to obtain a valid SbS frame, and the new systems will be also able to get both stereo views in Full-HD.

Finally, the next level contemplates the provision of Full-HD resolution per stereo view and the support for multiview video signal and depth-enhanced scenarios. This would require the use of new Set-top Boxes (STB), while the arrival of multiview displays to the consumer market will trigger the second generation of 3DTV systems.

2.4.3.2. Main perceptual effects

As aforementioned, the same infrastructure than for 2D video is currently being used for delivering 3D content. Thus, the same errors may appear in 3D video services. However, their impact on the end users' QoE may not be the same, due to the additional perceptual factors that are involved in the stereoscopic vision in comparison to the conventional watching experience. In essence, the main artifacts caused by the transmission of video content that have been reported by the observers are [17]:

- *Macroblocking* effects: Delays (jitter) and losses on the reception of the video stream (e.g., packet losses in IPTV services) could make the decoder unable to rightly reconstruct the video frames, so error concealment algorithms are applied. One of the most common mechanisms, usually called *slicing*, is based on the substitution of the erroneous macroblocks by those that have been correctly decoded in previous frames. This mechanism generally entails the appearance of incoherent squares or blocks in regions of the image, as depicted in Figure 2.23. Moreover, this effect mainly affects zones of the images containing moving elements, and it could propagate to several frames (until a new I-frame is correctly received), due to the use of reference frames in the encoding process. Although these effects already cause severe degradations of the QoE in 2D video, their impact in 3D content may be even more annoying, since they could cause binocular rivalry when the same region in both stereo views is affected differently [57, 212]. This effect is deeply investigated in the following section 5.3.
- *Screen freezes*: A simpler error concealment strategy just substitutes the whole frame affected by transmission errors with the last frame correctly decoded. Thus,

Figure 2.23: Effect of video losses on SbS video

this causes that a still frame remains in the screen until the next I-frame is correctly received. Also, video freezes may be caused by *buffering* events in internet-based video services, such as HAS.

- **Choppy transitions:** Short consecutive video freezes or framerate reductions (e.g., applying adaptive streaming techniques based on dropping frames, like *denting*) may cause that the video is not smoothly visualized, which could degrade drastically the visual experience of the viewers.
- **Outages:** These events are interruptions of the whole service for a relevant period of time, which may be caused by equipment failure (either in the customer or the network sides), but also by abrupt and severe losses of the video and audio signals.

Apart from these degradations, also audio distortions (or silences due to audio losses, called *glitches*) have been reported by the consumers to affect their QoE, together with coding quality degradations [17], that may appear in scenarios applying bitrate adaptations (e.g., HAS), which have been already discussed in the previous subsection 2.4.2.2. Finally, another important factor in video transmission services is the end-to-end latency, which is crucial in real-time communications (e.g., videoconferencing) or live events, and mainly takes place in internet-based services, such as P2P.

2.4.4. Display

2.4.4.1. 3D display technologies

Since the presentation of the first stereoscope by Charles Wheatstone in 1838 [218], a great development on technologies for visualizing 3D content has taken place. Espe-

cially important is the progress achieved in the last years thanks to decisive technological and hardware advances. This development, which is still ongoing [170], has allowed the apparition of various alternatives to watch 3D content, even in the consumer market.

A possible classification of 3D displays is based on their degree of parallax, having two-view, horizontal-parallax, and full-parallax displays [69]. The first is the simplest type of 3D displays, which provides one view for the left eye and one for the right eye of the observers to allow stereoscopic perception. On the other hand, horizontal-parallax displays provide simultaneously multiple views of the same scene, so the observers could see different perspectives changing horizontally their position within some defined regions. Finally, full-parallax displays provide variations in what is being perceived by the observers with both horizontal and vertical head movements.

Since various extensive reviews have been published covering in depth the evolution of 3D display technology [9, 48], only some the most popular and common 3D display systems are here described.

Two-view systems These displays are based on providing two different views, one for the left eye and one for the right eye, and different types have been developed depending on how these images are generated. Thus, these displays can be classified in four groups: location multiplexing, wavelength multiplexing, polarization multiplexing, and time multiplexing [69, 129].

- **Location multiplexing:** This technique is based on redirecting the left and right views to the corresponding eye through different optical channels, for example using two separate lenses or two different images. The first device developed using this approach was the stereoscope invented in 1838 by C. Wheatstone. As depicted in Figure 2.24a, it was based on the use of two mirrors oriented at 45 degrees (A and A' in the figure) with respect to two images (one for the left eye and one for the right eye) located in both sides of the device (E and E') [218]. An improved stereoscope was proposed by D. Brewster in 1849 [13], consisting on the use of two lenses, instead of mirrors, to provide the corresponding image to each eye, as depicted in Figure 2.24b. This primitive design is considered the basis of the development of current Head-Mounted Displays (HMD) (see Figure 2.25) [69], which use two micro-displays and optical components, such as waveguides [118] or holographic elements [132], to provide to each eye the corresponding stereo view [9].
- **Wavelength multiplexing:** This approach is based on using two different color channels to provide each view to the appropriate eye. In essence, the two views are filtered through two different color channels and the observers use glasses with the respective color filters in each eye (see Figure 2.26). Although near-complementary colors are commonly used (e.g., red and cyan), color rivalry and crosstalk effects appear and degrade visual experience of the viewers. Nevertheless, it is one of the earliest stereoscopic systems and has been widely used thanks to its simplicity and low cost. Also, more recent systems have been developed from this basic approach [98].

Figure 2.24: Stereoscopes

Figure 2.25: Prototype of a HMD [132]

- **Polarization multiplexing:** These two-view systems present simultaneously the left and right views using two projectors or displays with polarizing filters, so the observers should wear appropriate polarized glasses to make possible that each eye sees the corresponding stereo view, as depicted in Figure 2.27b. Although there are approaches for time-parallel presentation of both stereo views (e.g., at cinemas using two projectors), the most common technique in the consumer market is to interleave both views line-by-line using a filter film located in the screen that polarizes the horizontal lines with alternate polarization. Obviously, this approach entails a reduction of the image resolution (e.g., the vertical resolution is reduced to the half in line-by-line interleaving) [69].
- **Time multiplexing:** These two-view systems present the stereo views to the viewers alternatively, so the two stereo images do not coexist in the screen simultaneously as in the polarization-multiplexing approaches. Thus, it is necessary to use active shutter glasses (blocking alternately the light to each eye), synchronized with the display (usually by infrared systems), to see each time the corresponding

Figure 2.26: Anaglyph glasses

Figure 2.27: Schemes for time and polarization multiplexing displays [48]

view in each eye. These systems take advantage of the capability of the HVS to fuse the stereo views with a time-lag of up to 50 ms [129]. Although the spatial resolution of the images is conserved, it is necessary to, at least, double the framerate of the displays to reduce annoying flickering effects. Although this is one of the most common 3D display technologies in the consumer market, there are other time-multiplexing approaches (in the sense that both stereo views are not displayed simultaneously). For example, in cinemas it is possible to use only one projector that emit the left and right views sequentially with different polarizations, with the advantages that the observers can use a simple lightweight pair of polarized glasses and that there is no resolution loss in comparison with aforementioned polarization multiplexing [69].

Horizontal-parallax systems This type of displays do not require any kind of glasses to watch 3D content, since each view is sent to different spatial regions in front of the display. Thus, when the observer is located at the correct position, each eye receives the corresponding view. The main technologies used are lenticular lenses and parallax barriers, schematized in Figure 2.28. The first type is based on the use of adjacent vertical lenses located in front of the display limiting the visibility of each column of pixels to a specific zone. Therefore, the resolution of the display is divided by the number of zones (i.e., views). On the other hand, parallax barrier displays use vertical slits located in front of the screen, creating contiguous opaque and translucent regions and generating a set of viewing zones where the observers could experience stereoscopic perception. Also, in this type of displays, the resolution is divided by the number of views [69].

In essence, conventional multiview 3D displays have three important drawbacks. Firstly, the effective image resolution is divided by the number of views provided by the display. Secondly, there is always a discontinuity in view switching with respect to the viewing direction. Finally, they suffer from the accommodation-vergence conflict, which may cause visual discomfort [48]. To overcome these problems, researchers are working on the development of SMV displays.

Figure 2.28: Schemes of autostereoscopic displays

Figure 2.29: Super-multiview condition [187]

SMV displays are horizontal-parallax systems and, as any multiview visualization system, they generate a set of discrete views and distribute them over the viewing field in front of the screen, offering a series of viewing zones where the observers can perceive different perspectives of the scene as they move their heads [153]. These displays provide sufficient views for continuous horizontal parallax and for a comfortable viewing experience, since the distance between two adjacent views becomes smaller than the pupil size, as depicted in Figure 2.29, satisfying the super-multiview condition [187] (described in detail in the following subsection 2.4.4.2). To achieve this requirement, SMV displays require a large number of views, typically 80 or more [127]. In addition, to obtain a satisfactory image resolution, different technologies are being investigated, such as multi-projection systems or flat-panel systems [187].

Full-parallax systems Full-parallax systems provide both horizontal and vertical parallax, allowing a complete 3D viewing experience of the scene from any point of view. One of the main technologies that are being investigated are volumetric displays. These systems generate imagery that occupies a volume in space instead of a surface, composed by voxels instead of pixels. This is achieved by means of light-emitting, light-scattering, and light-relaying techniques [69]. Two types of volumetric systems are: multiplanar and solid-state displays. The former type of displays are based on integrating several 2D surfaces (like slices) into a 3D volume, thanks to the persistence of vision [182]. The latter type is based on the use of laser beams and mirror scanners to

generate solid state processes, such as two-step upconversion, allowing the generation of volumetric imagery [135].

Another approach to full-parallax 3D displays are integral imaging systems [95]. This type of displays could be considered as an extension of lenticular autostereoscopic displays, since, instead of using cylindrical lenses in front of the screen, they use semi-spherical lenses, so the light from the pixels is directed in different horizontal and vertical directions, generating horizontal and vertical parallax. Currently, these displays have many drawbacks to face, which make their cost not worthy for their performance and their limited added value in comparison with lenticular horizontal-parallax displays. For example, integral imaging displays suffer a greater resolution loss to provide also vertical parallax. In addition, these displays have limited viewing angle, depth of field, and quality provided by the costly lenses [69].

Finally, it is also worth noting the development regarding holographic systems for 3D visualization. This type of displays have been considered the ideal solution for a glasses-free 3D viewing experience, since they provide all depth cues to observers (although there are also horizontal-parallax approaches [69]). However, these systems still present some critical drawbacks that are focus of research. For example, holograms contain a huge amount of data that should be compressed and coherent light should be used for recording them [129]. Holographic displays provide more continuous parallax than multiview displays, since they use laser projectors and light modulators [175] or a screen of holographic optical elements [4], so they do not render a limited number of views.

2.4.4.2. Main perceptual effects

Apart from typical effects that appear in conventional displays, several artifacts that could degrade the viewers' QoE are caused by technical limitations of visualization systems related to 3D perception.

Crosstalk This is one of the most annoying effects produced by 3D displays, since it diminishes the viewers' 3D perception and could produce visual discomfort [137]. It could be observed, both in autostereoscopic and stereoscopic displays, as ghosting artifacts and double contours, caused when an image is shown to the corresponding eye and the image corresponding to the other eye is not completely isolated, so a residual image can be perceived. Many technological issues related to the displays could cause crosstalk. For instance, in autostereoscopic displays, crosstalk could appear due to incorrect viewing positions of the observers. On the other hand, in stereoscopic displays with polarized glasses, crosstalk could be caused by deficient filters or incorrect head position of the observers (when linear polarization is used), while in time-sequential displays it could appear due to the persistence of images in the screen or deficient occlusions of the shutter glasses [129]. Moreover, crosstalk is influenced by contrast, disparity, and screen parallax, which could limit the amount of depth of the sequences to provide a high QoE to the viewers [166].

Frame effect Also known as window violation, it appears when objects in the scene are represented in front of the screen (i.e., with crossed disparity) and they approach the screen frame. When the screen frame disrupts the objects, a conflict between two depth cues appear: the binocular disparity causing the perception of the object in front of the screen, and the information coming from the occlusion of the screen frame, which indicates that the object should be behind the screen [229]. In addition, when the objects approach the screen frame, it is possible that one eye do not see parts of the objects that the other eye sees. Thus, an incoherent 3D perception is caused, which considerably degrades the viewers' QoE. This effect should be avoided with a careful acquisition of the 3D content [130], although also the visualization in large screens (e.g., cinemas) may reduce its impact, since the observers are less conscious of the existence of the frame [81]. Moreover, some post-production techniques have been proposed to solve this issue, such as the change of the disparity of the content or adding floating windows, which are virtual frames positioned closer to the viewer than the content [229].

Accommodation-vergence conflict As described in the previous subsection, the majority of the 3D displays are based on showing the images addressed for each eye on the screen with a certain disparity between them. Depending on the technology, these images are separated so they are perceived by the corresponding eye and the HVS is able to fuse them and provide a stereoscopic perception of the scene. Therefore, the objects can be perceived at different depths, being represented in front of the screen (i.e., crossed disparity) or behind the screen (i.e., uncrossed disparity). However, unlike the natural stereoscopic perception in the real world, where the convergence and accommodation points coincide, in most of the 3D displays the accommodation mechanism establishes the focal distance in the screen, while the convergence point is fixed where the objects of the scene are virtually projected, as depicted in Figure 2.30. This abnormal difference between both elements, has been reported to cause visual discomfort [110]. It is also worth noting that the accommodation-vergence effect is affected by the diameter of the pupil, since it changes the focus. Therefore, brighter reproductions of the 3D content may reduce the visual discomfort [186].

Also, some 3D display technologies, as previously explained, are not based on the presentation of one stereo view to each eye, so the accommodation-vergence conflict does not appear, for example in holographic displays. In addition, as previously explained, SMV displays, which are currently under research, provide a view density that allows each pupil to capture more than one view. Thus, the super multiview condition is satisfied, which avoids the accommodation-vergence conflict permitting a more comfortable 3D visual experience. As depicted in Figure 2.31, since the pupil receives various views, the effective pupil diameter is reduced, which entails an enlargement of the depth of focus. Thus, although the accommodation focuses on the screen, the eyes converge on the projected objects, which are within the depth of field where the HVS allows a correct perception.

Figure 2.30: Scheme of the accommodation-vergence conflict [153]

Figure 2.31: Depth of field when viewing a conventional multiview display (above) and an SMV display (below) [187]

Other effects Stereoscopic displays also present some effects caused by the use of glasses to show each stereo view to the corresponding eye. For example, using active shutter glasses may produce disturbing flickering effects if the illumination of the room is not adequate. In addition, both active shutter glasses and passive polarized glasses reduce considerably the luminance emitted by the display, which may degrade the QoE of the viewers [19].

Other important issue is related to the loss of resolution that appears in multiview displays (since the total resolution of the screen is divided by the number of views) and stereoscopic displays with polarized glasses (because of the use of spatial multiplexing to show to each eye the corresponding view). These losses, could degrade the image quality perceived by the viewers due to the appearance of artifacts, like blurring effects.

Finally, it is also worth noting the possible appearance of other binocular artifacts, such

as shear distortions (the observers see the same images independently of their viewing position, entailing a wrong head parallax) [12], puppet-theater effect (as described in subsection 2.4.2.2, it is a size distortion characteristic of 3D images that especially appears with familiar objects, as human beings) [225], picket-fence effect (the observer can perceive brighter and darker vertical stripes in multiview displays due to a limited spatial multiplexing of the views) [12], and image flipping (when sharp transitions could be observed in multiview displays because the objects of the scene have been captured with a limited number of views) [12].

2.5. Evaluation of 3D QoE

Although the first approaches to evaluate the QoE of the users of 3D video technologies have been based on the methods and techniques developed for conventional video, the new factors that influence the overall 3D QoE in comparison with 2D video (e.g., depth perception, visual comfort, naturalness, immersiveness, viewing conditions, etc.) [19], make necessary to review the existing methods and develop new approaches to obtain reliable evaluations of the 3D visual experience. Thus, this section presents the advances in the alternatives for evaluating the 3D QoE, both by means of subjective assessment tests (based on the direct opinions of a number of observers about the system or application under study) or by objective metrics (which try to automatically estimate the quality that end users would perceive).

2.5.1. Subjective evaluation methodologies

Since audiovisual services are addressed to people and their visual experience is significantly influenced by subjective factors, the most accurate and reliable way to evaluate the QoE of the end users is by means of subjective assessment tests. In such experiments, the audiovisual content to be evaluated is shown to a number of observers who rate it, or expresses their opinions about it, according to a test methodology. One of the main advantages of these tests is the possibility of investigating mechanisms and responses of the HVS in the presence of visual stimuli. Thus, it is possible to extract useful outcomes to optimize different aspects of the audiovisual systems, such as the coding techniques for video content taking advantage of perceptual features of the HVS, like masking mechanisms. In addition, although the practical applications of these methods are limited, since they are highly time-consuming and require the participation of several observers, they are essential to calibrate and validate objective metrics developed to automatically estimate the QoE perceived by the end users.

The research work carried out during the last decades regarding the design of appropriate methodologies for subjective quality evaluation for conventional video has resulted in the standardization of some procedures to take into account depending on the application or system under study. In particular, the main recommendation for subjective tests with 2D video are: ITU-R BT.500 [89] stated for television pictures

Figure 2.32: Diagrams of presentation of sequences for DSIS methodology

Quality scale	Degradation scale	Comfort scale
5-Excellent	5-Imperceptible	5-Very comfortable
4-Good	4-Perceptible but not annoying	4-Comfortable
3-Fair	3-Slightly annoying	3-Slightly uncomfortable
2-Poor	2-Annoying	2-Uncomfortable
1-Bad	1-Very annoying	1-Very uncomfortable

Table 2.1: Five-grade quality, degradation, and comfort scales

and ITU-T P.910 [86] for multimedia applications, and the recent ITU-T P.913 for audiovisual quality in video distribution services in any environment [94]. In contrast, although some recommendations have been already published for 3D video, such as ITU-R BT.1438 [81] (directly based on the ITU-R BT.500 [89]) and ITU-R BT.2021 [92] (more complete considering specific factors of the 3D visual experience, like depth perception or visual comfort), the design of proper methodologies for subjective evaluation of 3D content is still under research, as it is reflected by the work carried out by the VQEG under the scope of the 3DTV project [205]. The main methodologies considered in these recommendations are:

- Double Stimulus Impairment Scale (DSIS) [89]: Also known as Degradation Category Rating (DCR) [86], is a double-stimulus method, so the observers watch both the reference sequences (i.e., undistorted) and the test sequences (containing certain effects under study). The most common alternative is to show to the participants firstly the reference sequence followed by the test sequence, and then they provide their evaluations, as depicted in Figure 2.32a. Another possibility is to show twice both the reference and the test sequences, as shown in Figure 2.32b [92]. Since this methodology is normally used to evaluate the robustness of a video system analyzing the visibility of certain degradations that it may introduce (e.g., transmission errors, coding artifacts, etc.), the degradation scale, as shown in Table 2.1, is used by the observers to provide their ratings. Usually, undistorted sequences are also used as test sequences to have reference ratings. It is worth noting that this method was used in the subjective evaluations of the 3DV standard [77].
- Double Stimulus Continuous Quality Scale (DSCQS) [89]: This method is very similar to the previous one, however, in this case the observer does not know whether the displayed sequences are the references or the test clips. The most common presentation pattern is shown in Figure 2.32b, although in some cases, if the test is carried out individually, the observer is allowed to see the sequences

Figure 2.33: Diagrams of presentation of ACR methodology

as many times as desired, within a time interval. Also, a quality scale is used to express the ratings, as the one shown in Table 2.1. This method is usually employed when the quality of the test sequences and the references is close, since it is very sensitive to small differences. This methodology has been also already applied in subjective tests with 3D videos, as in [158].

- **Absolute Category Rating (ACR)** [86]: It is a single-stimulus method (i.e., no explicit references are shown to compare with the test sequences) based on the presentation of the test sequences followed by an interval of time between them when the observers have to indicate their ratings, as depicted in Figure 2.33. Usually, during the evaluation period the screen shows a gray still image. Also, the scale used by the observers to provide their ratings is the five-grade quality scale shown in Table 2.1. It is also worth noting that it is possible to include undistorted sequences as test clips to allow a better analysis of the results, which is usually called ACR with hidden references. Moreover, this method has been used in various VQEG [206, 207] projects and in many tests with 3D video, as in [212].
- **Single Stimulus Continuous Quality Evaluation (SSCQE)** [89]: It is a single-stimulus method based on the continuous evaluation of a long test sequence (e.g., 20-30 minutes) using a slider bar, potentiometer, or a similar device. Thus, the observers evaluate the test sequence using a continuous quality scale. The continuous evaluation of the test sequence allows the detection of rapid quality changes, so this methodology is the most appropriate to study instantaneous quality fluctuations or degradations that vary during time. On the other hand, the devices employed for evaluating are complex and could cause undesired effects, such as that some observers may divert the attention to the sequences if they lose the reference position of the slider. It is worth mentioning that a variation of this method, called SSCQE with hidden reference removal, was used by the VQEG for the evaluation of the performance of objective metrics, based on the introduction of undistorted segments in the test sequences, acting as hidden references, which permit the equalization of the obtained results [145]. Also, this method has been already used with 3D video, as in [159] to evaluate different symmetric and asymmetric coding configurations.
- **Pair Comparison (PC)** [86]: This method is based on the comparison of sequences processed with different conditions, depending on the parameters or effects under study (e.g., different compression levels). In particular, the sequences are presented in pairs and the observers are asked to select the one they prefer, or to evaluate them according to a comparison scale as the one showed in Table 2.2. The presentation of the pairs could be done in time-sequential or time-parallel

-3: Much worse
-2: Worse
-1: Slightly worse
0: The same
+1: Slightly better
+2: Better
+3: Much better

Table 2.2: Comparison scale [89] **(a)** Time-sequential **(b)** Time-parallel**Figure 2.34:** Diagrams of presentations for pair comparison methodology

modes. On one side, in time-sequential presentation the pair of stimulus are presented on a single screen one after the other, as schematized in Figure 2.34a, and the observers indicate the evaluation after seeing both stimulus. On the other side, time-parallel presentation (also called side-by-side presentation) is based on the visualization of both stimuli in two identical screens simultaneously, as depicted in Figure 2.35. In this case, the synchronization of the two videos of each pair is crucial, as well as the use of two screens with the same model, size, and color and luminance calibrations. Although these technical issues make this alternative more complex, the time-parallel presentation has some advantages in comparison with the time-sequential, such as that the duration of the tests is reduced (as reflected in the time-pattern depicted in Figure 2.34b), and it is easier for the subjects to identify the differences between the two stimuli. It is worth noting that in both presentation modes, each pair should be presented twice in the two possible orders (e.g., stimuli A followed by stimuli B, and vice versa in

Figure 2.35: Position of two screens in time-parallel presentation

Figure 2.36: Rating scale used in SSMM

time-sequential presentation, and A-B and B-A in time-parallel presentation) to avoid possible effects of the presentation order and the screens. Moreover, this method has been also adopted by the VQEG for its 3DTV project [205].

It is worth noting that in some cases, when several test conditions are considered in the test, it is unfeasible to carry out all the possible comparisons between them during one assessment session, so some optimized designs have been proposed in which only a subset of the test conditions are compared. For example, the Optimized Rectangular Design (ORD) has been proposed by Li [120] and already used in 3D tests, which is based on the comparison of a subset of the test conditions according to a previous knowledge about their possible effects. In addition, although the results obtained with this methodology are usually based on how many observers preferred one stimulus against another, referred as Preference of Experience (PoE), there are statistical models, such as the Bradley-Terry (BT) model, that permit the conversion of these results to scores, similar to those obtained using rating scales [120]. This way it is possible to quantitatively compare different test conditions and, also, compare the results from PC method with other methodologies.

Other interesting approaches that, although not standardized, have been used in various subjective assessment tests are:

- Subjective Assessment Methodology for Video Quality (SAMVIQ) [11]: It was deigned from the DSCQS methodology to evaluate the performance of video encoders. It is a double-stimulus method that shows to the observers both an explicit and a hidden reference. It was originally developed to be used with a computer, so the evaluators can access to all the sequences and visualize them in the order that they prefer, and as may times as they want. Then, the observers provide their evaluations using a virtual slider and a quality scale. It is worth noting that one of the main advantages of this methodology is its reliability for evaluating low-quality sequences. This methodology has been also already used in 3D quality tests, as in [20].
- Single Stimulus Multimedia (SSMM) [134]: It is similar to the ACR methodology, so the presentation pattern is as depicted in Figure 2.33, but to reduce the context effect, each test sequence is showed twice during the same assessment session, thus evaluated twice by the observers. Then, the average score provided in both

evaluations is assigned to that test sequence. Usually, a quality scale with eleven values and five adjectives is used, as shown in Figure 2.36. This method has been used in the subjective evaluation of the SVC [75] and MVC [76] extensions of the coding standard H.264/AVC.

Furthermore, an interesting study was carried out by Pinson and Wolf [145] comparing various of the aforementioned methodologies, especially focusing on the differences between double-stimulus and single-stimulus approaches. In essence, one of the main distinctions refers to the less sensitivity of double-stimulus methods to the content, since showing an explicit reference may provide more precise and robust results against factors like the presentation order of the sequences. Nevertheless, the fact of showing a reference distances the observers from real-viewing conditions, which may not be appropriate when conclusions concerning the real usage of the services or applications under study are investigated. Another difference lies on the longer duration of the test sessions when double-stimulus schemes are used, since both the reference and the test sequences should be displayed. In addition, single-stimulus methods are more sensitive to context effect, which consists in the influence of the content previously watched on the ratings of the observers, and is influenced by short-term memory factors. Moreover, a similar study was carried by Kawano *et al.* [101] comparing the performance of ACR, DCR, and DSCQS methods with 3D video, concluding that the DCR method was the most stable, although, in terms of both stability and assessment time, the ACR method was the most effective.

It is worth noting that all the previous methodologies were conceived to be used in controlled environments to avoid the influence of external test conditions in the results, so the experiments could be repeatable in other moments and places. Thus, these methodologies specify various factors to set up the test environment, such as the illumination to be used in the room or the employment of gray curtains covering the walls to avoid reflections. However, these test conditions may distance the observers from real viewing situations, which may be not convenient for the evaluation of certain systems and applications, such as video distribution services, whose performance may be better to be evaluated in real conditions. In fact, the viewing environment is one of the context IFs which may affect the QoE [19]. In a similar vein, the selection of the test material for the experiments is crucial [143]. For instance, the aforementioned standard methodologies usually recommend the use of certain test videos that in many cases are very different to what is commonly watched in real viewing situations. For instance, the test content is normally more focused on testing some features of the applications under study than in being interesting to the observers, so short sequences (e.g., around 10 seconds) without audio are usually employed, which, in many cases, may not be appropriate to obtain representative results of realistic usage situations.

Therefore, a tendency has lately arose to research on methodologies considering realistic home-environment situations. In fact, this idea has been included in the recent recommendation for subjective testing ITU-T P.913 [94]. In essence, the objective of these approaches is to obtain more representative conclusions of the QoE of the end users of video systems addressed to the consumer market, since the participants are not only focused on detecting possible degradations, but they could be immersed in

the viewing experience.

For example, interesting approaches are based on the collection of the opinion of the end users of video delivery services by means of surveys. For instance, Cermack [17] used this alternative to investigate the degradations that are most commonly reported by the end users. Also, the acceptance of the consumer to those impairments was studied, taking even into account economic aspects, since end users' QoE is highly related to the price of the service [154]. In a similar manner, Krishnan and Sitaraman [106] investigated end-users behaviors with internet video, like engagement, abandonment, and repeated visualizations, using the traces acquired from a content delivery network.

Although the ideal way to study the QoE of the end users would be knowing their opinions in real-life usage, the control over some important factors, such as the observer, the content, or the environment are totally lost, so trade-off alternatives are still under research. For instance, Jumisko-Pyykkö and Hanuksela [99] carried out a subjective study related to video visualization in mobile devices considering three real-life contexts: a train station, a bus, and a cafeteria. Another approach was proposed by Staelens *et al.* [174] based on the evaluation of typical IPTV degradations in full-length movies by the participants directly in their homes. In addition, to stress the importance of carrying out the evaluations in real-life environments, the authors evaluated the same degradations with the ACR standard methodology, and compared the obtained results, concluding that some impairments were less perceptible in the real-usage case. Furthermore, Strohmeier *et al.* [181] conducted an experiment to evaluate the quality of 3D video in mobile devices considering both a controlled setting and a cafeteria, as a real-life environment. In addition, the Open Profiling Method was used, so the observers indicated their evaluations both using typical qualitative ratings and descriptions of their opinions using their own vocabulary. In essence, similar results were obtained in both test environments, although the observers seem to be more sensitive to crosstalk under the uncontrolled conditions of the cafeteria.

Another interesting method was proposed by Sullivan *et al.* [184] using realistic content in order to engage the observers and immerse them into a visual experience similar to what they encounter in real life. Thus, typical video content is shown to the participants and, apart to traditional quality ratings, distractor questions are posed to them, to focus their attention on the application under study. Although this approach is mainly focused on getting the observers closer to real-life conditions in relation to the content used in the test, the ideal usage of this methodology would be in test environments simulating realistic home-viewing conditions [144]. Therefore, in this sense, the *Content-Immersive Evaluation of Transmission Impairments* (CIETI) methodology, described in detail in this thesis in subsection 5.2, was proposed to evaluate typical effects that take place in video delivery systems trying to mimic, as far as possible, real viewing conditions, in terms of test content and environment.

Finally, it is also worth mentioning the recent activities for evaluating QoE using crowdsourcing, based on the execution of subjective tests by anonymous participants in their preferred environment [70]. Although this alternative brings important advantages, such as the use of realistic environments, the reduction of time and costs needed in comparison with traditional tests, and the possibility to expand the range of partici-

pants (e.g., geographically); some important problems also appear. For example, the control over the test settings is totally lost and the reliability of the user ratings could be dubious, due to cheating users. Thus, an active research work is currently going on to define robust methodologies for crowdsourcing testing.

To extract valuable conclusions from subjective tests, it is essential to statistically analyze the evaluations provided by the observers. Usually, since rating scales are employed, the Mean Opinion Score (MOS) is computed averaging the scores provided by the observers, as well as the standard deviation of the scores, and the Confidence Intervals (CI), which help to investigate the statistical significance of the results. Also, screenings of the results should be carried out to detect possible random evaluations and outliers [89]. In addition, further analyses could be carried out to explore the differences among the scores for the considered test conditions, such as *t-test* and analysis of variance (ANOVA) for normally distributed results, and Wilcoxon Signed-Rank and Mann-Whitney tests for non-normal distributions. Furthermore, correlation analysis could help to extract further conclusions from the results [179].

Although the principles of all the aforementioned methodologies, originally designed for subjective evaluation of 2D video, could be valid to evaluate the 3D QoE, some adaptations are required to appropriately consider all the agents implied in the 3D visual experience and to reliably evaluate the perceptual factors influencing 3D QoE. In fact, an active research work is being carried out by the VQEG towards the definition of proper methodologies for subjective evaluation of 3D video [205]. For example, one important aspect that should be additionally considered when evaluating 3D video is the influence of picture quality, depth perception, and visual discomfort on the 3D QoE, as well as secondary perceptual factors like immersiveness and naturalness (which are totally subjective and could vary among observers). In fact, the evaluation of these factors lead to multidimensional statistical analysis of the results, and modifications in the techniques for collecting the observers' ratings. For instance, alternative approaches are being investigated based on the measurement of physiological parameters when observers are visualizing 3D content, such as ocular responses and electroencephalographic signals [137].

Moreover, the equipment and environment settings may change when using stereoscopic displays with specific glasses that reduce the luminosity and, in the case of active glasses, may cause flickering effects when the ambient light is not appropriate. In addition, the viewing distance and the position of the observers should be taken into account, as well as the location of the 3D screen in relation to the wall behind it, since projecting objects further away than the wall may cause incoherent 3D perception. Finally, the duration of the assessment tests may be reduced because of the possible experience of visual fatigue [19].

2.5.2. Objective evaluation metrics

The development of objective quality metrics arises from the necessity of obtaining automatically estimations of the QoE perceived by the end users, without the direct

Figure 2.37: PSNR performance [219]

participation of several observers, since subjective assessment tests are costly, time consuming, and not valid for many practical applications.

Traditionally, the Peak Signal-to-Noise Ratio (PSNR), or Mean Square Error (MSE), has been used (and is still used) to measure the quality of an image or video, which is based on the direct comparison of the original and the test content. Although this measure has been widely used thanks to its simplicity and good performance in certain applications and particular degradations, it does not take into account perceptual properties of the HVS (e.g., masking mechanisms). For example, as depicted in Figure 2.37, the center and the right images have the same PSNR with respect to the left image (original), but the perceptual quality is worse for the center image, since the artifacts in the right image are located in the lower part, which is much more textured.

In addition to PSNR that measures the quality in the image domain, other metrics have been used to estimate the quality of an audiovisual service based on QoS parameters, such as packet loss rate, delays, bandwidth, etc. [21]. However, although these metrics provide information about the performance of the service, which affects the QoE of the end users, they also do not take into account perceptual factors of the HVS, neither how the problems that are detected in the network affect to the content watched by the end users (e.g., the information carried out in the lost packets could affect differently depending on the type of packet and coding structure).

Therefore, during the last years an extensive research work has been carried out to develop methods to automatically estimate the QoE of the end users, in concordance with the properties of the HVS in relation with the visualization of video content. In fact, this work has resulted in the proposal of several approaches for conventional video, while the first steps are being done towards 3D quality metrics. Nevertheless, although different perceptual factors should be considered, similar design principles are being followed to develop quality metrics for 3D video, so the following discussion is valid for

Figure 2.38: Generic diagram of an HVS model [219]

both types of content.

In essence, two approaches have been followed to develop quality metrics during the past years: the visual modeling approach and the engineering approach [222]. On one side, the vision modeling or psychophysical approach, consists in modeling the perceptual components of the HVS related to quality perception, such as color processing, contrast sensitivity, and spatial and temporal masking (see section 2.3), as depicted in the diagram of Figure 2.38 [199, 219]. To achieve this, these models are based on the results of psychophysical subjective experiments carried out to characterize the perception mechanisms of the HVS. Moreover, since these approaches cover many perceptual factors of the HVS, they could be applied in the evaluation of many different types of degradations, although they are very complex and usually highly computationally expensive, so their practical applications are limited (especially for real-time evaluation) [222]. Some early examples of these metrics are the model proposed by Mannos and Sakrison [125] implementing the contrast sensitivity, and the first color model proposed by Faugeras [43]. Also, more recent models are the Visible Differences Predictor (VDP) by Daly [33], the Just Noticeable Difference (JND) model by Lubin [124] (widely used in various audiovisual applications), and the Digital Video Quality (DVQ) model by Watson [216] (considered in the VQEG projects for developing video quality metrics [83]). An example of an HVS model considering stereoscopic vision was proposed by Jin *et al.* [97], based on a 3D discrete cosine transform and considering factors as masking effects and contrast sensitivity.

On the other side, the engineering approach stands with the objective of obtaining less complex metrics limiting the generality of the algorithms, and focusing on the analysis of the particular degradations that may take place in the considered application. For instance, it is known that the coding process could introduce blurriness in the images, so an algorithm specialized on the analysis of the edges could estimate the quality degradations. Although with less versatility than complete psychophysical models, these metrics also take into account perceptual factors of the HVS in the analysis of the degradations and the estimation of the overall quality. Some popular examples of these metrics are the Structural Similarity (SSIM) proposed by Wang *et al.* [215], which computes statistics related to the structural information of the image, or the Video Quality Model (VQM) proposed by Pinson and Wolf [146], which extracts various properties of the video (e.g., spatial gradient and contrast, color and temporal information) to provide an estimation of the perceived quality.

Furthermore, there is another possible classification of the quality metrics depending on the required data to compute the estimations. Thus, there are three types of metrics [222]:

- Full-Reference (FR): These metrics require access to the original content to estimate the quality of the test content, since they are based on the comparison between the reference image and the test image (or frame by frame in the case of video). The extensive research work carried out concerning this type of metrics has resulted in high performance algorithms that provide reliable estimations of the viewers' QoE. However, the availability of the reference content is limited in several cases, so the practical application of these metrics is reduced to offline evaluation of systems or algorithms for image or video processing (e.g., encoders). One example of this type of metrics is the aforementioned SSIM.
- Reduced-Reference (RR): These metrics were developed with the objective of extending the range of applications of FR, so they do not need to access to the whole original content, but only to some parameters. Thus, certain features (e.g., coding settings, spatial and temporal indicators, statistical parameters, etc.) are extracted or computed from the original content, and they are compared with the same parameters from the test content to provide a quality estimation. Although they do not provide a performance as high as FR metrics, they could be used in quality monitoring scenarios, which is one of the main applications of video quality metrics [1], provided that the transmission of additional data over the network is possible. Some interesting examples of this type of metrics are the RR variant of the aforementioned VQM, or techniques based on watermarking [214].
- No-Reference (NR): These metrics do not require any type of information from the original content to carry out the quality estimation. Therefore, they are more versatile to be used in many practical applications than the previous alternatives. On the contrary, since the quality estimation is done working only over the test content, they are based on the detection of the effects of previously known impairments, and their performance is much worse. In fact, this has been reflected in the results of the VQEG for developing RR and NR metrics, since none of the proposed NR metrics fulfills the specified performance requirements [206]. In addition, this type of metrics could be classified into three groups: NR-Pixel-based (NR-P) if they are applied in the image domain (i.e., over the decoded frames of the video), NR-Bitstream-based (NR-B) if they work over the encoded bitstream, and NR-Hybrid (NR-H) if they consider both possibilities. A detailed review of the state-of-the-art on NR was carried out by Hemami and Reibman in [66].

To know the performance of the quality metrics it is necessary to compare their estimations of QoE with results from subjective assessment tests carried out with the same test material. In general, the validation of the metrics is carried out following some standard guidelines, which are based on the computation of correlation coefficients between the estimations and the subjective results, the (RMSE), and the outlier ratio [82]. In addition, previously to these computations it is possible to require a transformation of the results (e.g., using linear, polynomial, exponential, or logistic functions) to properly analyze the performance of the metrics. However, it is important to take into account the design of the subjective tests in order to extract valid conclusions of the performance of the objective metrics [152], since bias effects may appear in the experiments.

Figure 2.39: Diagrams of the metrics according to the required data from the reference

Finally, it is worth noting the guidance of standardization bodies, such as VQEG (with the current projects 3DTV, JEG-Hybrid, MOAVI, etc.) and ITU [28], in the development of reliable objective metrics for VQEG evaluation of 2D and 3D video.

3 Content production

3.1. Introduction

This chapter addresses the study of the aspects related to the generation of 3D video content. Therefore, firstly, a review of 3D video databases available for the research community and employed in quality evaluation activities is carried out. In addition, an outline of the state-of-the-art of studies considering the effects of 3D content creation on the viewers' QoE is presented.

Secondly, due to the lack of freely available 3D sequences for research purposes, a dataset of S3D videos was generated in collaboration with the IVC laboratory of the *Université de Nantes*. Thus, the process of generation of this freely-available dataset of high-quality 3D videos is described, which was carried out to contribute to the accessibility to 3D content for research activities. In fact, as mentioned later in this chapter, it has been already widely used in different works related to 3D video.

Moreover, a subjective evaluation is also presented, dealing with typical degradations that may appear during the video acquisition and processing. This test was aimed at understanding the impact of these degradations on the QoE, and may provide useful guidelines to follow when creating 3D content. The presented results are based on a large-scale subjective experiment carried out by various laboratories collaborating with the VQEG. In fact, since this experiment is still ongoing, preliminary analyses and results are presented, which may be completed once all the participant laboratories report their results.

Finally, some conclusions related to the contributions carried out in this chapter are provided.

3.1.1. SoA 3D databases

The availability of public video databases is a major requirement for the research community. In particular, it makes possible the access to useful content for developing video processing algorithms and systems. Moreover, it allows comparing the results of different approaches and studies using the same test data. However, in contrast to the large amount of monoscopic video databases [220], only a few have been already published with 3D sequences.

For instance, Goldmann *et al.* [50] captured various short Full-HD sequences with two consumer camcorders installed on a rig with parallel configuration, considering

different distances between both cameras. In addition, a database of stereoscopic Full-HD was published by Corrigan *et al.* [27] with the aim of facilitating the research on post-production techniques, which was captured using two cameras mounted on a side-by-side adjustable rig, and covering many of the scenarios characteristic to broadcast footage. Another example is the RMIT3DV dataset, published by Cheng *et al.* [23], consisting of uncompressed 3D Full-HD videos comprising a wide range of shooting environments, lighting conditions, textures, motion, etc. Furthermore, the DEIMOS database was generated by Klíma *et al.* [104] from different types of sources (e.g., blu-ray discs, 3D camcorders, satellite reception and static images), as an open-source database of images and videos for diverse applications, such as multimedia, television, security, assistive technology, biomedicine, astronomy, etc. Another interesting database was published by Xing *et al.* [224] composed of videos for analyzing crosstalk effects in stereoscopic and autostereoscopic displays, as well as other factors influencing the 3D QoE, as the camera baseline. Also, it is worth mentioning the database created by Hanhart and Ebrahimi [64] for visual attention modeling, providing eye tracking information of various S3D sequences; as well as the dataset presented by Scharstein and Szeliski [160] for evaluating the performance of depth estimation algorithms.

Apart from these datasets, some other collections of 3D videos have been used in standardization activities, as the development of the 3DV coding standard [77], and international projects, like the Mobile3DTV [169], which provide a detailed characterization of several sequences, although not all of them are freely available.

Moreover, an extensive work was carried out by Qualinet to generate a compilation of datasets for several research purposes in relation with image and video processing, including 3D image and video content, which could be found in [149].

3.1.2. Quality evaluation

To evaluate perceptual factors related to the creation of 3D content, the main alternative is to carry out subjective tests, as reflected by the number of studies already presented by the research community. For example, some experiments have been carried out to evaluate the effects of acquisition parameters, such as the distance between stereo cameras as investigated by Goldmann *et al.* [50]. Also, Ijsselsteijn *et al.* [73] studied the effects of the convergence, focal distance, and the vertical parallax, taking into account the appearance of keystone distortion and depth plane curvatures when using toed-in camera configurations. Moreover, interesting studies have been carried out by Yamanoue *et al.* [225] analyzing the impact of puppet-theater and cardboard effects caused by the use of convergent camera configurations, by means of subjective assessment tests.

In addition, some studies have been presented evaluating the depth quality of 3D videos [113], as well as the influence of features of the captured content, such as motion activity, on the 3D visual experience [121, 172].

Regarding the evaluation of 3D video processing techniques, some studies have been also published, for instance, evaluating depth adjustment methods to avoid visual

discomfort caused by excessive disparity [103], and analyzing the performance of 2D-to-3D conversion approaches [107, 148].

Although estimating the impact of content creation effects on the perceived QoE is a highly complex task, so subjective assessment tests are more appropriate, there are some works proposing automatic evaluations of these effects. For example, Yamanoue *et al.* [225] proposed a model to evaluate the impact of puppet-theater, and Masaoka *et al.* [126] developed a system to estimate spatial distortions in 3D images, such as puppet-theater and cardboard effects.

Furthermore, to control the appearance of annoying effects during the content creation, monitoring tools, as proposed by Grau *et al.* [53], could be used.

3.2. Generation of a freely available database of high quality stereoscopic 3D videos

The objective of the work presented in this section was to provide a high-quality freely available 3D video database, composed of 3D Full-HD stereo sequences, and especially designed to be used in 3D video subjective tests by the research community. The resulting database, called *Nantes-Madrid-3D-Stereoscopic-V1* (NAMA3DS1), is available at the website [74].

Therefore, one of the main aspects considered in the generation of the set of videos was to record representative scenes of typical contents watched in real-life environments (e.g., movies, sports, broadcasting news, etc.). So, this would entail obtaining, to a certain degree, interesting sequences for the observers that would participate in subjective tests. Thus, diverse shooting conditions were considered, such as indoor and outdoor sequences, and close (<5 m), medium (<10 m), and long recording distances. In addition, 3D effects that may be visually attractive for the observers (e.g., variations in depth, objects appearing in front of the screen, etc.) were also contemplated.

Besides, also technical issues were taken into account, regarding challenging features for coding algorithms (e.g., colorfulness, motion, texture, etc.), as well as perceptual effects that may influence the 3D visual experience of the observers (e.g., amount of disparity, motion in depth, etc.).

To describe the main properties of the resulting sequences, preview frames of the videos are shown in Figure 3.1 (the frame number is indicated in parenthesis), and exhaustive information of each sequence is provided in Table 3.1, including a content description, shooting conditions, camera parameters, and depth data. The parameters included in the table are described in the following subsections, where also a detailed description regarding the equipment and generation of the complete dataset is provided.

3.2.1. Equipment

The camera used to record the sequences of the database was the Panasonic AG-3DA1E [136]. Basically, this camera is equipped with a pair of twin-lenses, which are

Sequences	Barrier gate	Basket	Boxers	Hall	Lab	News report	Phone call	Soccer	Tree branches	Umbrella
Duration	16 s	16 s	16 s	16 s	16 s	16 s	16 s	16 s	16 s	13 s
Scenes	1	1	2	1	1	1	1	2	1	1
Condition	Outdoor	Indoor	Indoor	Indoor	Indoor	Indoor	Indoor	Indoor	Outdoor	Outdoor
Description	Car and barrier gate	Basket ball training	1. Boxer warms up 2. Boxing training	Persons meeting in a hall.	Two lab assistants working	News report mimic	Phone call in an office	1. 2 players score 2. Goal keeping	Tree leaves and wind	Person playing with an umbrella
	Medium	Long distance	Close distance	Long distance, high angle	Close distance	Close distance	Close distance	Long distance	Medium distance, low angle	Close distance
C	73	n. a	n. a	99	n. a.	50	36	n. a.	61	60
Z	0	n. a	n. a	0	n. a.	0	n. a.	n. a.	0	0
Source compression	No	Panasonic H.264 HP	Panasonic H.264 HP	No	Panasonic H.264 HP	No	No	No	No	No
SI	59	71	50 (50)	82	53	53	36	89 (89)	101	74
TI	21	41	56 (19)	5	12	4	13	38 (21)	14	19
DSI	20.42	11.72	24.4	17.02	17.77	21.58	21.56	24.70	23.02	17.02
DTI	15.43	9.69	18.03	6.97	10.04	8.71	11.78	18.08	13.63	15.24
Coding α	7.89	11.66	11.36	8.20	8.94	4.68	6.63	10.73	8.52	6.57
Coding β	9.88	2.97	3.01	7.94	9.34	25.99	19.98	3.76	3.09	9.38
D+	6	-14	9	7	6	16	17	7	3	5
D-	9	26	3	-3	22	6	15	10	9	17
Disparity histogram

Table 3.1: Characteristics of the sequences of NAMA3DS1 dataset

Figure 3.1: Sequences previews

adjusted and synchronized to avoid vertical ($<1.2\%$), angular, and rotational deviations, as well as brightness mismatches between left and right views. In addition, these lenses have a motorized $5.6\times$ zoom, providing a focal length range from 4.2 mm to 23.5 mm, and an aperture range from F1.8 to F2.4. Also, since it is a semi-professional camera aimed at providing natural-looking contents, the optical axes of the lenses are separated a distance close to the interpupillary distance (i.e., 60 mm approximately [51]).

The camera parameters used in each sequence, regarding white balance, zoom factor (Z), and convergence (C), were adjusted depending on the shooting conditions. These parameters, detailed in Table 3.1, provide useful information about the recorded scenes, but they may be insufficient for some applications, such as multi-view rendering or depth estimation algorithms, which were out of the focus of this work. It is worth noting that the zoom factor range covers from wide angle ($1\times$, $Z=0$) to long focus ($5.6\times$, $Z=99$), while the convergence ranges from 2.2 m ($C=0$) to infinite (parallel configuration, $C=99$).

When the shooting environment was appropriate, the sequences captured by the camera were transmitted via dual SDI streams to a *Clearview Extreme* system from *Video Clarity*, which made possible to store them in uncompressed format and 1920×1080 progressive Full-HD resolution per view with 25 frames per second. However, in some cases it was not feasible to use this system in the shooting environment, so some sequences were stored directly in the camera, in two separate SD cards *SanDisk Extreme* class 10. This entailed a compression of the videos with high-profile H.264/AVC encoder, 1080p25 resolution per view, and a maximum bitrate of 24 Mbps (average on these sequences: 21 Mbps). The latter alternative was used with sequences *Basket*, *Boxer*, and *Lab*, while the sequences *Barrier gate*, *Hall*, *News report*, *Phone call*, *Soccer*, *Tree branches*, and *Umbrella* were stored in uncompressed format.

3.2.2. Depth map generation

In order to complete the generated dataset, the depth maps of the recorded sequences were computed. In particular, they were generated from the captured stereo views using the disparity estimation algorithm proposed by Chambolle *et al.* [18], which is based

Figure 3.2: Depth maps previews (frame 0)

on a first order primal-dual convex optimization. For this purpose, this algorithm was adapted to stereo matching, so occlusion areas were taken into account in the optimization process.

In particular, two depth maps were computed for each frame (one using the left view as reference image, and another one using the right view) and they were stored in individual *.y16* files, where disparities are represented by 16 bits (one bit for the sign, nine bits for the integer part, six bits for the fractional part).

To provide some examples, a preview of the depth maps of each sequence is shown in Figure 3.2.

3.2.3. Content diversity

As recommended by international standards, like the ITU-T recommendation P.910 [86], a characterization of the sequences of the database was carried out. In particular, the following aspects were analyzed:

- **Spatial complexity:** The Spatial Perceptual Information (SI) factor was computed (on the luminance component of the left view) according to the guidelines described in ITU-T P.910 [86]. Mainly, it is based on the standard deviation over the pixels of the Sobel-filtered frames.
- **Temporal complexity:** Similarly, following the ITU-T P.910 [86] recommendation, the Temporal Perceptual Information (TI) factor was computed on the luminance component of the left view. In essence, it is based in the motion difference between adjacent frames.
- **Depth complexity:** The SI and TI factors were computed in the depth maps resulting from the process described above (subsection 3.2.2), obtaining the Depth Spatial Information (DSI) and Depth Temporal Information (DTI) factors. These indicators provide an idea of the spatial and temporal complexity of the depth of the scenes.
- **Coding complexity:** Taking as a reference the parameters proposed by Oelbaum and Diepold [133], the parameters offset (β) and slope (α) were computed obtaining a characterization of the sequence quality at a certain bitrate. These are the

Figure 3.3: Spatial, temporal, and coding characteristics

fitting parameters obtained from a linear mapping of the logarithm of the bitrate and the PSNR measured on the luminance component of the sequences.

All these indicators are shown in Table 3.1, together with the disparity histograms, and maximum crossed ($D+$, objects projected in front of the screen) and uncrossed ($D-$, objects projected behind the screen) disparities.

In addition, the distribution of the sequences according to their spatio-temporal and coding complexity are depicted in Figure 3.3a and Figure 3.3b, respectively. As shown, the set of sequences are spread over a wide range in those dimensions, reflecting the heterogeneity of the NAMA3DS1 database.

3.2.4. Applicability of the dataset

The NAMA3DS1 dataset has been already used by the research community for different purposes regarding 3D video. For example, VQEG have considered this dataset for their 3DTV activities [205]. In addition, it has been used for different subjective studies regarding 3D: a cross-lab study analyzing the impact of display technology and test environment [122], a study of the preference between S3D video over 2D video [114], an evaluation of stereoscopic imaging systems considering visual attention [65], and an analysis about the impact of chromatic variations in the 3D QoE [142].

Furthermore, it has been used for developing and testing objective quality metrics [34, 39, 191], developing models for visual saliency [211] and visual attention [49], and for eye-tracking [42] and watermarking applications [2].

3.3. Subjective evaluation of typical degradations

This section presents a subjective evaluation of some typical degradations caused by the processes of capture and post-production of the video through the 3DTV distribution chain. These degradations were evaluated under the scope of a wider test designed by the VQEG within the 3DTV project [205]. The objectives of this experiment are, on one side, to establish a ground truth database for Quality of Experience in 3DTV (GroTruQoE3D) regarding measurement methodologies, and to evaluate and determine a set of verified assessment methods for 3D video to contribute to the standardization in ITU recommendations. Thus, the presented test is included in the first step towards these objectives, which provides a subjective dataset of videos with 3DTV typical degradations, exhaustively evaluated using the PC method.

This large-scale experiment is carried out in collaboration with several laboratories that follow specific guidelines and use a common dataset to obtain reliable and valuable results. In essence, the results of the GroTruQoE3D evaluation may be used to verify different issues, such as the performance of existing subjective quality assessment methods, the impact of different perceptual measurement scales, and the influence of observer training on the results; and will help in the development of new methodologies [209].

It is worth mentioning that the whole inter-lab experiment, which is currently still ongoing, considers the evaluation of many other test conditions that are not included in this section, since, to analyze them, it is necessary to obtain the results from all laboratories. Therefore, the results presented in the following subsection 3.3.2 are derived from the tests carried out at *Universidad Politécnica de Madrid* (UPM) after the statistical analysis that were applicable at the moment of writing this section. Nevertheless, in subsection 3.3.2.2, some other preliminary results are provided.

3.3.1. Subjective test setup

3.3.1.1. Evaluation methodology

The subjective assessment method used was PC, in which, as described in subsection 2.5.1, the test sequences are presented in pairs and the observers should indicate which one provides a better QoE. Thus, after watching a pair of tests sequences, the observers had a period of time in which they had to answer the question: “Which one do you prefer?”. In the particular case of the test carried out at UPM, this was done using a keyboard when a message, over mid-gray background, was shown on the screen indicating the observers to vote (the message remained in the screen until the observer pressed a selection button).

Both time-parallel and time-sequential configurations for PC (see subsection 2.5.1) were allowed to be used in the participant laboratories. In particular, time-parallel configuration was used at UPM.

The test sessions started with an exhaustive explanation of the test methodology and purposes. In addition, some pair sequences were showed to the observers as a training

Figure 3.4: Test environment at UPM

Display	Model	LG DM2350D-PZ
	Size	23"
	Resolution	1920x1080
	Refresh rate	60 Hz
	3D technology	Polarized glasses
Viewing distance		4.5H
Voting interface		Keyboard (left and right arrow)

Table 3.2: Test equipment at UPM

for the formal test, in order to make them familiar with the methodology. Then, the test session consisted of three parts in which a number of pairs were evaluated by the observers. Between each part there was a break, so the observers could rest to avoid fatigue. Thus, the total duration of a test session was one hour approximately. Also, in this test, only one observer participated at a time in each test session.

3.3.1.2. Environment and equipment

The test environment was set following, as close as possible, the recommendations specified by the ITU-R BT-500 [89]. In particular, gray curtains were installed in the walls, and controlled faint ambient light was allowed to avoid disturbing reflections. As depicted in Figure 3.4, where a photograph of the test environment is shown, two screens were used side-by-side, and the viewing distance was set to 4.5H, as a compromise between vertical and horizontal resolution [209]. The main properties of the display, which were calibrated using the *X-rite i1 Display Pro* tool, are shown in Table 3.2. In addition, a high-performance computer was used with SSD hard drives, a graphic card *Nvidia Quadro FX 580*, and a specialized stereo player developed by IVC laboratory to assure a smooth playback of the sequences.

3.3.1.3. Test material

The source sequences used to generate the test material were selected to have a wide representation of different properties, such as coding complexity, motion, brightness, disparity, etc. In addition, natural content recorded with stereoscopic cameras as well as animation videos were considered. Some details of the sequences are shown in Table 3.3. All the stereoscopic videos were stored in *avi* containers at 25 fps, *yuv422* format, and Full-HD for both views. Moreover, all of them have a duration of 16 seconds (i.e., 400 frames).

Although several test conditions (18 distortions or Hypothetical Reference Circuits (HRC)¹ and 11 source sequences) were considered in the whole VQEG test [209], under the scope of the present work, the attention is focused on those described in Table 3.4.

In particular, in HRC1 the right view has a gamma offset to the left view, so brighter parts of the images are emphasized (black level remains unchanged), which simulates an incorrect luminance calibration of the cameras. HRC2 and HRC3 present a horizontal disparity offset of the left and right views with opposite sign. This entails a depth shift of the content, since the disparity is changed. Concretely, the shift of -30 pixels for HRC2 and 30 pixels for HRC3 were introduced symmetrically, so the total disparity offset was -60 for HRC2 and 60 for HRC3, as depicted in Figure 3.5a. A similar process was applied to HRC4 but in vertical direction, simulating a wrong camera calibration, which usually causes difficulties to fuse both stereo views. Again, the shift of -20 pixels has been introduced symmetrically, so the total disparity offset was -40 pixels. Also to simulate strong camera calibration errors, the left view in HRC5 was processed to add geometric deformations, in particular, related to displacing the corners. HRC6 is based on the conversion to 3D of monoscopic videos, which can be obtained by a geometrical stretching transformation. Concretely, a simple algorithm proposed in [230] was used. As depicted in Figure 3.6, it is based in three steps: the production of a second view from the input 2D video using *Avisynth* and *Virtualdub*, a framerate reduction skipping one frame, and a cropping of the video to separate both stereo views. In the case of HRC7-HRC9, they were obtained encoding the sequences with H.264 with the reference software JM 18.2. In particular, HRC8 was processed after decoding applying a edge enhancement filter, and HRC9 is obtained taking the left view of the videos generated for HRC7. In addition to HRC1, HRC10 simulates a wrong luminance calibration of the cameras, adding a brightness offset to the left view. Finally, HRC11 and HRC12 are the undistorted references for 3D and 2D video, respectively.

In addition, it is worth noting that, due to the amount of test conditions to evaluate in the global GroTruQoE3D project, the design of the tests was done taking advantage of the participation of various laboratories, so the test conditions were distributed among the different laboratories [209]. Thus, some HRCs were not evaluated for all sources at each lab, since this would have resulted in unfeasible tests sessions. However, there is a common set of test conditions evaluated by all the participant laboratories, which allows to analyze the factors that may influence the results in different laboratories.

¹According to the VQEG terminology, an HRC is a fixed combination of a video encoder at a given bitrate, network condition and video decoder.

3.3. Subjective evaluation of typical degradations

ID	Description	Preview	Origin
SRC1	Car and barrier gate.		NAMA3DS1
SRC2	Animation: old man dancing on platform.		http://www.elephantsdream.org
SRC3	A woman standing in the foreground and a fountain in the background.		National Institute of Information and Communication Technology
SRC4	Tree leaves and wind.		NAMA3DS1
SRC5	Person flying a drone.		<i>Université de Nantes</i>
SRC6	Basket ball training.		NAMA3DS1
SRC7	News report mimic.		NAMA3DS1
SRC8	Animation: Two people attacked and projected in a capsule.		http://www.elephantsdream.org
SRC9	People meeting in a hall.		NAMA3DS1
SRC10	Women walking in a park.		National Institute of Information and Communication Technology
SRC11	Water taxi is sailing at the wharf.		National Institute of Information and Communication Technology

Table 3.3: Source sequences

HRC	Description
1	Asymmetric gamma at 0.5
2	Horizontal disparity offset -30
3	Horizontal disparity offset +30
4	Vertical disparity offset -20
5	Graphical distortion with Stirmark
6	2D to 3D
7	3D encoded with H.264 (QP=44, GOP IBBP64)
8	Edge enhancement at 40% (encoded with H.264, QP=32, GOP IBBP64)
9	2D encoded with H.264 (QP=44, GOP IBBP64)
10	Brightness at 80%
11	Reference 3D
12	Reference 2D

Table 3.4: Test conditions **(a)** Horizontal disparity offset **(b)** Vertical disparity offset**Figure 3.5:** Disparity offsets [209] **Figure 3.6:** Processing chain for 2D to 3D conversion [209]

Concretely, the common set was formed by nine HRCs (HRC1-HRC9) and two source sequences (SRC1 and SRC2) [209]. Although they are not in the common set, so a different processing of the results is necessary, HRC10-HRC12 have been also considered to complete the study.

Moreover, the ORD design (see subsection 2.5.1) was applied to reduce the total number of possible comparisons to those more useful and interesting [120]. Thus, firstly, a preliminary test was carried out by experts to rank the HRCs according to their potential impact on the QoE and to select the pair comparisons to consider. This resulted in the following rank ordering (from best to worst QoE) of the HRCs considered in the current test: {12 11 3 2 8 6 1 4 5 10 7 9}. Also, the R_{ORD} matrix shown in Table 3.5 was obtained, where only the HRCs considered in the present study have been indicated. This matrix reflects the comparisons carried out, since only the HRCs located in the same column or row in the R_{ORD} matrix are compared in the tests [120]. Further details of the design could be found in [209].

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

Table 3.5: R_{ORD} matrix

Finally, an important aspect regarding the presentation of the test sequences, is that they were differently randomized for each observer with the restrictions that the same content was not shown consecutively, and the presentation orders (i.e., stimulus A in the left screen and stimulus B in the right, and vice versa) were balanced for all observers.

3.3.1.4. Observers

A total of forty observers participated in the test at UPM, all of them having normal or corrected-to-normal vision. This was checked prior to the test using the Snellen chart for visual acuity, the Ishihara test for color vision, and the Randot test for stereo vision. All of them were naive observers, so their work did not involve video picture quality and they were not experienced assessors. In addition, they had not participated in a subjective quality test over a period of six months before this experiment.

In particular, 21 males and 19 females took part in the test, with a distribution of ages between 18 and 49, and an average of 23.

3.3.2. Experimental results

As aforementioned, the total number of test conditions have been distributed among all the participant laboratories, so, at this moment, only preliminary results can be provided, until all the participant laboratories complete the tests. Thus, to analyze

Figure 3.7: Comparison of the test conditions

the effects of the considered HRCs, it is only possible to apply the BT model to the test conditions included in the common set, while for the rest of them just a raw-data analysis could be performed. It is worth mentioning that twenty observers evaluated the pairs of the common set in each laboratory.

3.3.2.1. BT-model

Results at UPM The global results obtained with the BT model considering the test conditions included in the common set (i.e., grouping the evaluations for HRC1-HRC9 and SRC1-SRC2) are shown in Figure 3.7. Since the BT model provides relative scores, it is necessary to establish a reference setting its BT-score to zero, in order to carry out the comparisons among all the considered conditions. Thus, HRC5 (geometric distortions) was taken as reference, since it was expected to be the most annoying degradation caused in the content creation process, and the one with the lowest BT-score.

As it can be seen, although there is no statistical significance among the different geometric effects considered in the tests (HRC2-HRC5), the graphical distortions simulated in HRC5 seem to be the most annoying. Also, it is worth noting the closeness among the BT-scores obtained for horizontal and vertical disparity offsets (HRC2-HRC4), since it may be expected that vertical misalignments could cause more visual discomfort. In addition, the results show that all these geometric effects critically degrade the viewers' QoE as reflected by the comparison of these results with HRC8 (an encoded version of the original 3D video processed with an edge-enhancement filter, which presents the highest rating and is expected to be worse than the 3D original videos [209]). On the other hand, as reflected by the results for HRC1, degradations

Figure 3.8: Comparison of the test conditions for both sources

caused by a deficient luminance calibration of the cameras may be less disturbing than geometric distortions.

In addition, the geometric effects seem to similarly degrade the QoE as a poor coding quality of the 3D video, as derived from the comparison of the BT-scores of those HRCs and HRC7. This may be caused by the perceptual effects of a coarse quantization in the encoding process, since severe artifacts, such as blocking effects, may appear even producing difficulties in fusing both stereo views.

Regarding the 2D to 3D conversion (HRC6), although it provides better results than a coarse encoding of the 3D quality, there is no appropriate reference to compare within the common set, so a deeper analysis of the raw data from the tests was carried out and described in next subsection 3.3.2.2.

As it can be seen in Figure 3.8, similar behaviors are observed in the BT-scores obtained for both sources and the global BT-scores in Figure 3.7. Although it seems that for SRC2 bigger differences are achieved between the BT-scores of the HRCs with respect to the reference (HRC5), it is also important to notice that there were less agreement among the observers evaluating SRC2, as indicated by the greater CIs. This fact may be caused by the type of content of SRC2, since animated 3D videos could provide a more comfortable and better visual experience to the observers. Also, this type of content may provoke that the degradations caused by the encoding process or asymmetric gamma could be less perceivable. Moreover, although not statistically significant, HRC4 seem to disturb less than HRC2 and HRC3 in SRC2 compared to SRC1, so the impact of a vertical misalignment may be influenced by the disparity characteristics of this sequence.

Lab	Display	Technology	Size	Test method
IVC	Philips 46PFL9705H 46	Shutter glasses	46"	Time-parallel
T-Labs	Samsung UE40ES8000/8090	Shutter glasses	40"	Time-parallel*
Yonsei	Not reported	Not reported	-	Time-sequential
INSA	Samsung SMART LED-TV	Shutter glasses	40"	Time-sequential

Table 3.6: Settings of the other test laboratories

Comparison among laboratories The same test was also carried out in other laboratories with the aim of completing a large scale experiment. Although the same conditions regarding the test environment and observers were requested for all laboratories, different equipments and configurations of the PC methodology were allowed. The main details regarding these differences are shown in Table 3.6.

It is worth noting that a different method of the time-parallel PC, described in 2.5.1, was used at T-Labs. Concretely, both stimuli of each pair to compare are played in the same screen simultaneously, but only one of them is shown in the screen in a specific instant (i.e., as if the other is running in background). Then, the observers could freely switch between one stimuli to the other (pressing a key on a keyboard). This way, the observers are able to compare a pair of videos permuting as many times as they want between both, and indicate their preference when the playback is finished. One of the main advantages of this methodology is that, switching quickly between the two videos, the observers could perceive small differences. On the contrary, this method presents the drawback that, since only one of the videos is rendered at a time, the observers could miss some of the degradations if they appear when they are watching the other stimuli.

The results obtained in the different laboratories for the test conditions evaluated in the common set are shown in Figure 3.9 (where HRC5 has been maintained as reference in each case). It can be observed a similar behavior in all laboratories for test conditions regarding conversion from 2D to 3D and coding configurations (HRC6-HRC9). However, in the case of UPM the artifacts introduced by the encoding process seem to be less perceptible than in the other laboratories. This could be caused by the size and type of display (polarized technology which entail a loss of resolution) that could hide some of these degradations.

On the contrary, for the rest of HRCs there is more discrepancy among the laboratories. Especially, regarding the impact of geometric effects (HRC2-HRC5) the differences among the BT-scores for these HRCs are very small at UPM, while in the rest of the laboratories horizontal disparity offsets (HRC2 and HRC3) are significantly better than geometric distortions introduced with *Stirmark* (HRC5). This worse performance of the disparity changes may be caused also by the size of the display and the viewing distance at UPM, that may produce more visual discomfort when geometric distortions degrade the video, but also may be motivated by the polarized technology of the display, since more crosstalk (which also significantly changes with small movements of the viewer's head) could be perceived with these degradations affecting the disparity.

Finally, also there are some differences regarding the effect of having asymmetric

Figure 3.9: Comparison among the different laboratories

gamma (HRC1), although this may be caused by the presentation method used in the tests, since UPM and IVC (which used time-parallel presentation seeing both videos of each pair at the same time) present a similar trend against the rest of the laboratories, where worse results for this test conditions were obtained.

3.3.2.2. Analysis of raw data

Apart from the HRCs included in the common set, other test conditions were evaluated in the VQEG tests that are interesting in this analysis of impact of 3D content creation effects. However, as aforementioned, since the global tests are still ongoing, only preliminary analyses over the raw data obtained from the experiments can be carried out. Thus, Tables 3.7-3.13 show the results of the observers' preferences from the pair comparisons in each lab, for those source videos in which the HRCs have been already evaluated.

Firstly, to get a notion about the observers' preference of 3D video over 2D video, the comparisons about the references of both formats (HRC11 and HRC12) have been collected in Table 3.7. In essence, the results show an important preference for 2D content at UPM, while at other laboratories 3D content seem to provide a better QoE, although more results are needed to confirm this. The cause of the worse performance of 3D video obtained at UPM may be the properties of the display, especially the small size, used in the tests (as also reflected by the results described in subsection 4.2.2).

As aforementioned, the analysis over the common set does not provide enough data to

Lab	Source	Pref _i	Pref _j
UPM	2	12	28
	3	5	35
	5	5	35
	7	6	34
	10	16	24
	Global	44	156
IVC	10	26	14
T-Labs	3	21	19
	9	28	12
	Global	49	31

Table 3.7: Comparisons between HRC11 and HRC12

Lab	Source	Pref _i	Pref _j
UPM	8	40	0
	10	37	3
	11	35	5
	Global	112	8
IVC	11	32	8
T-Labs	1	31	9
Yonsei	7	28	12
	9	35	5
	Global	63	17

Table 3.8: Comparisons between HRC11 and HRC6

Lab	Source	Pref _i	Pref _j
UPM	4	38	2
	8	39	1
	11	37	3
	Global	114	6
IVC	5	31	9
	9	31	9
	Global	62	18
INSA	11	25	15

Table 3.9: Comparisons between HRC12 and HRC6

Lab	Source	Pref _i	Pref _j
UPM	9	14	26
IVC	11	32	8
T-Labs	10	30	10
Yonsei	1	33	7
	6	33	7
	Global	66	14

Table 3.10: Comparisons between HRC11 and HRC8

Lab	Source	Pref _i	Pref _j
IVC	6	6	34
	8	2	38
	9	3	37
	Global	11	109
T-Labs	9	1	39
Yonsei	6	9	31

Table 3.11: Comparisons between HRC10 and HRC1

properly investigate the performance of the conversion of 2D content to 3D (HRC6). Thus, the results from the direct comparisons between this test condition and the 3D reference (HRC11) carried out by the observers are shown in Table 3.8. In fact, a clear preference for the natural 3D content could be observed. In addition, the low QoE provided by the converted 3D content is confirmed with the results of the comparison with the 2D reference (HRC12), shown in Table 3.9, since a significant preference for the monoscopic versions is demonstrated. These results are in line with the outcomes of other studies comparing converted and natural 3D content, like the study presented by Kulyk *et al.* [107].

Similarly, the results from the common set were not sufficient to deeply study the effects of the edge enhancement applied over the encoded 3D video in HRC8. Thus, Table 3.10 shows the results of the comparisons between this test condition and the 3D reference (HRC11). As it can be seen, there is a general trend of a significantly better QoE provided by the 3D reference, except for the results obtained at UPM. Although this test condition was only evaluated over one source video, so more data is necessary to find a solid cause, probably the size of the display and the polarized technology may hide the effects of coding artifacts producing that the observers perceive a similar QoE than with the 3D reference.

Finally, since it was not included in the common set, the effects of an asymmetric brightness (HRC10) are analyzed with the results from the direct comparisons of this HRC with other test conditions related to incorrect camera calibration: asymmetric gamma (HRC1), vertical disparity offset (HRC4), and geometric distortions (HRC5). The obtained results are shown in Tables 3.11-3.13, respectively, which show a less perceptible effect of the asymmetric gamma, and an unclear preference against HRC4

Lab	Source	Pref _i	Pref _j
UPM	1	22	18
IVC	1	23	17
	2	18	22
	4	18	22
	7	28	12
	Global	81	79
T-Labs	4	25	15
	8	14	26
	Global	37	41
Yonsei	1	7	33
INSA	6	13	27
	7	18	22
	Global	70	50

Table 3.12: Comparisons between HRC10 and HRC4

Lab	Source	Pref _i	Pref _j
IVC	2	27	13
	6	27	13
	11	16	24
	Global	70	50
T-Labs	6	10	30

Table 3.13: Comparisons between HRC10 and HRC5

and HRC5. Thus, it is possible that the asymmetric brightness could be as annoying as geometric modifications, but more results are need to state this conclusion.

3.4. Conclusions

The generation of high-quality 3D contents is one of the most crucial aspects to provide a satisfactory QoE to the viewers. Actually, the lack of well-recorded videos and the appearance of poor-quality 3D sequences is considered one of the most important facts that have hindered the success of the recent attempt to establish 3D video technology in the consumer market.

Furthermore, there is still a reduced availability of 3D video datasets free from license restrictions for research purposes in relation to subjective assessment of 3D QoE. Thus, in this chapter, the generation of a high-quality freely available S3D video database has been presented. In essence, it has been designed to be used by the research community for 3D video applications, covering a wide range of sequences concerning spatial, temporal, depth, and coding complexity; as well as a broad variety of representative contents typical to those watched in real conditions. In fact, the applicability of the

dataset has been demonstrated by many research works that have been presented by the community using it for different purposes, such as subjective tests, and development of visual models and objective quality metrics.

Moreover, both in professional and consumer environments, several degradations could appear in the process of content creation. From incorrect camera calibrations to annoying disparity effects or poor techniques for processing the acquired 3D video. Thus, this chapter presented a subjective test that has allowed the study of the impact of some degradations caused in the content generation process, such as disparity offsets, geometric distortions, brightness or gamma mismatches between views, or 2D to 3D conversions. Preliminary results have been exposed and some relevant conclusions were obtained, although the presented test, carried out in collaboration among different laboratories under the scope of the VQEG project GroTruQoE3D, is still ongoing. In essence, the main points that arise from the results are related to the significant annoyance of geometric distortions and brightness asymmetries, caused by incorrect camera calibrations, as well as the requirement of a precise and conscious method for converting monoscopic to 3D content to achieve satisfactory results. Also, the use of different displays and methods for presenting the test sequences using PC, have allowed to analyze the influence of the visualization technology and the discriminability of certain degradations.

4 Coding

4.1. Introduction

This chapter addresses the evaluation of the impact of coding effects on the viewers' QoE. With this aim, firstly, a review of the state-of-the-art of subjective studies and objective metrics developed to analyze the effects of coding degradations in 3D video is presented, complementing the technical and perceptual issues described before in subsection 2.4.2.

Then, in section 4.2, a subjective experiment carried out to evaluate the impact on the 3D visual experience of different coding settings (e.g., encoder implementations, bitrates, etc.) is presented. In addition, one of the main objectives of this experiment was to investigate the use of different representation formats for stereoscopic content in distribution services providing 3D video to home environments, such as SbS, FCC, and 3D Full-HD. Therefore, the subjective experiment was carried out in two different laboratories, at UPM and IVC, using test environments and equipments similar to real domestic situations. Specifically, different displays were used in both laboratories, in terms of technology and size, to analyze the QoE provided by the different alternatives concerning formats and coding systems to bring 3D video to households. In addition, the visual experience provided by 2D content was compared with the aforementioned formats, to analyze the added value expected from 3D video.

Finally, some conclusions are presented in section 4.3.

4.1.1. Quality evaluation

The compression of the data when video content is encoded could cause some quality degradations, as previously described in subsection 2.4.2, so several studies have been proposed in the literature analyzing the impact of these effects in the viewers' QoE. In particular, among the numerous subjective tests analyzing coding degradations, the validation studies carried out during the development of coding standards should be pointed out, such as HEVC [7], SVC [134], MVC [204], and 3DV [78].

Apart from these experiments, some interesting studies have been carried out, for example, analyzing the effects of V+D formats [198], and showing that the HVS is less sensitive to distortions affecting the depth map than those degrading the quality of video color. Moreover, other studies have analyzed and compared the performance of different coding schemes (simulcast and MVC) and settings with S3D video [213].

Similarly, Lebreton *et al.* [114] investigated the preference of S3D content over 2D considering different coding configurations. Finally, it is worth mentioning the studies that have been carried out investigating the subjective performance of asymmetric coding of 3D video, using different quantization parameters or resolutions (spatial or temporal) for each stereo view [158, 176], or applying this idea to depth map coding [198].

On the other side, regarding objective quality metrics, the extensive research work carried out during the last decades in relation to conventional video has led to the development of some metrics that provide high performance evaluating the impact of coding degradations [24]. In fact, ITU has included the VQM metric in the recommendation ITU-T J.144 [84]. In addition, although mainly FR metrics are used for offline evaluation of coding systems, such as the new HEVC encoder [210], some NR metrics have been also proposed to be used in quality monitoring systems [66].

In contrast, only a few metrics have been published to evaluate the coding quality of 3D videos. In fact, the first approaches were based on the direct use of 2D metrics [227], or applying some modifications to these metrics to take into account 3D video features, as proposed by Politis *et al.* [147] using VQM. Nevertheless, some specific metrics for 3D video have been already developed. For instance, Gorley and Holliman [52] presented a 3D quality model based on analyzing both stereo views considering the visual sensitivity to contrast and luminance changes. Another example was proffered by Seo *et al.* [164], based on the analysis of blocking and blurring effects in S3D video, considering the quality differences between both stereo views.

4.2. Subjective evaluation of 3D coding artifacts

The experiment presented in this subsection was carried out under the scope of the JEDI (Just Explore Dimensions) ITEA2 european project [96] with the aim of evaluating different encoders, bitrates, and image formats for 3D video broadcasting. The test was performed in two laboratories (the IVC laboratory of the *Université de Nantes* and the Dem-3DTV of the UPM) evaluating several test conditions, so it was divided in two experiments:

- Experiment 1: It was focused on the comparison between the PoE of different test conditions, thus, all the HRCs considered in the test were included.
- Experiment 2: Planned to analyze the influence of video content on the observers' PoE, so all the source sequences were considered, but only a subset of representative HRCs were included.

As detailed in the following subsections, Experiment 2 is considered a subset of Experiment 1, so the same test setup was used in both experiments, except the test conditions and observers.

Figure 4.1: UPM test environment

4.2.1. Subjective test setup

4.2.1.1. Evaluation methodology

The methodology used in the test was the PC method, so the observers were asked to select which video they preferred after seeing each pair of sequences. Since some of the considered test conditions were difficult to differentiate due to their close quality, the time-parallel PC method (described in subsection 2.5.1) was used to have a higher discriminability on stimuli with similar quality levels [116]. Thus, two displays were placed side-by-side to show the pairs of sequences. As in the test described before regarding content production degradations in section 3.3, this was done using a keyboard when a message was shown on the screen indicating the observers to vote.

The test session included a previous training in which the test methodology was exhaustively explained to the observers and they practiced with five training pairs of sequences. In addition, each whole test session, which lasted approximately one hour, was divided into two parts with a similar duration, and a break time of about five minutes in-between.

4.2.1.2. Environment and equipment

The test environment was set up trying to mimic, as far as possible, real home environments, in order to obtain more reliable results than using standardized environments, which could distance the observer from typical viewing conditions. In particular, the test configuration used at UPM is shown in Figure 4.1. The ambient illumination of the test room was almost constant and fitted to avoid reflections on the screen that could affect the observers judgments. In addition, the equipment used at UPM laboratory was the same employed in the tests described before regarding content creation (see Table 3.2). Thus, the viewing distance was set to 90 cm, approximately, which corresponds to three times the height of the screen of the displays (3H).

Source	Castle	Rome
Preview
Description	A camera is panning behind some old arches which overlooks a castle.	Camera pan showing a fountain (Fontana di Trevi).
SI	59	57
TI	15	22
DSI	5.75	5.1
DTI	0.9	3.74
D+	18	12
D-	16	17

Table 4.1: Source sequences

4.2.1.3. Test material

Six stereoscopic Full-HD video sequences were selected as source material, in order to cover a wide range of video characteristics, such as spatial properties (textured versus uniform areas, contours and gradients, etc.), temporal properties (amount and type of parallax and scene motion, etc.), and depth properties (small and large depth budget, distribution of the depth budget, pop out effects, etc.). In particular, four of the source videos were chosen from the the NAMA3DS1 database: *Barrier*, *Soccer*, *Tree branches*, and *Umbrella*. Their main properties are detailed in Table 3.1. The characteristics of the other two source sequences are shown in Table 4.1. The framerate for all video sequence was 25 fps.

The HRCs considered in the test, whose properties are shown in Table 4.2, were chosen to study the effects of the coding process in typical broadcasting systems, in terms of encoding, allocated bitrates, and 3D video formats (see subsection 2.4.2). Specifically, in addition to 2D in Full-HD, three stereoscopic video formats were used: 3D Full-HD, SbS, and 3D FCC. Moreover, three test reference conditions were considered for 3D Full-HD, 3D SbS and 2D Full-HD, while the rest of HRCs were obtained using different encoders and typical bitrates for broadcasting systems. In particular the three considered encoders were:

- E0: Hardware implementation of a 3D stereo real-time encoder (MVC) designed to be used at the head end of the distribution system for online contribution services. Thus, high quality of the source video and low delays were strongly requested.
- E1: Hardware implementation of a 3D stereo real-time encoder (MVC) for video distribution (e.g., using IP, satellite, or terrestrial networks) with strong requirements on bandwidth saving.
- E2: Similar to E1, except that the implementation is based on the standard

HRC	Encoder	Standard	Image format	Bitrate (Mb/s)
1	Ref.	-	3D Full HD	-
2	Ref.	-	3D SBS	-
3	Ref.	-	2D Full HD	-
4	E0	H.264/MVC	3D Full HD	20
5	E1	H.264/MVC	3D Full HD	12
6	E1	H.264/MVC	3D Full HD	9
7	E1	H.264/MVC	3D Full HD	6
8	E2	H.264/AVC	3D SBS	12
9	E2	H.264/AVC	3D SBS	9
10	E2	H.264/AVC	3D SBS	6
11	E3	H.264/MVC	3D Full HD	12
12	E3	H.264/MVC	3D Full HD	9
13	E3	H.264/MVC	3D Full HD	6
14	E3	H.264/MVC	3D FCC	12
15	E3	H.264/MVC	3D FCC	9
16	E3	H.264/MVC	3D FCC	6
17	JM18.2	H.264/AVC	3D Full HD	9
18	E2	H.264/AVC	2D Full HD	12
19	E2	H.264/AVC	2D Full HD	9
20	E2	H.264/AVC	2D Full HD	6

Table 4.2: Test conditions

H.264/AVC used to encode 2D and SbS contents.

- E3: Software implementation of a 3D stereo encoder (MVC) for offline contribution service (e.g., IPTV, content delivery networks, or video-on-demand servers) with strong requirements on quality or bandwidth saving.
- JM version 18.2: Reference software of the H.264/AVC encoder provided by the Joint Video Team (JVT) of the ITU. It should be noted that the reference software is not optimized to perform close to real-time encoding.

As aforementioned, different sets of HRCs were used in both experiments. In particular:

Experiment 1 All HRCs were considered in this experiment, and the ORD was used to reduce the number of comparisons to carry out (see subsection 2.5.1), since evaluating all possible pairs would be infeasible in terms of duration of the assessment sessions. Thus, the HRCs considered in the comparisons are detailed in the following R_{ORD} matrix, which was designed following two aspects. Firstly, comparisons between test conditions specially interesting in this study were taken into account, such as high quality conditions, different encoders, and image formats at a given bitrate. Secondly, the results from a previous visual verification of the stimuli by experts were used to know the possible impact of each test condition on the observers' QoE (more details

can be found in [120]). In particular, the HRCs were ranked according to their impact on the QoE, resulting in the following order (from best to worst expected QoE) {1, 4, 5, 11, 14, 6, 12, 15, 17, 2, 8, 9, 7, 13, 16, 10, 3, 18, 19, 20}. Also, as aforementioned, following the ORD, only the pairs located in the same column or row in the R_{ORD} matrix are compared.

1	4	3	16	13
2	5	11	14	9
8	17	12	6	15
18	20	19	10	7

Table 4.3: R_{ORD} matrix for Experiment 1

Considering the six source videos and the fact that each pair should be presented twice in different order (e.g., stimuli A in the left display and stimuli B in the right display, and vice versa) to avoid the effect of possible differences between both displays, a total of 840 comparisons should be done for one complete observation. This entails that one observer should take part in a session of approximately 3.5 hours, which is impracticable. Thus, groups of eight observers were formed that performed a complete observation, so each observer of these groups evaluated 105 pairs. This led to reasonable sessions of approximately forty minutes of viewing time. It is worth noting that grouping the observers to evaluate all the test conditions entails the assumption that the voting is independent of the observers, which is analyzed in the following subsection 4.2.2, comparing the results of Experiment 1 with Experiment 2, since the latter one is a subset of Experiment 1, in which one observer performed the whole subset test.

The video pairs used in the tests were randomized taking into account that one complete observation was conducted by eight observers and the following restrictions:

1. The presentation order for each HRC should be as balanced as possible (i.e., HRC_i is not always presented on the left).
2. The presentation order of the video sequences should be as random as possible. In particular, no observer watches the same source content in two consecutive presentations.
3. The number of source videos for each observer should be as balanced as possible (e.g., no observer watches only two source contents in the test).
4. The frequency of occurrence of each HRC should be as balanced as possible for each observer (e.g., HRC_i is not balanced if one observer watches HRC_i only once but another observer watches HRC_i several times).

Experiment 2 Since the objective of Experiment 2 was to analyze the influence of video content, all source videos were considered. However, nine out of the twenty HRCs (concretely: 1, 2, 3, 5, 7, 10, 16, 18, 20) were selected, to make the pair comparison test feasible (within one hour). Again, using the ORD, this led to the evaluation of 108

Experiment	Exp.1	Exp.2
Number of observers	32	30
Gender (m/f)	26/6	19/11
Age range (average)	21-49 (29.9)	23-52 (31.5)

Table 4.5: Observers at UPM

pairs by each observer. As in Experiment 1, the HRCs were selected after a previous visual verification of the stimuli by experts [120], with the objective of comparing different encoders (uncompressed references, E1, E2, E3), different bitrates with the same encoder (HRC5 and HRC7, HRC18 and HRC20), and different image formats (3D Full HD, SBS, FCC, 2D Full HD). This resulted in the following R_{ORD} matrix. Finally, since an imbalance of the randomization of the stimuli would affect the paired comparison results, restrictions 1 and 2 described above were also considered.

1	2	5
20	7	10
3	16	18

Table 4.4: R_{ORD} matrix for Experiment 2

4.2.1.4. Observers

The detailed information of all observers who participated in the tests at UPM laboratory are listed in Table 4.5. In particular, 32 observers carried out the Experiment 1, while 30 subjects took part in Experiment 2, all of them having normal or corrected-to-normal vision. To assure this, a previous visual screening was carried out to each observer, based on the Snellen chart for visual acuity, Ishihara test for color vision, and Randot test for stereo vision.

4.2.2. Experimental results

In the following paragraphs, the main results from the experiments carried out are presented. Firstly, the BT model has been used to analyze the results from Experiment 1 at UPM, grouping the evaluations provided by the observers for all sources. Thus, perceptual scores and 95% CIs have been computed, which allow the comparison among the different HRCs considered in the test. In particular, a comparison among the different encoders and 3D video formats has been carried out, in addition to an analysis of the preference of 2D or 3D content.

Secondly, the results from Experiment 2 at UPM have been analyzed applying the BT model to the evaluations provided by the observers for each source video, to study the influence of the video content.

Figure 4.2: Comparison of the different encoders and formats

Then, a comparison between the results from the experiments at UPM and IVC has been carried out, analyzing the outcomes that could lead to similar conclusions in both laboratories, and the different factors that could have caused distinct results.

Finally, the validity of the assumption considered in the design of Experiment 1 based on grouping the observers to get complete observations is analyzed, comparing the results from both experiments in both laboratories.

Comparison among encoders and formats for 3D video The BT-scores obtained from the results of Experiment 1 at UPM are shown in Figure 4.2, where the score for HRC16 (E3 encoder in FCC format with 6 Mbps) has been taken as a reference (BT-score set to zero) to compare the evaluated test conditions, since it was the lowest score. As it can be seen, considering the different bitrates for this encoder both in FCC and Full-HD formats, as well as for E1 in Full-HD, there are no statistically significant differences between encoding with 9 Mbps or 12 Mbps, while using 6 Mbps causes perceptible distortions that significantly degrade the QoE. In fact, it seems that approximately 9 Mbps could be a saturation point, since, especially using E1, the perceived quality is similar with encoded and raw 3D Full-HD video.

A different behavior is presented by E2 in SbS format, which does not provide statistically significant differences among the three different bitrates considered in the tests. This may indicate that using more than 6 Mbps to encode SbS content won't assure an improvement of the quality perceived by the viewers, while when using Full-HD or FCC formats, higher bitrates may be needed. In this sense, it is also worth noting that SbS format at 6 Mbps could provide a similar performance than using Full-HD at 12 Mbps, since no significant differences were found among E2-SbS, E1-FullHD, and E3-FullHD. This difficulty in noticing differences between SbS and 3D Full-HD is also reflected in the comparison of the scores obtained for the conditions with uncompressed content (HRC1 and HRC2). This is a very important aspect since, even using time-parallel

Figure 4.3: Comparison among the 2D and 3D test conditions

PC method, differences between 3D Full-HD and SbS are almost imperceptible, so viewing 3D Full-HD in real home-environment viewing conditions may not provide an improvement on the viewers' QoE.

Other important aspect is the low performance of the FCC format, since using the same encoder (E3) provides significant lower quality than Full-HD, probably due to not using spatial filtering in the downsampling process involved in the generation of FCC content. In fact, E3-FCC at 9 Mbps performs similarly (considering CIs) than E3-FullHD at 6 Mbps, which may indicate that it may be necessary to use a bitrate 50% greater for FCC than for 3D Full-HD.

Finally, some other factors that should be mentioned are the low performance of JM18.2 reference software in comparison with the other encoders in 3D Full-HD (it was only compared at 9 Mbps due to the limited number of HRCs that was feasible to evaluate); the better performance of the hardware encoder E1 than the software encoder E3 with 3D Full-HD video, especially significant at 9 Mbps and 6 Mbps; and the performance of the encoder for contribution E0 that provides a similar quality than viewing uncompressed format.

Comparison among 2D and 3D conditions The results obtained from Experiment 1 at UPM for the HRCs related to 2D content are shown in Figure 4.3, together with the reference test conditions and HRCs related with encoded 3D video, to study the effects of stereoscopic visualization.

Firstly, it is worth noting that there are no significant differences between the results obtained for 3D (both Full-HD and SbS) and 2D with uncompressed video. Moreover, there is a general trend indicating that the observers preferred 2D content over 3D, especially significant when the coding quality is low. Although from these observations

Figure 4.4: Comparison among the different contents

it is difficult to know the causes of this preference for monoscopic content, one possible cause may be the size of the screens used in the tests, since, probably, they are not big enough to provide a high-quality depth perception. It is known that the size of the screen has a strong impact on the perceived depth, and usually large screens provide more depth quality and immersiveness, leading to a more complete 3D visual experience [26]. In addition, a poor depth perception together with the fact that 2D content is usually more visually comfortable, could make insufficient the expected added value of 3D content. It is worth noting that similar results have been observed in the analysis presented in subsection 3.3.2 in the previous chapter. Nevertheless, other aspects, such as the influence of the content or the display technology will be addressed in the following paragraphs.

Influence of video content The results obtained from Experiment 2 at UPM are depicted in Figure 4.4. Since the objective was to analyze the influence of the video content on the effects of the evaluated HRCs on the observers' QoE, the BT model was applied over the results for each source video independently. As previously, the BT-score obtained for HRC16 was set to zero in each source video to establish a reference. Due to the fact that in the test there were no direct comparisons between two different source sequences (i.e., the two screens showed for each pair of HRCs the same source), it is not possible to carry out a direct comparison between BT-scores of different contents. Nevertheless, in addition to the effects of the different HRCs for each source, it is possible to compare the differences between the BT-scores of two HRCs for different sources (e.g., observing the difference between the BT-scores of HRC1 and HRC20 for *Barrier*, and the difference between the same HRCs for *Soccer*, and comparing these values).

As it can be observed, sources *Barrier* and *Soccer* show in general lower BT-scores and smaller variations among the considered contents. This may be caused by the fast in-depth motion that both sequences have, with objects moving from the background to the foreground presenting high disparities in some instants. On one hand, this could have attracted the attention of the observers influencing the perception of some artifacts, and, on the other hand, the in-depth motion and high disparities may have caused difficulties in fusing both stereo views or visual comfort, affecting the observers' evaluations.

Apart from this, the main interesting aspects to investigate the content influence may be the following:

- **Formats comparison:** As it can be seen in the figure, none of the contents shows statistically significant differences between 3D Full-HD and SbS in uncompressed format (HRC1 and HRC2, respectively), confirming the difficulty in perceiving differences between both 3D video formats.
- **Comparison between 2D and 3D:** Observing the BT-scores for HRC1 and HRC3, it is shown that there are only slight significant differences between viewing the content in 3D or 2D (in uncompressed format) for the sources: *Barrier*, *Rome*, *Soccer*, and *Umbrella*. In all of them, the 3D version was preferred which may indicate a better 3D quality than the other contents, probably due to a poor or excessive depth in *Castle* and *Tree Branches*, respectively.
- **Encoding aspects:** None of the source videos shows significant differences between 3D Full-HD at 12 Mbps, 3D Full-HD at 6 Mbps, and SbS at 6 Mbps, which may be caused by the difficulty in perceiving resolution losses and coding artifacts in small stereoscopic displays with polarized glasses. It is worth noting that for *Tree Branches*, encoding the sequence with any of the considered configurations drastically distorted the quality (the results for all the encoders are similar) in comparison with the uncompressed versions, due to its spatial complexity. Also, as aforementioned, very small differences could be appreciated in the cases of *Barrier* and *Soccer*. Only in the other three contents, to use 3D FCC format is significantly worse. It is also worth noting that the encoding configurations for 2D content provide similar quality than the raw versions of the source sequences.

Comparison with IVC laboratory The same test was also carried out in the IVC laboratory of the *Université de Nantes*, with the following differences in the test setup:

- **Environment and equipment:** as at UPM, the test environment at IVC was set simulating real home-viewing conditions, as depicted in Figure 4.5. However, some variations were accepted between both laboratories. For example, concerning the furniture, room distribution, and ambient light, which at IVC was changing according to the sun light, as there were two windows behind the screens. Regarding the equipment, instead of using small polarized displays as in UPM, two 46" stereoscopic displays with active shutter glasses were used at IVC, as detailed in Table 4.6. In this case, the synchronization of the two displays and the glasses is a critical issue, so two high-performance PCs running a 3D payout

Figure 4.5: IVC Test Environment

Display	Model	Philips 46PFL9705H
	Size	46"
	Resolution	1920x1080
	Refresh rate	400 Hz
	3D technology	Active shutter glasses
Luminance		Living room condition
Viewing distance		1.72 m
Voting interface		Three buttons (left, right, validate) on a touch screen

Table 4.6: Test equipment at IVC

software, specifically developed for this, task were used. In addition, a customized firmware was used to guarantee the synchronization of the displays with the glasses [120].

- **Observers:** The detailed information of the observers that took part of the tests at IVC is listed in Table 4.7. The main differences with the observers that participated in tests at UPM are related to the balance between males and females and the age distribution.

The results obtained from Experiment 1 for all HRCs at IVC and UPM are shown in Figure 4.6 for comparison. In particular, some of the observations aforementioned with respect to the results obtained at UPM are confirmed at IVC, for instance:

- The difficulty in perceiving the differences between 3D content in Full-HD and SbS formats are ratified comparing the results at IVC for HRC1 and HRC2.
- There are no significant differences between coding 3D Full-HD at 9 Mbps and

Experiment	Exp.1	Exp.2
Number of observers	32	30
Gender (m/f)	17/15	15/15
Age range (average)	20-66 (29)	20-53 (32)

Table 4.7: Observers at IVC

Figure 4.6: Comparison between results from IVC and UPM

12 Mbps, in fact, using 9 Mbps provides similar quality than uncompressed 3D Full-HD video.

- There is no significant differences among the three bitrates used to encode SbS video.
- Encoding SbS at 6 Mbps provides similar performance than encoding 3D Full-HD with 12 Mbps.
- Encoding 3D Full-HD video at 6 Mbps provides similar results than using FCC at 9 Mbps, as the results with the encoder E3 show. Thus, for FCC it may be necessary to use a bitrate 50% greater than for 3D Full-HD.
- In both laboratories the hardware encoder E1 seems to perform better than the software implementation of E3.
- The encoder E0, developed for obtaining high-quality content for contribution, provides similar quality than uncompressed video.

On the other hand, in some other cases the results at both laboratories are different, for example:

- The results obtained for 2D conditions (HRC3, HRC18, HRC19, and HRC20) were much better at UPM, which, apart from implicit factors of the observers, could be caused by the display size and technology. In essence, small stereoscopic displays with polarized glasses provide a lower 3D quality of experience and, probably, a better quality in 2D in comparison with the displays used at IVC, based on active shutter glasses that could introduce negative issues, such as flickering effects.
- The performance of the encoder JM18.2 (HRC17) at UPM is generally worse than in comparison with the other MVC encoder, while at IVC similar results

are obtained for all encoders. Although a possible cause for this issue could be the different display technology used in both laboratories, further analysis is needed.

- At UPM, using the encoder E3 both in 3D Full-HD and FCC formats there were no significant differences between using 12 Mbps or 9 Mbps. However, at IVC, there is statistically significant difference between using those bitrates for FCC format, which may be due to the fact that in a bigger stereoscopic display with active shutter glasses the artifacts introduced by the process of generating the FCC content and the encoding process may be more noticeable.

Concerning the Experiment 2, Figure 4.7 shows the results of both laboratories for each source video, from where the following observations arise:

- Formats comparison: Again, these results confirm the similarity between 3D Full-HD and SbS in uncompressed format since only *Soccer* shows significant difference between HRC1 and HRC2, being SbS preferred to 3D Full-HD.
- Comparison between 2D and 3D: In contrast to the results from UPM, at IVC 3D was preferred to 2D (comparing HRC1 and HRC3) for all sources, although the differences were not statistically significant in the cases of *Rome* and *Soccer*. This confirms the aforementioned statement regarding the best performance of 3D videos at IVC, where a bigger stereoscopic screen was used.
- Encoding aspects: It is worth noting that at IVC the differences among the different encoding configurations seem to be more perceptible than at UPM, probably due to the displays used, since the resolution loss in polarized displays could hide coding artifacts, as also revealed in other subjective tests [122]. For example, there are significant differences between using 3D Full-HD at 12 Mbps (HRC5) and 6 Mbps (HRC7) in the case of *Tree Branches* and *Soccer*, which are the most spatially complex sequences (see Table 3.1 and Table 4.1). In addition, for *Tree Branches* using 3D Full-HD at 12 Mbps is significantly preferred than using SbS at 6 Mbps (HRC10), while for *Soccer* using SbS is preferred than using 3D Full-HD at 6 Mbps, reflecting a different influence of the coding artifacts and the resolution loss depending on the sequence. Moreover, the distortions that appear using 3D FCC format may be also more perceptible at IVC, since this format is significantly worse for the majority of the contents. Finally, also more differences are found with respect to UPM comparing the conditions related to 2D content for uncompressed and encoded videos (HRC3, HRC18, and HRC20).
- Content aspects: It is worth noting that for *Tree Branches* and *Umbrella*, the results for many of the HRCs are significantly better at IVC than at UPM (comparing the differences with the reference HRC16). This could be caused by the size and the loss of resolution of the polarized display that entail a general poorer QoE, since these two sequences present a very high spatial complexity (see Table 3.1 and Table 4.1). Besides, the BT-scores follow a similar trend for the rest of the contents in both laboratories.

Validation of the tests As aforementioned, to make feasible the execution of Experiment 1, the total number of observers that participated was divided into groups of

4.2. Subjective evaluation of 3D coding artifacts

Figure 4.7: Comparison results between IVC and UPM for different contents

Figure 4.8: Monte-Carlo simulation histogram

eight, so each observer did not evaluate all the test conditions, but each group completed a whole observation. This supposed the assumption that the evaluations were independent from the observers, which needed to be verified comparing the results of Experiment 1 and Experiment 2, since the latter is a subset of the former, in which all the observers evaluated all the test conditions.

Therefore, the *Fisher's-mid-p-value* was applied on the raw results of the common set of the two experiments, since the sample size is unbalanced and large. The results of the test provided indications of the validity of the tests, since only one of the eight pairs compared in both experiments showed significant difference in IVC (HRC10-HRC18, $p=0.0403$), and only two in UPM (HRC1-HRC3, $p=0.0058$ and HRC10-HRC20, $p=0.043$). Moreover, specifically the two pairs significantly different at UPM are related to comparisons between 3D and 2D conditions, that as it has been studied in the previous paragraphs may be caused by the size and technology of the displays used. In addition, as shown in the comparison of the results from both laboratories, similar conclusions have been obtained in the main observations of the experiments, especially related to the comparison between formats and among the different encoders.

Nevertheless, in order to determine if the number of statistically different pairs between both experiments is statistically significant, a Monte-Carlo simulation was carried out according to the method proposed in [120]. Thus, in each iteration (1000 trials were computed) all the observers from both experiments were randomly divided into two groups (according to the settings of each experiment), and the *Fisher's-mid-p-value* was run. This way, the histogram of the ratio of significantly different pairs shown in Figure 4.8 was obtained, characterized by an average ratio of 0.096, and a maximum of five significantly different pairs (i.e., a maximum ratio of $5/8=0.625$). The set of small probability events is composed by those events with probability less than 5% (i.e., occurrence less than 50). Therefore, having two or less significant different pairs is not a low probability case, which supports the validity of the assumption.

4.3. Conclusions

Many different alternatives for representing and encoding 3D videos have been developed during the last years, including different options for stereoscopic and multiview formats, as well as for depth-based representations. Nevertheless, the research on coding techniques for multiview and depth-based videos is still ongoing, with the objective of achieving efficient compressions to feed current and future displays to provide a better 3D visual experience than stereoscopic technologies. In the meantime, stereo coding systems have allowed to rapidly introduce 3D video in the consumer market, taking advantage of the existing technologies for conventional video.

In particular, this chapter presented an analysis of different representation formats for S3D video, as well as different encoding configurations, by means of an inter-lab subjective experiment. In essence, these tests were carried out in two different laboratories that were set up simulating realistic home-viewing conditions to obtain more reliable results. In addition, a wide high-quality 3D video database was used and different display technologies were employed in the tests.

Among other results, some outstanding conclusions should be pointed out. Firstly, the difficulty in perceiving differences between 3D Full-HD and SbS formats, reflecting the good performance of the frame-compatible alternative taken to rapidly introduce 3D video technologies in households. Using typical bitrates for video broadcasting (e.g., 6 Mbps), SbS format provides a satisfactory QoE, which is shown by the comparison of the obtained results with this format and the same 3D videos in 3D Full-HD and FCC formats, in uncompressed and encoded configurations.

Furthermore, it is important to mention the possible influence of the display size and technology used for visualizing 3D video, since, as shown by the results from the tests, a poorer 3D QoE could be perceived with small displays. In fact, in many cases, watching the videos in 2D was preferred to 3D. Moreover, the technology of the display could affect in the perception of encoding degradations, as indicated by the results obtained with polarized screens that entail a loss of resolution in comparison with stereoscopic displays with active shutter glasses.

Although the 3D video formats considered in these tests have allowed the introduction of 3D video technology in the consumer market with acceptable performance in some cases, the research on new encoding methods for multiview and depth-based formats is required, in combination with a development in display technologies, to provide a high-quality 3D visual experience to the consumers, fulfilling their requirements and expectations, to achieve the definite establishment of 3D video.

5 Transmission

5.1. Introduction

This chapter deals with the evaluation of the impact of transmission events on the QoE of the end users of 3D video distribution services. Therefore, to introduce the problematic of analyzing and estimating the quality perceived by the viewers, a review of the state-of-the-art is presented, considering both subjective experiments and objective metrics proposed in the literature.

Then, a novel methodology for subjective assessment tests, called CIETI is proposed. Specifically, it is specially designed to evaluate the impact of transmission events that take place in video delivery systems for conventional and 3D video. So, it is based on the imitation of typical home-viewing conditions (e.g., concerning test content, environment, etc.) to obtain representative results, since these systems are addressed to consumers. Thus, this methodology contributes to the current trend of presenting alternatives to traditional standard methodologies for subjective testing, which distance the observers from real-usage cases, and do not always provide reliable results in these applications. In addition, the proposed methodology was designed taking into account that the conclusions extracted from the test should be helpful for the development of video quality monitoring tools, which is one of the main challenges within the field of QoE evaluation.

After presenting the basis of the methodology, its application to different transmission scenarios is presented through three subjective tests. In particular, the first one investigates the impact of typical errors in the distribution of 3D video over IPTV networks. Besides, the other experiments are focused on the evaluation of different strategies to transmit 3D video through adaptive streaming services, such as coding quality changes, discarding frames, asymmetric coding, or switching between 2D and 3D formats. Thus, the results provide some indications of the impact of these events, which may be helpful in the technological development of 3D video delivery solutions and objective metrics for quality estimation.

Furthermore, to test the validity of the proposed methodology, it has been compared with the standardized ACR methodology to evaluate a subset of conditions of the experiment related to 3D over IPTV.

Finally, the conclusions extracted from these works are presented.

5.1.1. Quality evaluation

As described in subsection 2.4.3, the transmission events taking place in video delivery systems could have an important influence on the QoE perceived by the end users. Therefore, to understand the impact of these events on the visual experience of the viewers, subjective assessment tests are crucial, as they provide knowledge about the HVS, information about the performance of the processing techniques applied to broadcasted content, and a basis to develop objective quality metrics (e.g., to be applied in network monitoring).

In fact, numerous subjective studies have been published during the last years in relation with conventional video, especially analyzing the effects of video transmission over packet-based networks. For instance, Lin *et al.* [199] investigated the visibility of video packet losses, considering the coding standards MPEG-2 and H.264/AVC, as well as diverse configurations and patterns of packet losses. Also, Barkowsky *et al.* [5] carried out an interesting study considering also a wide range of test conditions for realistic HDTV transmission scenarios, whose results are freely available for developer purposes. In addition, De Simone *et al.* [35] carried out an interesting analysis of the impact of transmission errors on the quality of video encoded with H.264/AVC considering different channel models and packet loss rates. Besides, other studies have analyzed the impact of other types of transmission errors, such as video freezes by Pastrana-Vidal and Gicquel [138], and audio losses by Hands [63], studying the relation between audio and video quality. Finally, it is also worth noting the studies of the transmission errors in home-viewing conditions, also previously mentioned in subsection 2.5.1, looking for more representative results, such as the one presented by Cermack [17] using consumer surveys, or the one carried out by Staelens *et al.* [174], proposing a novel methodology to carry out subjective tests at households using long test sequences.

Furthermore, some studies have been already published analyzing the effects of novel distribution services based on adaptive streaming (see subsection 2.4.3), which are gaining popularity in the last years. For instance, Thang *et al.* [197] studied the optimum number of coding quality levels that could be used in an adaptive video system by investigating the JND levels existing in the quality range of mobile video content and taking into account the perceptual factors of the HVS. Another approach was presented by Cranley and Murphy [29] considering the incorporation of the effects of framerate and resolution adaptations on the user's perception to obtain the configuration that maximizes the QoE. Also, another interesting study was presented by Robinson *et al.* [156], in which the impact of various patterns of quality variations was studied. Finally, to complete this summary, a recent review of subjective studies concerning HAS carried out by Garcia *et al.* [47] could serve as a good reference.

However, in contrast to the numerous studies of transmission errors in conventional video, fewer works have been carried out concerning 3D video. As an example, it is worth noting the analysis of different transmission errors and scenarios carried out by Strohmeier *et al.* [180]. In addition, Wang *et al.* [212] performed a multi-laboratory experiment investigating the impact on the QoE of various transmission events with simulcast and MVC 3D video. Furthermore, another study was proposed by Yasakethu

et al. [227], where transmission errors were simulated in 3D video broadcasts, in which S3D and V+D videos were asymmetrically encoded using SVC.

Moreover, much more limited are the studies published in relation with the use of adaptive streaming strategies for the transmission of 3D video. It is worth noting the comparative analysis of SVC strategies for adaptive streaming that was carried out by Gurler *et al.* [55]. In particular, different coding configurations of SVC, H.264/AVC and MVC were compared, showing that dynamic rate scaling of only one view provides sufficient rate adaptation capability and better overall compression efficiency compared to scaling both of the views. However, asymmetric coding strategies present some limitations that should be taken into account, since coding configurations causing severe differences between both stereo views should be avoided [112].

Besides, regarding the evaluation of the impact of transmission events using objective metrics a similar trend is observed. Nowadays, many alternatives have been proposed for conventional video, while only a few deal with 3D content. In particular, although the interest in these scenarios is focused on NR metrics, since the access to the original content in real video distribution systems is unfeasible, some works have been carried out using various popular FR metrics to analyze transmission errors [162]. Moreover, it is worth noting the inclusion in the recommendation ITU-T J.341 [88] of the VQuad quality metric, which analyses and combines various perceptual degradation (e.g., blockiness, *tiling*, blurring, and jerkiness) that make it very useful for evaluating transmission errors. Nevertheless, some interesting works have been carried out towards the development of NR metrics, especially to be applied in quality monitoring tools. For instance, one of the most popular metrics used with this purpose is the Media Delivery Index (MDI) [217], although the provided results do not correlate well with the end users' QoE, since it is based in network parameters, such as packet loss rate and packet jitter. Thus, other approaches that provide better estimations of the visual experience of the viewers have been proposed. For example, worth noting are the works by Reibman [151] and Raake [150] developing video quality metrics for monitoring systems. In fact, this problem has been considered in standardization activities by the ITU, resulting in the publication of two monitoring approaches: the recommendation P.1201 that is a packet-based metric [90], and the P.1202, which also allows decoding the signal [91]. Furthermore, in this sense, the work carried out by the VQEG through the MOAVI project should be pointed out.

Regarding 3D video, after the activities concerning subjective experiments carried out during the last years by the research community analyzing the related perceptual issues, the first approaches have been presented for automatically estimating the quality of stereoscopic videos. Intuitively, the first proposals of objective quality metrics for 3D video were based on the use of metrics originally developed for 2D video. For instance, Hewage *et al.* [67] proposed the employment of 2D quality metrics in each one of the stereo views or V+D videos. Also, Lebreton *et al.* [115] applied standard quality metrics for monoscopic video to 3DTV sequences impaired with coding and transmission errors. Similarly, Politis *et al.* [147] proposed a 3D metric based on the VQM developed for 2D video. One step forward was given with the appearance of quality metrics especially developed for S3D content, as in the algorithm proposed by

Seo *et al.* [164] that considers blocking artifacts, blurring in edge regions, and video quality difference between the stereo views.

5.2. Proposed subjective methodology

As aforementioned in subsection 2.5.1, there are recent approaches defending that the evaluation of degradations that may appear in the video distribution chain should be evaluated respecting, as far as possible, real viewing conditions, since this will allow to obtain meaningful, reliable, and representative conclusions about what they would perceive in their real-life experiences. In fact, the ideal manner would be to obtain direct subjective reports from their homes. However, this is obviously not feasible, apart from using surveys [17] or network traces [106], which generally do not provide sufficient information. Thus, the design of subjective assessment tests would be influenced by this objective, mainly in the following aspects:

- The test material should be similar to that usually watched by people in their homes (e.g., movies, sports, news, etc). In addition, the sequences should be longer enough to attract the attention of the observers. This way, as it happens at households, the viewers would be interested on the content and immersed in the watching experience, and not only focused on detecting the impairments, which may alter their perception of certain degradations. In this sense, also the use of video sequences with audio (unlike the sequences commonly used in standard tests) is recommended, since this is the usual way of watching video in real situations.
- The sequences should be shown to the observers following a single stimulus procedure, which means that no unimpaired reference is presented to them to compare with the test video. Showing an explicit reference would distance the observers from real viewing conditions, so they cannot be considered for tests associated to home environment scenarios.
- The evaluation should be carried out in a nearly continuous way, since the effects of transmission errors are highly dependent on the instant when they occur, and they are not stationary. This aspect discards the use of standard methodologies (e.g., ACR) based on the evaluation of short test sequences once they have been entirely watched, during periods without video on the screen. Also, other approaches like SSCQE methodology may be not adequate since a strictly continuous evaluation is not usually required, and it is complicated to implement (e.g., sliders or complex devices to record the observers' evaluations).

These aspects make not appropriate the majority of the international standard methodologies (e.g., the procedures proposed in ITU-R BT.500 [89] or ITU-T P.910 [86]), since they were designed to evaluate the performance of video coding algorithms and, in many cases, some conditions (e.g., the use of short test sequences of approximately ten seconds, without audio, explicit references, standardized test environments, etc.) distance the observers from real viewing situations. Besides, some of these aspects have been

Figure 5.1: Diagram of the structure of the test sequences

Figure 5.2: Examples of frames of the first three segments of the test sequence

taken into account by some researchers in their proposals towards subjective assessment methodologies that bring closer the test conditions to real viewing situations. For instance, the methodology proposed by Staelens *et al.* [174] is based on the use of full-length movies in tests carried out in real-life environments. However, in this case the evaluation of the degradations is provided when the movie is finished (after approximately 120 minutes), which may be an excessive time to remember the impact of each distortion. Also, little data is gathered for such long assessment sessions [144]. Also, the control over the subject and the environment is totally lost with this methodology. Furthermore, the method proposed by Sullivan *et al.* [184], also considered the use of immersive test content with long sequences (i.e., one minute) in test environments simulating typical home-viewing conditions. However, since only one test condition is inserted in each test sequence, the number of degradations to study in a feasible assessment session is very limited.

Therefore, taking this into account, a novel methodology, called *Content-Immersive Evaluation of Transmission Impairments* (CIETI), is proposed to evaluate the impact of typical transmission artifacts. As aforementioned, this methodology is based on the use of long test sequences. Thus, to evaluate different test conditions within the same long continuous stream, the processed video sequence is composed alternating processed and original segments, as depicted in Figure 5.1. In the processed segments the test conditions under study are inserted, while during the original (undistorted) segments the observers can carry out the evaluation of the distortion introduced in the immediate previous segment. To indicate to the observers when and which segment they have to evaluate, the original segments display a number in the right-bottom corner of the screen, as depicted in Figure 5.2. In addition, a first segment is used to indicate to the observers the beginning of the test, and provide an implicit quality reference, thus it is also left unimpaired and marked with a zero. This way, a nearly-continuous evaluation of the quality of the sequence without losing the continuity of the video (as it would happen using a similar method to ACR since the evaluation is done during a period in which the screen turns gray), can be carried out.

In addition, the following aspects can be implemented depending on the specific test design:

- Hidden references: In some cases, no degradations were inserted in the evaluated segments, so they can be considered being hidden references for comparison with other test conditions.
- Duration of test sequence: The total duration of the sequences could be set depending on the particular conditions of the test, taking into account the limited duration of the subjective experiment sessions. Typically, sequences with a duration of, at least, five minutes have been considered.
- Duration of the segments: The duration of both the processed segments and the evaluation segments should be adjusted depending on the considered test conditions and the ratings that the observers should provide. For instance, to evaluate adaptive streaming mechanisms it may be necessary to have long processed segments to show the complete quality variation [196]. Also, for 3D studies it may be required to have a longer evaluation segment so the observers could rate the different perceptual dimensions (i.e., image quality, depth quality, visual comfort, etc.).
- Evaluation scale: Usually, the evaluated test conditions are degradations caused by the video processing chain, so the evaluations were done using the five-grade impairment scale (see Table 2.1) [89]. Nevertheless, other scales could be used depending on the conditions considered in tests. For example, in [196], an extension of the degradation scale was used to include also positive scores, since improvements of quality were considered in a test related to adaptive streaming.
- Collection of the evaluations: For simplicity, in the tests carried out until now, the observers provided their ratings using a questionnaire with numbered boxes where they marked their evaluations for the corresponding test conditions [32]. Although some observers could need to avert their gaze from the screen to write their score, this would not affect the result of the evaluations since during the evaluation period there are no impairments in the test video. However, other methods could be easily applied and should be investigated to obtain less intrusive subjective tests.
- Additional variants: In some cases, to evaluate different test conditions over the same segment could be interesting. Thus, different variants of the processed test video could be generated and evaluated. However, this may entail the requirement of more participants or longer tests.
- 3D video: This methodology is also applicable to the evaluation of the impact of transmission errors on the 3D QoE. In fact, apart from the global quality perceived in a certain processed segment, other factors could be evaluated. For example, in the specific case of using a questionnaire, a method to evaluate also the visual discomfort has been proposed [61]. Specifically, the observers are normally asked to write a cross on the box corresponding to the score of the impairment scale and to the evaluated segment, as depicted in Figure 5.3. Then, in the case that the observed impairment caused them some kind of discomfort

SCALE	1	2
Imperceptible		
Perceptible but not annoying		
Slightly annoying		
Annoying	X	
Very annoying		/

Figure 5.3: Example of questionnaire

or annoying visual difficulties to correctly perceive the 3D effect, they should write a single line instead of a cross. This way it is possible to identify which impairments could cause visual discomfort.

Finally, it is also worth noting, that, apart from the subjective studies carried out considering different scenarios (e.g., IPTV, 3DTV, adaptive streaming, etc.), the design principles of this methodology make it very useful to help in the design of monitoring tools to automatically estimate the perceived quality [140], especially the use of realistic long sequences divided into segments that could correspond to monitoring windows used in this type of mechanisms.

5.3. 3D video over IPTV

This section presents a study of the impact of typical transmission errors in frame-compatible 3D video delivered over IPTV networks. In particular, an extensive study of video and audio distortions considering a broadcasting scenario based on SbS 3D video was carried out. Also, to compare the effects of these degradations with respect to 2D video, the test sequences were evaluated by the observers both in monoscopic and stereoscopic formats. Furthermore, other factors related to the 3D viewing experience were evaluated, like naturalness, sense of presence, and visual discomfort.

The following subsections describe the details of the experimental setup and the analysis of the obtained results.

5.3.1. Subjective test setup

5.3.1.1. Evaluation methodology

Since the objective of the study was to evaluate transmission errors that may appear in real 3DTV delivery systems, the CIETI methodology was used in the assessment test, in order to obtain representative results of what the end users perceive at their homes.

The evaluation tests were carried out organizing different assessment sessions of thirty minutes approximately that started with an explanation of the tests focused on clarifying the objective of the experiments and familiarizing the subjects with the evaluation

methodology. With this aim, a training sequence, different to those used for evaluation, was presented as an example containing the possible transmission errors. Then, the observers evaluated each test sequence generated according to the structure shown in Figure 5.1, using the five-grade degradation scale, shown in Table 2.1.

In addition, in order to measure the performance of 3D content, the subjects also evaluated the naturalness, sense of presence, and degree of visual fatigue for each stereoscopic test sequence. In particular, the five-grade quality scale [89] (see Table 2.1) was used to rate the naturalness and sense of presence, while the fatigue was measured considering the five-grade scale defined by Yano *et al.* [226] (ranged from: 5-”My eyes are not tired” to 1-”My eyes are tired”) for eye tiredness and a similar one for headache. A paper questionnaire was used to collect the evaluations (similar to Figure 5.3), where there was also space available for the observers to freely describe other types of discomfort felt during the tests, like headache or dizziness.

Furthermore, the same test sequences were evaluated by the observers in 2D and 3D, so the impact of the considered transmission errors could be compared. In essence, the stereoscopic version of each test sequence was evaluated after rating the same video in 2D. Also, after the evaluation of the sequences in both formats, the observers were asked to indicate their preference between them.

5.3.1.2. Environment and equipment

The subjective tests were carried out at the UPM, in a laboratory especially fitted-out to satisfy the recommendations given in ITU-R BT.500 [89] for assessments in home environments.

In particular, the display used for presenting the sequences was a 42” Panasonic stereoscopic television, with Full-HD resolution. For visualizing 3D sequences, the observers wore the active shutter glasses associated to this TV set. In addition, the display was configured in standard mode with default settings and without any image processing implemented by the TV, trying to simulate domestic usage conditions. Also, special caution was taken controlling the ambient illumination to avoid disturbing reflections and possible flickering effects using shutter glasses.

Moreover, the subjects were positioned at a viewing distance of three times the height of the screen (3H). Also, the videos used in the assessment sessions were played using a high-performance PC and the *Nvidia 3D Vision Video Player* for displaying both stereoscopic and monoscopic content.

5.3.1.3. Test material

For this test, a typical broadcasting architecture was considered, so the source videos used in the tests were two SbS sequences of five minutes of duration. As previously explained in section 5.2, these source videos were divided into segments of six seconds to introduce the test conditions in alternate segments (see Figure 5.1) and generate the test sequences. The main characteristics of the content used for the test are described

Source	Format	Content	Depth
SRC1	1920x1080p 23.97fps	Movie. Some slow segments with dialogs. Some others with fast camera movement.	In general, the objects are projected within a small range of disparities around the screen, while the majority of the objects of interest have a disparity close to zero.
SRC2	720x576p 25 fps	Documentary. Slow action. Some segments with camera panning. Only music as soundtrack.	Although some objects are projected in front of the screen, the majority of the objects present uncrossed disparity. Generally, the range of disparities is larger than for SRC1.

Table 5.1: Test Sequences **Figure 5.4:** SI vs TI

in Table 5.1. In addition, to provide a better characterization of the temporal and spatial properties of the contents, the SI and TI factors of the original segments in which the test conditions were introduced was computed according to the ITU-T P.910 recommendation [86]. As it can be appreciated in Figure 5.4, a wide range of spatial and temporal complexities have been considered for the test clips. The transmission of the video over a lossy network can cause several distortions affecting video and audio, entailing degradations of the quality perceived by the end users. In particular, as shown in Table 5.2, different types of errors were considered in the test, taking into account the most typical degradations reported by the end users of video distribution systems [17]. Specifically, different patterns of each type of error were considered, which are detailed in Table 5.3, to understand the impact of easily identifiable degradations to help in the development of quality models for monitoring tools.

For video losses, "Min" is used to indicate that only B-frames not used as reference are affected. Also, the percentage of the frame that has been corrupted is shown. It is worth noting that when video packets are lost, the error concealment algorithm of the decoder substitutes the affected macroblocks for the corresponding ones of the previous frame,

Error Type	Description
R	Bitrate drops
F	Framerate drops
E	Video losses producing <i>macroblocking</i> . The losses could affect different fractions of the frames and various lengths were considered
V	Video freezes of different durations
A	Audio losses of different durations
AV	Video freeze combined with audio loss

Table 5.2: Considered transmission errors

Code	Duration	Pattern	% Frame
E1	1	Min	100
E2	1	90%	25
E3	1	90%	50
E4	1	90%	100
E5	3	90%	50
E6	5	90%	50
E7	3	25%	50
E8	5	25%	50
R1	6 s	1/2	-
R2	6 s	1/4	-
F1	6 s	1/2	-
F2	6 s	1/4	-
V1	1 s	-	-
V2	2 s	-	-
V3	6 s	-	-
A1	Min	S	-
A2	0.5 s	S	-
A3	2 s	S	-
A4	6s	S	-
A5	Min	3 losses / 2 s	-
A6	Min	7 losses / 6 s	-
AV	6 s	Audio/Video loss	-

Table 5.3: Specific error patterns considered in the tests

thus the effects of video losses propagates until the next I-frame. Thus, the duration indicates the number of consecutive GOPs, and the pattern represents the percentage of the frames of each GOP affected by packet losses (depending on the position of the lost packet within the stream).

Moreover, concerning bitrate and framerate drops, the pattern represents the reduction with respect to the reference bitrate and framerate of the sequences.

Finally, for audio losses, "Min" indicates that a single audio packet is lost, and, regarding the pattern, "S" indicates a single loss, while in the rest cases it is shown the number of losses per time interval.

Moreover, the test sequences were encoded using typical IPTV configurations and RTP packetization. Thus, the H.264/AVC standard was used, setting a GOP length of 24 frames following the structure IBBBP, and setting all I-frames as Instantaneous Decoder Refresh (IDR) pictures. Also, the reference bitrates used for encoding the videos were 8 Mbps for the Full-HD video SRC1, and 4 Mbps for SRC2.

Another important aspect was the use of four different variants of the test sequences, introducing different degradations (from Table 5.3) of the same type (according to Table 5.2) in the same processed segment of each variant. For example, E2, E3, E4, and V1 could be inserted in the first processed segment of each variant, to analyze the effects of different video losses in the same content. Thus, this way it is possible to investigate the relation between the results for the different realizations of a certain impairment, as well as the results for different impairments in the same content.

5.3.1.4. Observers

The subjective experiment was carried out in various stages, so in total participated 42 observers (7 female, 35 male), all of them having normal visual and stereoscopic acuity, and color vision. The ages of the subjects ranged between 20 and 48 years old, with an average age of 27. In addition, a maximum of four subjects took part in each assessment session.

The results obtained in the subjective tests were screened to discard possible random evaluations analyzing the correlation of each observer votes with the average of the scores provided by the other observers and taking into consideration the scores that the observers provided to the hidden references.

5.3.2. Experimental results

5.3.2.1. Impact of transmission errors

To analyze the impact of the transmission events considered in the tests, the MOS values provided by the observers were grouped according to the degradations described in Table 5.3 for each content and globally (i.e., combining both source videos). The obtained results are shown in Figure 5.5 for video losses, Figure 5.6 for audio losses,

Figure 5.7 for bitrate and framerate drops, and Figure 5.8 for outage events. In the figures, the MOS scores provided by the observers for the hidden references inserted in the test sequences of both contents are represented with dashed lines, in order to provide an anchor for the impact of the considered degradations. In fact, in the present study the absolute MOS values are less important than the relation among the different transmission errors.

Video losses Firstly, regarding the degradations caused by video losses, the following observations can be pointed out from Figure 5.5a:

- When the video losses do not affect I-frames the effects are lighter, as in the case of losing a whole B-frame (E1), whose impact is reduced and nearly imperceptible.
- Generally, losing a greater part of the information corresponding to an I-frame means distorting more the video quality (comparing E2, E3, E4, and V1). The caused effects are especially annoying when the header information is also lost (impairment V1), since the rest of the GOP could not be synchronized and a video freeze occurs, lasting the same as one GOP (one second). Another interesting aspect is the reduced visibility of video packet losses not affecting a large part of the bitstream of I-frames, as showed by the high scores provided to the degradation E2. This is due to the fact that losing the last part of the bitstream of the frame affects the lower part of it, and depending on the content the perceptibility could not be substantial. This observations are clearly reflected in the results for SRC1, while for SRC2 the effect of video losses are less perceptible, which is mainly caused by the lower motion activity of the content (see Figure 5.4), that makes the error concealment of the decoder to hide better *macroblocking* effects.
- Regarding multiple losses (E5 to E8), since only a few patterns were considered, it is difficult to extract clear conclusions. In addition, these impairments are highly dependent on the content and the performance of the error concealment of the decoder. For instance, losing a greater part of the GOP (the loss of the I-frame takes place before -E5 and E6- or after -E7 and E8- within the GOP) could cause more severe degradation depending on the content, as shown by the different results for SRC1 and SRC2. Also, although the losses may affect more consecutive GOPs (comparing E5 vs. E6 and E7 vs. E8) the impact could be very similar. Thus, although these impacts are as annoying as losing one entire GOP (E4) and less annoying than video freezes, further research is needed to find possible ways to model these behaviors.

To further analyze the impact of the considered impairments depending on the specific video content, deeper analysis of the results have been carried out, examining the MOS values provided by the observers to each impairment in each test segment of the sequences, thanks to the use of different variants. To interpret correctly the graphs, the vertical axis indicates the MOS, while in the horizontal plane one axis contains the different impairments of the set (according to Table 5.3), and the other shows different segments of the test sequences where the impairments were introduced (i.e., SRC1-1 is the first segment where the corresponding impairment was inserted in SRC1).

(a) Global results

(b) Single video losses in different segments

(c) Multiple video losses in different segments

Figure 5.5: Impact of video losses

The results shown in Figure 5.5b regarding single video losses support the second observation aforementioned, since losing a great part of the frame causes more annoyance, while losing also the packet header entailing a video freeze is the worst case. Also, as shown by the results for the segments of SRC2, the low motion activity helps the error concealment to hide the effects of video packet losses. Similarly, concerning multiple losses, the results shown in Figure 5.5c generally reveal a similar annoyance among all the considered patterns, but also the content could make these impairments less perceptible as reflected by SRC2-22 and SRC2-15.

Audio losses Secondly, from the results for audio losses shown in Figure 5.6a, two effects could be pointed out:

- The duration of the audio loss (and the consequent silence) is clearly related to the degradation of the audiovisual experience (comparing the impairments A1 to A4).
- Although only a few patterns of periodic losses have been analyzed, it seems that having multiple losses during a certain time interval causes less degradation than having a longer single loss during that interval. Also, the effects of these multiple losses are similar to those caused by continuous losses of similar overall duration (comparing A5 and A6 with A2 and A3, taking into account that a single loss lasted about 200 ms, A5 contains an overall silence of 0.6 s and A6 of 1.5 s). Nevertheless, further investigation with more loss patterns is required.

These two observations are generally supported by the results per segment shown in Figure 5.6b and Figure 5.6c.

Bitrate and framerate reductions Concerning the impact of bitrate and framerate drops, Figure 5.7a shows that decreasing the framerate is significantly more annoying than coding quality reductions. In addition, the expected result that a greater reduction of the bitrate entails a worse quality, is especially reflected with SRC1. However, due to the content characteristics of SRC2 (lower resolution, low motion activity, etc.) the coding quality reductions are hardly perceptible by the observers, which also influence the global results in the comparison between degradations R1 and R2. Although the considered bitrates for encoding the sequences are representative of normal 3DTV broadcasting scenarios, they may be too high so that after the bitrate drops the quality is close to the original content.

In fact, the influence of the video content in the impact of these degradations is also reflected in Figure 5.7b, where the MOS scores provided by the observers for some segments of both test sequences are shown. Although the predictable impact of these degradations is shown in SRC1-5, in the rest of the cases the temporal and spatial properties of the content could hide the degradations caused by coding quality reductions (SRC1-3, SRC2-3, and SRC2-5), and even those caused by framerate reductions, as it happens in SRC2-3, where a static view of a landscape is visualized.

(a) Global results

(b) Single audio losses in different segments

(c) Multiple audio losses in different segments

Figure 5.6: Impact of audio losses

(a) Global results

(b) Bitrate and framerate drops in different segments

Figure 5.7: Impact of bitrate and framerate reductions

Outage events Regarding the tested outage events, again, the longer the impairment (silence or audio freeze) the worse the degradation, as shown in Figure 5.8a. Although the statistical significance of the differences among the considered impairments is not shown with the current analysis, valid conclusions could be extracted due to two factors:

- Firstly, this type of degradations are very noticeable (they have durations longer than one second) and extremely annoying (even in the cases of the shortest impairments).
- Secondly, fewer impairments of these types were inserted in the test sequences, since their effects were more expected than in the rest of the cases.

Analyzing the results in individual segments from Figure 5.8b, the influence of the particular video content on the impact of the considered degradation is reflected. As it can be observed, the effects of an audio loss and a video stop of the same duration depends on the information carried on the video or the soundtrack (having dialogs as in SRC1-13 or only music in SRC2-13).

(a) Global results

(b) Impact of outage events in different segments

Figure 5.8: Impact of outage events

5.3.2.2. Comparison between 3D and 2D

A comparison of the results obtained for monoscopic and stereoscopic visualizations of both contents is shown in Figures 5.9-5.12, for video losses, audio losses, bitrate and framerate drops, and outage events, respectively. As it can be seen, very similar results were obtained in all cases, although slightly better scores were obtained for 2D video with the most severe *macroblocking* effects, while better results were obtained for 3D content in the cases of bitrate drops and video freezes.

To analyze more extensively the impact of the perceptual effects (i.e., *macroblocking*, coding quality drops, silence events, etc.) caused by the considered degradations, the evaluations provided by the observers were grouped according to Table 5.2. Then, a more exhaustive statistical analysis was carried out to compare the results for 2D and 3D content. Firstly, the rating results provided by the observers for each error were grouped in one sample for monoscopic video and another for 3D video (two samples for each distortion). Then, the Kolmogorov-Smirnov test [25] was applied to find out that the samples were not normally distributed at the 5% significance level. Thus,

Figure 5.9: Impact of video losses in 2D and 3D

Figure 5.10: Impact of audio losses in 2D and 3D

non-parametric methods for comparing each two samples were considered. Specifically, the Wilcoxon Signed-Rank test [25] was used, and the results are shown in Table 5.4.

One of the main aspects to analyze is the effect of video losses (E) in frame-compatible 3DTV, where the broadcasting system for conventional video is used. This implies that the frame containing the stereo views is encoded using H.264/AVC, and when a video packet is lost, the error concealment technique of the decoder acts without taking into account that it is S3D video. This fact may cause that the blocking effects caused by video losses appear in different regions of each stereo view. Therefore, very different content could be shown in corresponding regions of each view (as depicted in Figure 2.23). Then, when the 3D video is displayed, those content differences could difficult the fusion carried out by the HVS of both stereo views, causing binocular

Figure 5.11: Impact of bitrate and framerate reductions in 2D and 3D

Figure 5.12: Impact of outages in 2D and 3D

rivalry and visual discomfort to the observers. This is reflected in the better MOS results obtained for monoscopic video affected by video losses in comparison with 3D video, which have statistical significance at the 1% level. Similar conclusions were reported by Barkowsky *et al.* [6] in a simulcast broadcasting scenario.

In contrast, better results were obtained for 3D video for the rest of the transmission errors. Results from the Wilcoxon tests showed statistical significance for the comparisons related to bitrate drops (R) at 1% level and video freeze (V) at 2% level. This could be attributed to the added value provided by the factors related to the depth perception of 3DTV. However, no statistical significance was found in the comparison of the results for framerate drops (F) and video freeze combined with audio loss (AV). In the first case, this is probably due to the annoyance of the alteration of the natural

Error	MOS-3D	MOS-2D	z-value	p-value
R	4.649	4.309	-3.273	0.0011
F	2.96	2.8	-1.927	0.054
E	3.398	3.502	2.775	0.0055
V	1.892	1.694	-2.403	0.0162
A	3.041	2.975	-1.879	0.0602
AV	1.458	1.208	-	0.1250

Table 5.4: Results of Wilcoxon signed-rank test

video pace, while in the latter, the loss of all the information of the audiovisual content could be the cause. In the case of audio losses, although the results are near to the statistical significance, it is difficult to get conclusions of the different effects of audio losses evaluating 2D and 3D video content, so further investigation is needed [144].

Finally, comparing the MOS of the different transmission errors analyzed, it can be seen that a similar trend is obtained for conventional video and 3DTV. As it can be expected, the most annoying distortion is the loss of audio together with video freeze (AV), as the observers miss all the multimedia information. Analyzing the effects of the loss of video and audio packets, it can be seen that the audio losses (A) could be more annoying than video losses (E), when at least some video content is correctly displayed. However, when the video losses cause video freeze (V) the subjective quality is worse in comparison to that perceived during audio losses (A).

5.3.2.3. Other results

3D performance The performance of the 3D video system considered in the experiments is analyzed from the results of the evaluation of naturalness, sense of presence, and visual discomfort. In Figure 5.13 the obtained MOS values for both source videos are presented. The results show that these factors are highly dependent on the video content and the creation of the video. As it can be seen, the SRC1 satisfies better the observers expectations for 3D video, obtaining greater values of naturalness and sense of presence.

A critical factor related to the experience of viewing 3D video which is being extensively studied is the visual discomfort felt by the observers [6, 226]. Figure 5.13 also shows the results obtained from the evaluations of headache and visual fatigue as explained in the methodology description in subsection 5.3.1.1 (higher scores indicate less annoyance), revealing that SRC2 could cause more visual discomfort than SRC1, probably due to a worse 3D content production.

In addition, these scores could be also influenced by the time that the observers spent in the evaluation of the tests sequences, since they were provided after the complete visualization of them. In fact, for both source videos more than half the subjects felt visual fatigue in some degree. Furthermore, a 15% of the observers felt headache or slight dizziness after seeing any of the 3D source sequences. Finally, some observers

Figure 5.13: Results for 3D performance

Source	Preference (%)	
	3D	2D
1	50	50
2	21.05	78.95

Table 5.5: Results for preference between 2D and 3D presentations

reported that other factors were also annoying, like the reduction of luminance while seeing 3D video wearing shutter glasses, or a very high degree of scene activity.

Preference between conventional and 3D video After seeing the same sequence in monoscopic and stereoscopic formats, the observers were asked to indicate their preference between both. The obtained results are summarized in Table 5.5. While the same number of observers preferred 2D and 3D content for SRC1, in contrast, the majority of the observers preferred the monoscopic version of video SRC2. Various factors influence the decisions of the observers, like the content of the sequences and a proper creation of the video. The results are in concordance to those showed in Figure 5.13, and indicate that the viewers will only switch to 3D technology when an added value with respect to viewing conventional video is provided without causing discomfort.

5.4. Adaptive streaming of 3D video

This section presents the studies carried out to analyze the effects of adaptive streaming strategies in 3D video. In particular, two subjective tests were performed.

Firstly, in Experiment 1 a comparison of the perceptual effects on the end users' QoE of different strategies was done, considering in particular four adaptive streaming tech-

niques: framerate reductions, bitrate reductions, asymmetric coding, and switching between 3D and 2D.

On the other hand, Experiment 2 was focused on the study of the effects on the QoE of various bitrate adaptation patterns used in adaptive streaming scenarios under varying network conditions.

Therefore, the following subsections present the details of the test setups used in both experiments, as well as the analysis of the obtained results.

5.4.1. Subjective Experiment 1

In this experiment, a comparison of the perceptual effects on the end users' QoE of different strategies was done by means of a subjective assessment test according to settings described in the following subsections. In particular, the considered strategies are based on framerate reductions, bitrate reductions, asymmetric coding, and switching between 3D and 2D.

5.4.1.1. Subjective test setup

Evaluation methodology Since the objective of the study was to evaluate the impact of degradations caused by adaptive streaming strategies in 3D video delivery scenarios, the CIETI methodology was used in the tests (described in the previous section 5.2), to respect, as far as possible, real-viewing conditions.

Due to the fact that only possible degradations of the QoE were considered, the five-grade degradation scale (see Table 2.1) [89] was used by the observers to provide their ratings. Also, to collect them, the observers had to mark their scores in questionnaires with boxes corresponding to each segment composing the test sequences. In particular, the observers were asked to write a cross in the box corresponding to the score of the impairment scale and to the evaluated segment.

Also, in the case that the observed impairment caused them some kind of discomfort or annoying visual difficulties to correctly perceive the 3D effect, they should write a single line instead of a cross, as depicted in Figure 5.3. This way it was possible to identify which impairments could cause visual discomfort.

Thus, the assessment sessions were based on the evaluation of the test sequences after a detailed explanation of the purpose of the test and the evaluation methodology.

Environment and equipment The experiments were carried out in a room at the UPM with controlled ambient light. The sequences were displayed in a stereoscopic 42" Full-HD Panasonic television set (with default settings), and the observers used the corresponding active shutter glasses of the set. According to the recommendations of the manufacturer, the viewing distance was set to three times the height of the screen (3H).

Source	Format	Content	Depth
SRC1	1920x1080p 24 fps	Movie containing segments with slow motion and dialogs, and others with fast motion activity.	In general, the objects are projected within a small range of disparities around the screen, while the majority of the objects of interest have a disparity close to zero.
SRC2	1280x720p 50 fps	Football match with commentaries as soundtrack.	Generally, the video present uncrossed disparity. Only some effects (e.g., score, time, etc.) are projected in front of the screen.

Table 5.6: Test sequences

Figure 5.14: SI vs TI

In order to assure a smooth visualization of the videos, a high-performance PC was used to playback the test sequences. Since SbS stereoscopic sequences were used in the tests, as detailed in the following subsection, the *VLC* player was employed and the functionality to show the content in 3D of the TV set was used, specifying the input format of the test sequences.

Test material In concordance with the objectives of the tests, two SbS videos were selected with contents like those typically watched in real domestic situations, whose main properties and a description of the depth characteristics of the videos is provided Table 5.6. In essence, the content of SRC1 was quite diverse, having dialog scenes and action intervals with high motion activity, while SRC2 was a football match. Also, the spatial and temporal characterization of the segments in which test conditions were inserted is depicted in Figure 5.14, by means of the SI and TI factors [86].

In particular, from these sources and according to the CIETI methodology described in section 5.2, two test sequences were created with a total duration of five minutes and segments of six seconds, so the test conditions were randomly introduced in alternate segments.

Code	Description
R1	Bitrate drop of both views to the half of the reference bitrate.
R2	Bitrate drop of both views to a quarter of the reference bitrate.
F1	Framerate drop to the half of the reference.
F2	Framerate drop to a quarter of the reference.
D1	Switch to 2D coded with the half of the reference.
D2	Switch to 2D coded with a quarter of the reference.
AS1	Asymmetric coding. Left view encoded with 3/8 of the reference bitrate and right view with 1/8.
AS2	Asymmetric coding. Left view encoded with 3/16 of the reference bitrate and right view with 1/16.
CS1	Video freeze of 2 seconds and skip.
CS2	Video freeze of 4 seconds and skip.
CC1	Video freeze of 2 seconds and continue.
CC2	Video freeze of 4 seconds and continue.

Table 5.7: Considered strategy patterns

Also, the sequences have been encoded using the standard H.264/AVC, with a reference bitrate of 12 Mbps and a coding configuration based on closed GOPs of approximately one second of duration and IBBBP structure.

Taking into account the existing adaptive streaming strategies for conventional video and the approaches for transmitting 3D content, four mechanisms, described in subsection 2.4.3, were considered in the test. These techniques allow the creation of versions of a video sequence with lower bitrate to be transmitted in cases when the QoS of the network is reduced, with the objective of affecting as less as possible the QoE perceived by the end users. In particular, the investigated strategies for 3D video were: *denting* (i.e., discarding frames), changing the coding quality, asymmetric coding of the stereo views, and switching to 2D video.

Additionally to these strategies, *buffering* events were also evaluated in the tests, considering both cases of maintaining the end-to-end latency (i.e., freeze and skip) and maintaining the playback stream (i.e., freeze and continue). Evaluating these artifacts may be helpful, not only to measure their impact on the end users' QoE, but also in the design of adaptive streaming systems, in aspects like the length of the chunks.

Therefore, various patterns of these events were considered in the test, which are detailed in Table 5.7. Specifically, two situations of bitrate drops were considered: one to the half of the reference bitrate (i.e., 6 Mbps) and other to a quarter of the reference (i.e., 3 Mbps). Therefore, two patterns for each mechanism were simulated corresponding to these two cases to make an equitable comparison. Also, different durations of the video freezes for the two playback possibilities were considered.

Figure 5.15: Global results for the considered strategies

Besides, as in the subjective experiment regarding 3D video over IPTV, four variants of the test sequences were created to analyze the influence of the content on the impact of the considered events.

Observers A total of 57 observers (13 females and 44 males) participated in various test campaigns, recruited among the students and personal of the university. Their ages ranged between 20 and 44, with an average age of 24. All of the observers reported normal color vision, and visual and stereoscopic acuity. In addition, a maximum of five subjects participate in each session.

The ratings provided by the subjects were screened to detect possible random evaluations, or misunderstandings with the evaluation procedures. This was done controlling the scores provided to the hidden references, and discarding those observers that did not provide high values [86].

5.4.1.2. Experimental results

The evaluations collected from the observers that participated in the subjective test were processed to obtain the MOS and the 95% CIs of the values. The obtained results are presented in the following figures.

Performance of the considered adaptive streaming strategies Specifically, in Figure 5.15 a comparison of the subjective ratings obtained for the considered adaptive streaming strategies for both test videos are shown (represented with bars), together with the global results grouping the evaluations for both contents (represented with a line). In addition, the dashed lines represent the MOS scores provided by the observers for the hidden references of each content, which are useful to compare the perceptual

impact of the degradations caused by the considered mechanisms for adaptive streaming.

As it can be seen, in general, when more severe bitrate reductions are applied in the considered mechanisms the consequent degradations are more annoying for the observers (comparing the results for the strategies ending in “1” and “2”). However, some exceptions are worth noting. Firstly, for both contents, the bitrate reductions obtained switching between 3D and 2D have a very similar impact, as reflected comparing the results for D1 and D2. The reason behind this could be that, although the coding bitrate was considerably reduced, it was high enough to provide a satisfactory QoE for 2D content. This fact contrasts with the results obtained for coding bitrate reductions in 3D (R1 and R2), since the degradations are much more annoying and could also provoke difficulties in the observers’ stereoscopic perception.

Secondly, the other exception is observed in the results of applying asymmetric coding to SRC2, since a very similar impact was perceived with AS1 and AS2, in contrast to the results obtained for SRC1. This may be caused by the specific properties of the content, since, in general, SRC2 present less motion activity and wider uniform regions (e.g., the grass of the field), which could cause a less annoying impact of the coding degradations that appear with this strategy.

Observing the global results it is possible to compare the performance of the different adaptive streaming strategies considered in the study, and analyze which of the strategies could be preferred from the end user point of view. As it can be seen, switching to 2D and reducing the bitrate by symmetric coding of the views are the best options when the bitrate drop in the network is not too severe. However, this latter option does not provide a good performance in the case of a reduction to a quarter of the original bitrate. In this situation, the best option is to transmit a monoscopic version of the video with lower coding quality than the original, since as the results show, the annoyance of the transition from 3D is limited. Nevertheless, it is obvious that this effect could be worse depending on the disparity and 3D perception related to the visualized sequence. In comparison to this case, the severe bitrate reductions applied to the right view in the case of asymmetric coding (AS1) causes a worse performance of this strategy.

Finally, regarding the results for the cases of reducing the framerate (F1 and F2) it is not reliable to analyze the global results, since both test videos have different framerates. Nevertheless, observing the MOS obtained for both contents, it is evidenced that the option of discarding frames produces a framerate reduction which is very annoying, since the natural flow of the video is corrupted. In this sense, it is also worth noting, that in the case of SRC2 the annoyance is significantly reduced, due to the fact that the original framerate of the sequence is 50 fps. In this case, the performance of this mechanism could be as good as those provided by switching to 2D or reducing the bitrate similarly for both views.

Impact of buffering events Apart from the adaptive streaming strategies, the results obtained for the analyzed video freeze patterns are shown in Figure 5.16. Although

Figure 5.16: Results for video freeze events

the overlap of some confidence intervals of the values corresponding to the two video sources could indicate that the differences are not significant, the important comparison is mainly between global results for the two strategies and a certain duration. As expected, the longer the duration of the freeze, the worse the degradation of the users QoE is. However, in the case of skipping frames after the video freeze, the impact of both durations considered in the tests is similarly annoying. Therefore, the option of continuing after the video freeze from the last displayed frame keeping the continuity of the video is preferred. However, these aspects should be further studied, since in the case of streaming live events the increase of the end-to-end latency could be a critical factor. For example, in the case of a football match, it could happen that the viewer sees a goal with a certain delay respect to the reality. On the other hand, when prerecorded content is streamed, this option could be the best for the user visual experience.

Influence of video content To further analyze the impact of the considered strategies with regard to the video content, Figures 5.17-5.19 show the MOS scores obtained for different segments of the generated test sequences. Thus, the impact of the different strategies applied over the same video content (since four different variants were used) could be compared.

In particular, Figure 5.17a shows the results of the considered mechanisms applying a bitrate reduction to the half of the reference, which confirm the aforementioned observations regarding the better performance of switching to 2D and reducing the coding quality, as well as the reduced impact of a framerate reduction in SRC2. Similarly, Figure 5.17b, shows the results obtained applying bitrate reductions to the quarter of the reference, still reflecting a good performance switching to 2D and severe annoyance with asymmetric coding and framerate reductions.

In addition, the good results obtained switching to 2D are reinforced with the scores shown in Figure 5.18, where the effects of this strategy are compared with asymmetric

(a) Reductions to 50%

(b) Reductions to 25%

Figure 5.17: Comparison of bitrate reductions per segment

Figure 5.18: Comparison between asymmetric coding and switching to 2D per segment

Figure 5.19: Comparison of freezing events per segment

Figure 5.20: Visual difficulties for each type of degradation

coding. As it could be seen, it may be preferable to lose the 3D effect than watching stereoscopic content with poor quality.

Finally, Figure 5.19 shows the results for different segments of both sequences with frame freeze events. Although they are dependent on the motion activity of the specific content, in general frame freeze critically degrades the viewers' QoE. Also, as aforementioned, taking into account the conditions of the test, it is better not to skip frames after the freeze event, although the impact of this effect should be investigated in more realistic situations, since it could be crucial in live streaming scenarios.

Visual discomfort In addition to these results, the observers were asked to write a single line in the questionnaire if they experienced vision difficulties with the evaluated segment, as explained previously in the setup. Thus, Figure 5.20 shows the percentage of the observations that indicated visual difficulties for each type of the considered degradations (both reduction patterns have been grouped for each degradation). As it can be seen, asymmetric coding is the mechanism that causes more visual difficulties, since in some cases the differences between both views may cause binocular rivalry. This effect may also appear in cases where symmetric coding is applied with very poor quality. Also, the discontinuities caused by framerate effects also could cause visual discomfort, while the results obtained for switching to 2D may reflect more the perception of the change than a visual annoyance.

In addition, after completing the tests the observers were asked to indicate whether they felt visual fatigue resulting that 42% of the observers reported slight headache or ocular strain.

5.4.2. Subjective Experiment 2

In contrast to Experiment 1, in this test mainly bitrate changes produced by coding quality variations were considered, together with switching between 3D and 2D

Value	Description
+4	Very satisfactory
+3	Satisfactory
+2	Slightly satisfactory
+1	Perceptible but not satisfactory
0	Imperceptible
-1	Perceptible but not annoying
-2	Slightly annoying
-3	Annoying
-4	Very annoying

Table 5.8: Evaluation scale

content. In addition, not only degradations events were used as test conditions, but also situations where the bitrate increases were examined, as detailed explained in the following subsections.

5.4.2.1. Subjective test setup

Evaluation methodology As in the previous experiment, the CIETI methodology was used in the test. However, differently as described in section 5.2, in this case, since quality degradations and improvements were considered, as exposed in the following, the traditional five-grade degradation scale [89] was not appropriate. Thus, an extended scale was proposed in which four levels indicating a satisfactory experience or a positive change were added, as shown in Table 5.8. So, the observers were asked to visualize and evaluate the whole processed segment, taking into account the possible changes between the beginning and the end of the segment.

The duration of the test sessions was limited to approximately thirty minutes to avoid fatigue of the observers. The sessions started with a careful description of the test and the methodology, paying especial attention on the use of the evaluation scale. With this objective, a set of training sequences were shown to the observers to familiarize them with the dynamics of the test and the range of video quality levels considered.

Then, the observers carried out the evaluation of the test sequences, indicating their ratings in paper questionnaires, like Figure 5.3. In addition, similarly as in Experiment 1, the observers were asked to write a cross in the corresponding box of the evaluated segment, except in the cases when they felt some kind of visual discomfort that they had to write just one line (e.g., “/”). Finally, after evaluating all the test sequences, the observers were asked to rank the test sequences according to their global watching experience.

Environment and equipment This test was also carried out at the UPM in the same laboratory than Experiment 1. Thus, the 42” Full-HD Panasonic television set (with default settings) and the corresponding active shutter glasses were used. Also, to

ID	Type of content	Format	Properties	Depth
SRC1	Movie	1920x1080p 24fps	Sudden motion, fast changing the relatively dark scenes. Dialogs and ambient sound as audio.	In general, the objects are projected within a small range of disparities around the screen, while the majority of the objects of interest have a disparity close to zero.
SRC2	Animation	1920x1080p 24fps	Smooth motion of objects is dominant, static background. Dialogs and ambient sound as audio.	Generally, the elements are projected with uncrossed disparity. The video presents a reduced depth range around the screen.
SRC3	Sport	1280x720p 50fps	Football, average motion, wide angle camera sequences with uniform camera panning. Commentaries as audio.	Generally, the video present uncrossed disparity. Only some effects (e.g., score, time, etc.) are projected in front of the screen.
SRC4	Documentary	1920x1080p 24fps	Average motion of sea animals in significantly dark scene. Commentaries and music as audio.	Along the video the objects could be presented in a wide range of disparities around the screen, even with large crossed disparity.

Table 5.9: Test sequences

playback the test videos, a high-performance computer was used running *VLC* player. So, as in Experiment 1, the built-in feature of the TV was used to visualize the S3D sequences.

In addition, the viewing distance was set to approximately three times the height of the screen ($3H$), and the ambient light was carefully controlled to avoid disturbing reflections and possible flickering effects that may appear when using active shutter glasses to watch 3D content.

Test material The video sources considered for the test were selected among typical contents that people usually watch at their homes. Thus, four video sequences were chosen, whose main properties are summarized in Table 5.9, together with a description of their depth properties. In addition, the spatial-temporal characterization of the segments used to introduce test conditions is shown in Figure 5.21, by means of the SI and TI indicators [86].

Regarding the test conditions, four quality levels were considered in the tests, as shown

Figure 5.21: SI vs TI

Code	Q1	Q2	Q3	Q4
Bitrate (Mbps)	2	4	6	8

Table 5.10: Quality levels

in Table 5.10. The coding of the sequences was done using the standard H.264/AVC and closed GOPs of one second. Also, three types of behavior of the adaptive streaming strategies under variable network conditions were considered: increasing, decreasing, and oscillation of the video bitrate. Additionally, different patterns for each of those behaviors were investigated, as described in Table 5.11. As it can be seen, for increasing and decreasing the bitrate, gradual and rapid coding quality changes were considered, together with abrupt changes after *buffering* effects, and changes between monoscopic and stereoscopic presentations. As in Experiment 1, after the buffering, the video can continue without dropping the following frames (i.e., increasing the end-to-end latency) or it can continue with the frame corresponding to the actual instant, dropping those frames that should have been presented during the freeze (i.e., maintaining the end-to-end latency). Besides, regarding bitrate oscillations, different patterns have been considered. Finally, some anchor conditions were used for each quality level shown in Table 5.10, so no bitrate changes were applied in them.

Also, all sources were stereoscopic SbS videos. However, to compare the effects of the test conditions in 2D and 3D video, monoscopic versions of these sources were also used. In particular, they were obtained taking the left view of the SbS. This entails an approximation that each view allocates half of the global bitrate. Thus, for example, when considering a 2D sequence at 2 Mbps, it was originally encoded in SbS format at 4 Mbps, but only the left view is shown. Furthermore, it is also worth noting that those test conditions marked with asterisk were only applied to S3D content.

Taking this into account, the test sequences, according to the CIETI methodology, were generated dividing five-minute (approx.) clips of the source videos into segments with test conditions and evaluation segments (according to Figure 5.1). In order to properly evaluate the adaptive streaming effects, the test conditions were inserted in segments

Type	Code	Properties	Diagram
Increasing quality	I-GR	Gradually (increasing the quality from Q1 to Q4)
	I-RP	Rapidly (increasing the quality from Q1 to Q4)
	FI	Following by frame freezing (starting with Q1, followed by one chunk of video freezing and later presenting Q4)
	FI-D	Following by frame freezing and frame drop (starting with Q1, following by one chunk of video freezing and frame dropping, and afterward presenting Q4)
	I-23*	Converting from 2D to 3D (starting from Q1 in 2D and later converting to Q4 in 3D)
Decreasing quality	D-GR	Gradually (decreasing the quality from Q4 to Q1)
	D-RP	Rapidly (decreasing the quality from Q4 to Q1)
	FD	Following by frame freezing (starting with Q4, followed by one chunk of video freezing and later presenting Q1)
	FD-D	Following by frame freezing and frame drop (starting with Q4, following by one chunk of video freezing and frame dropping, and afterward presenting Q1)
	D-32*	Converting from 3D to 2D (starting from Q4 in 3D and later converting to Q1 in 2D)
Quality oscillation	OSC1	Periodically changing in small steps (between Q3 and Q2)
	OSC2*	Periodically, changing from 3D to 2D (oscillation between Q3 in 3D and Q2 in 2D)
	OSC3*	Periodically, changing in big steps (oscillation between Q4 and Q2)
	OSC4	Non-periodically changing in small steps (between Q3 and Q2)

Table 5.11: Test conditions

of ten seconds, while the evaluation segments had a duration of six seconds. This way the observers had enough time to carry out their evaluations and the test conditions were longer enough to be visualized (if they were not noticeable was because of perceptual issues and not because they were too short). Thus, the resulting test sequences contained 19 test conditions. Moreover, in order to make possible the implementation of gradual changes, the segments were divided into independent *chunks* (according to the adaptive streaming nomenclature as described in subsection 2.4.3) of two seconds. Finally, to make possible the generation of these type of test sequences with several changes, the evaluation segments (when the observers had to write their scores) were used to adjust the subsequent segments containing test conditions. Thus, the observers were asked to not take into account in their evaluations anything that they could perceive during the evaluation segments. In addition, to avoid the possible influence of these effects on the observers evaluations, the test conditions were inserted in the middle of the corresponding segments, leaving some seconds at the beginning and the end of the segment as “safety margins”. In essence, the following adjustments were applied:

- Presenting 2D and 3D test sequences separately would have resulted in too long assessment sessions. Therefore, the test sequences were generated inserting both monoscopic and stereoscopic segments. To facilitate the visualization of the test sequences, the transitions between monoscopic and stereoscopic segments were done during the evaluation segments (except in the cases of the test conditions concerning these transitions). This way, the transition was done while the observers were indicating their opinion about the previous segment in the questionnaire, so they did not perceive big changes, and the perceptual adaptation of the HVS was easier.
- To minimize the possible annoyance that several changes may cause to the observers, test conditions based on increasing the bitrate were followed by test conditions with bitrate reductions (obviously, with an evaluation segment in the middle of both). When this was not possible due to the distribution of all test conditions in the test sequence, test conditions with similar quality were placed subsequently, and the slight transition between both qualities was done smoothly during the evaluation segment between them.

Observers A total of 45 observers (9 females and 36 males) participated in the tests, all of them with normal or corrected-to-normal vision. Also, their ages spanned between 21 and 49 years, with an average age of 29 years. It is worth noting that 45% of the observers have participated previously in other subjective tests, so they can be considered more experts than the rest of them, and that 68% of the observers already watched 3D videos more than three times. Also, the maximum number of participants in the same tests session was established to four, although in the majority of the sessions only three observers participated simultaneously.

Prior to any analysis, the scores provided by the observers were screened to discard possible random evaluations according to the recommendation ITU-R BT.500 [89], resulting in the acceptance of the evaluations of all the participants.

(a) Results for each content and globally

(b) Comparison between 2D vs. 3D

Figure 5.22: Results for decreasing the quality

5.4.2.2. Results

The MOS values obtained from the scores provided by the observers for the test conditions described in Table 5.11, are analyzed in the following paragraphs. In particular, the MOS were computed for each content individually, and globally grouping the scores for all contents. Also, the 95% CIs, calculated according to ITU-R BT.500 [89], are represented.

Decreasing the quality The results obtained for those events containing a bitrate reduction are shown in Figure 5.22a, generally reporting degradation on the the QoE,

as expected. As the global results reflect, there is no statistically significant difference between reducing the coding quality abruptly or gradually, while switching to 2D to save bandwidth seems to be the best alternative (which was also shown by the results of Experiment 1 in subsection 5.4.1.2).

Going in depth investigating the effects for each content, the results show that bitrate reductions were more annoying for SRC1 (movie) and SRC3 (soccer), since, as shown in Figure 5.21), generally the test segments of these videos present higher spatial-temporal activity, which may provoke more noticeable coding degradations. Also, for SRC4 (documentary) the bitrate reductions were annoying for the observers, which in this case may be caused by the degradation of the 3D perception caused by coding artifacts, since this sequence present a higher depth complexity with objects projected in front of the screen with large disparities. On the other hand, in the case of SRC2 (animation) these reductions were hardly perceptible (even being perceptible but not satisfactory). This could be caused by the encoding of computer graphics sequences, that may provide better quality thanks to the low texture and clear content. Also, reducing to 2D may not degrade drastically the QoE, especially in scenes with high motion activity or big disparities.

Moreover, the results also show similar annoyance of buffering events followed by a bitrate reduction, as much for dropping frames or not. Also, with these test conditions, the QoE of the observers seems to be less degraded with SRC2.

In addition, the effects of these test conditions with 2D and 3D content are compared in Figure 5.22b. As observed, the degradation of the QoE is similar with both video formats, since no significant differences could be reported.

Increasing the quality The results for test conditions related to increasing the quality are shown in Figure 5.23a. As it can be seen, observing the global MOS values, although, as expected, positive results were reported, no significant differences were found among increasing the quality gradually, rapidly, or switching from 2D to 3D. This latter fact cast doubts on the additional value of visualizing 3D content in comparison of having high-quality monoscopic content, although it is highly dependent on the original content.

Also, no significant differences are reflected among the different contents. Nevertheless, in SRC2 (animation) the quality increments seem to be more satisfactory, while in SRC4 (documentary) increasing the quality did not improve the QoE of the observers, probably due to perceptual effects regarding the 3D quality of the original content, since some scenes had excessive disparity.

In addition, lower scores than expected were obtained for these test conditions, which, together with the wide CIs, could reflect some difficulties or reticence experienced by the observers in using the positive range of the proposed scale. This could be also evidenced comparing the obtained scores with those related to the events of decreasing the bitrate (see Figure 5.22a), which span a wider range of values and show smaller CIs. This fact is further investigated in the following, with an analysis of how the observers used the rating scale.

(a) Results for each content and globally

(b) Comparison between 2D vs. 3D

Figure 5.23: Results for increasing the quality

Similarly, regarding those test conditions in which the quality increased after a buffering effect, the negative effect of the frame freeze prevailed over the quality improvement. In addition, although there are no significant differences, dropping frames after the buffering effect seems to degrade more the QoE than continuing with the corresponding frame. However, as concluded from the results of Experiment 1 (see Subsection 5.4.1.2), these test conditions should be analyzed with different and more realistic test setups, since latency is crucial in these events.

To compare the effects of the considered strategies in 2D and 3D content, the results for the test conditions related to increase the quality evaluated in both contents are shown in Figure 5.23b. As it can be observed, no statistically significant differences can be reported increasing gradually or rapidly the quality, similarly as with the comparison

(a) Results for each content and globally

(b) Comparison between 2D vs. 3D

Figure 5.24: Results for quality oscillations

of the considered buffering events. However, this should be further investigated, since buffering events could have annoying effects in 3D video due to the possible appearance of abrupt changes of disparity.

Quality oscillation As shown in Figure 5.24a, the patterns considered for the adaptive streaming strategies in bandwidth oscillation conditions were almost imperceptible in terms of quality for the observers, as reflected by the global results. Analyzing individually the results for each content, the oscillations were more perceptible for SRC1 (movie) and SRC2 (animation). It is also worth noting than switching alternatively

(a) Visual discomfort vs. TC.

(b) Visual discomfort for each content

Figure 5.25: Results regarding visual discomfort

between 2D and 3D formats may not impair more the QoE of the observers than coding quality oscillations.

A similar behavior could be observed with 2D content, as depicted in Figure 5.24b, since the quality oscillation events considered were almost imperceptible in both video formats. An important factor that may be taken into account is the influence of scene changes, since it may hinder the detection of quality oscillations [156]. Nevertheless, more investigations should be done regarding bandwidth oscillation conditions, since only a few cases (both with coding qualities and change patterns) were considered in this test.

Visual discomfort To investigate the visual discomfort that some test conditions could have provoked to the observers, those evaluations marked by the participants with one line (as mentioned in the description of the evaluation methodology in subsection 5.4.2.1) were counted for each considered event. The results are represented in Figure 5.25a showing the percentage of evaluations reporting visual difficulties. Also, the number of the source sequence (according to Table 5.9) with the highest visual discomfort in each test condition is displayed at top of the bars.

Content Test Condition	SRC1 D-GR	SRC1 OSC1	SRC4 D-GR	SRC4 OSC1
MOS with visual discomfort	0.80	2.85	1.63	3.55
MOS without visual discomfort	1.55	3.78	2.2	4.15

Table 5.12: Comparison between MOS of subjects with and without visual discomfort

As it could be observed, the events related to decreasing the coding quality (even followed by buffering events) are the most annoying in terms of visual discomfort, together with periodically switching the quality in small steps (OSC1), since the appearance of coding artifacts may cause binocular conflicts, as the structural information of the images could be severely affected.

In addition, switching between 2D and 3D in oscillating conditions also cause visual discomfort, since the disparity is changed alternatively and may produce perceptual difficulties. Moreover, switching from 2D to 3D to increase the quality (I-23) is also reported of causing visual difficulties, since the change of disparity could be easily perceived.

Regarding the video content, SRC4 (documentary) seems to be the most annoying, probably due to excessive disparities causing difficulties to fuse both stereo views, as well as due to the appearance of several objects projected in front of the screen. This result is also observed in Figure 5.25b, where all the test conditions causing visual discomfort were collected by contents. In particular, SRC4 (documentary) is the content that causes more visual discomfort, followed by SRC1 (movie), which also presents dark scenes and high motion activity (see Figure 5.21). On the other hand, SRC2 (animation) is the most comfortable content, since creating 3D videos by computer graphics present less complications than nature videos, and the disparity could be easily controlled in each sequence to avoid causing visual discomfort.

Furthermore, to investigate the evaluation of visual discomfort with the proposed methodology and its influence on the quality scores provided by the observers, the MOS of the observers reporting visual discomfort is compared with those not indicating it. In particular, the test conditions causing visual discomfort, to at least five observers, were selected. The results are shown in Table 5.12, for the two contents and test conditions causing more visual discomfort. As it can be seen, experiencing visual discomfort significantly influences the overall rating of the observers' QoE.

Analysis of the scale Since a novel rating scale was proposed in this test, its performance was further investigated. Thus, Figure 5.26 shows the distribution of votes for each score of the scale. The main observation that should be pointed out is the lower usage of the positive range of the scale in comparison with the negative. Various factors could influence this behavior, such as the overall 3D quality of the original (undistorted content), which could be not satisfactory enough for the observers.

In addition, probably to have nine levels could entail difficulties to the observers to discern among them. Furthermore, extending the standard impairment scale could

Figure 5.26: Observers' vote distribution

introduce uncertain levels, for instance, it could be difficult to decide whether to rate a slightly perceptible event as “Perceptible but not satisfactory” (+1) or “Perceptible but not annoying” (-1), as it has been shown in the results for oscillating events and decreasing quality with SRC2. Therefore, further investigation is needed to investigate the effects of the rating scale and find the most appropriate scale for this type of tests, following the ideas like the study carried out by Huynh-thu *et al.* [72] .

5.5. Validation of the methodology

In this section, the validity of the CIETI methodology to evaluate the QoE of the end users of 3D video distribution systems is analyzed. With this aim, a subjective test was carried out to compare the proposed methodology with the standardized ACR method. In particular, this subjective test was done in parallel with the Experiment 1 described in section 5.4.1. Thus, a similar test setup was employed, as described in the following subsections, to evaluate, with both methodologies, the impact of various transmission errors in S3D video, as well as other factors related to the 3D visual experience, like visual comfort.

5.5.1. Subjective test setup

5.5.1.1. Evaluation methodology

As aforementioned, both CIETI and ACR methodologies were used to evaluate the same test sequences, in order to compare the results obtained from both alternatives.

Therefore, the evaluation of the test sequences was carried out in different sessions of approximately 30 minutes. At the beginning of each session a detailed explanation of the experiment was provided to the subjects, including the aim of the tests to analyze the impact of transmission errors in 3DTV, the type of errors, and especially, the basis of the assessment methodologies.

Code	Type	Description
R1	Bitrate reduction	Reduction of the coding bitrate to the half.
F1	Bitrate reduction	Reduction of the framerate to the half.
D1	Bitrate reduction	Switch to 2D. Equivalent to reduce the original global bitrate to the half.
AS1	Bitrate reduction	Asymmetric coding equivalent to reduce the original coding bitrate to the half.
E1	Video packet loss	Loss of one whole b-frame.
E2	Video packet loss	Affecting one whole I-frame, affecting 1/3 of the GOP.
E3	Video packet loss	Affecting one whole I-frame, affecting 1 GOP.
V1	Video freeze	Duration of 1s.
CS1	Outage	Duration of 2s. The video continues with the corresponding frame to the actual moment.
CC1	Outage	Duration of 2s. The video continues with the subsequent frame of the frozen image.

Table 5.13: Considered strategy patterns

The rest of the session consisted of, firstly, the presentation of the test sequences to be evaluated with the proposed methodology, and later, the sequences generated to be evaluated following the ACR procedure were displayed.

5.5.1.2. Environment and equipment

In essence, the same test environment and equipment as in Experiment 1, detailed in section 5.4.1 were used in this test.

5.5.1.3. Test material

The same video sources were used in this test as in the Experiment 1 regarding adaptive streaming, so their main properties are shown in Table 5.6. In addition, although some of the considered degradations were described in section 5.4.1 as they were used in Experiment 1, other artifacts related to transmission errors in 3D videos delivered over IP networks (with RTP packetization) were considered in the present study. Thus, all the considered degradations are detailed in Table 5.13.

Specifically, different patterns of video packet losses were considered, causing *macroblocking* effects with different severity. These effects are the result of a typical error concealment technique of the decoder that substitutes the lost macroblocks by the corresponding from previous frames that have been correctly decoded. So, depending on which packet of the stream is lost, the effects could propagate affecting more frames within the GOP. Also, video freezes were taken into account simulating different durations, and in combination with audio loss (outage events). In addition, other typical degradations were analyzed, which are caused by the strategies used by

service providers to reduce the amount of data to transmit when the performance of the network is reduced. These adaptive streaming techniques were *denting* (framerate reduction), reducing the coding quality, switching to 2D, and using asymmetric coding.

Therefore, to generate the test sequences used for the CIETI methodology, the source videos were divided into segments of six seconds and these degradations were randomly introduced in alternate segments, resulting in five-minute test sequences as depicted in Figure 5.1.

On the other hand, in the case of the sequences to be evaluated using ACR methodology, they were composed of impaired segments extracted from the two test sequences created for the other methodology. Thus, two different test sequences were generated (according to Figure 2.33), having both of them the same test segments but in different randomized orders to compensate contextual effects. At the beginning of the sequences a message indicating the start of the test was displayed. In addition, the message in gray background between each impaired segment contained a number indicating the observers in which box of the questionnaire they had to mark the evaluation. The length of the impaired segments and the messages was six seconds, which is the same duration used for test and evaluation segments in the proposed methodology.

5.5.1.4. Observers

A subset of 41 participants from Experiment 1, detailed in section 5.4.1, took part in the evaluation of the test conditions using the ACR methodology.

5.5.2. Experimental results

5.5.2.1. Comparison between ACR and the proposed methodology

The results obtained in the tests were statistically analyzed to carry out the comparison between both assessment methodologies. Specifically, the MOS of the evaluations provided by the observers for each one of the considered degradations were computed. In the case of the ACR evaluations, since each test segment was evaluated twice by each observer, the mean value was taken as the resulting MOS [134].

In Figure 5.27a the results obtained from all the impairments considering the degradations inserted in both source videos are shown, except for the reduction of framerate (F1). Since the source sequences have different framerates, the effects of this degradation should be analyzed separately for each video, as depicted in Figure 5.27b.

As it can be seen in Figure 5.27a, slightly better scores were obtained with the CIETI methodology in the cases of *macroblocking*, which can be caused by the fact that they could be less perceptible, since a long continuous sequence is used, the observers are not focused on detecting the impairments, but on the content of the video.

A similar conclusion could be extracted from the comparison in the cases of coding bitrate reductions (R1 and AS1). In these cases, the higher values obtained with

(a) Results for transmission events

(b) Framerate reduction

Figure 5.27: Comparison of the results with ACR and the proposed method

the CIETI methodology could be also due to the fact that the observers perceive the reduction with respect to the original bitrate as it would happen in real-life situations. On the contrary, with ACR, observers only evaluate the coding quality of the segment without any contextual reference to the previous content of the video.

Interesting results were obtained for video freezes (V1) during one second (duration of a GOP), which could be caused by a packet loss affecting the header information and impeding the decoding of the whole GOP. As it can be seen, the results are also better in the case of the CIETI method, and similar to the scores for a video packet loss causing *macroblocking* propagated along all the GOP. This could be due to the fact that the effects of *macroblocking* in 3D video are very annoying causing binocular rivalry [57].

Concerning the outage degradations causing longer freezes and also the loss of audio (CS1 and CC1), better ratings were obtained with the ACR methodology. This could be due to the severe temporal interruption of a long continuous sequence that the observers are following in the case of the proposed methodology. In ACR, the observers just evaluate an error stop event within a six second segment, with no contextual reference to the rest of the sequence.

Finally, almost similar results were obtained for the degradation based on the switching to a monoscopic version of the 3D video (D1). In fact, it mainly caused annoyance due to the transition itself, more than the loss of 3D effect, which is reflected in the high MOS values.

In Figure 5.27b, the results obtained with the two evaluation methodologies for the framerate reductions (F1) and for each sequence are shown. As it can be seen, very similar results were obtained since the effect of discontinuous transitions between adjacent frames is equally annoying in both cases. The higher framerate of SRC2 causes the better ratings in this case.

To analyze the statistical significance of the differences between the evaluations with ACR and the proposed method, a non-parametric analysis was carried out. This was done because the Kolmogorov-Smirnov test [25] was firstly applied to the sets of obtained values, and it shown that they were not normally distributed at the 5% significance level. Therefore, since the sets of values were not exactly paired, the Mann-Whitney test was applied [25]. As a result, statistical significance was obtained, at the 5% significance level, in the cases of bitrate reduction (R1: $p=0.0148$, $z=-2.438$), and error stop (i.e., video freeze and audio loss) continuing the playback skipping to the frame corresponding to the actual instant (CS1: $p=0.04$, $z=2.054$). Especially the effects of these types of degradations are highly dependent on the contextual reference to the content previously seen. For the other degradations, the fact that there is no statistical significance reflects a similar performance of CIETI and ACR methodologies. It is worth noting that these results are in line of the conclusions extracted by Staelens *et al.* [174] from the comparison of their test method in real-life environments with ACR, since, although fewer error types were considered in that comparison, the obtained results reflected a different perception of certain impairments with each methodology.

In addition, Figure 5.28 shows the scatter plot of the results obtained with ACR and

Figure 5.28: Scatter plot of the results with ACR and CIETI

CIETI methodologies for each content and for the global results grouping both. Moreover, the PLCC for SRC1, SRC2, and global results are: 0.975, 0.882, and 0.958, respectively. These results, together with the outcomes of from the Mann-Whitney analysis, support the validity of the proposed methodology for reliable evaluation of transmission events simulating, as far as possible, real viewing conditions.

5.5.2.2. Visual comfort

The results obtained from the evaluations of visual discomfort are represented in Figure 5.29, where it is shown the percentage of the observers that expressed that they felt visual difficulties within the analyzed degradations. As it can be seen, *macroblocking* effects and asymmetric coding of the stereo views (AS1) are the most annoying impairments in this sense. In the case of *macroblocking* effects it has been shown that these impairments could cause binocular rivalry and, consequently, visual discomfort [57]. Regarding asymmetric coding, when the bitrates used for each stereo view cause a considerable loss of structural information of the images, binocular rivalry could also be produced to the observers.

Another aspect extracted from the tests is that in the case of ACR more observers felt visual discomfort. This could be probably due to the fact that the HVS should adjust the stereoscopic perception in each test segment. In fact, the visualization of the message segments (without depth effect), and the evaluation of test segments with probably very different disparity could cause visual discomfort. On the other hand, with the CIETI methodology, the adjustment of the 3D perception by the visual system is smoothly done during the sequence. However, similar effects could appear in the cases when the observers need to avert their gaze from the screen to write on the questionnaire. Thus, more appropriate ways to recollect that scores are focus of further research.

Figure 5.29: Percentage of observers that felt visual discomfort

5.6. Applications of the results from the subjective tests

The subjective tests presented in this chapter were designed to understand the elements that can affect the visual experience of the end users of distribution video services. This has been the first step in the path to develop objective metrics applicable to these type of streaming scenarios in order to automatically estimate the QoE perceived by the viewers. In fact, various activities have been carried out in this sense.

For example, the subjective test regarding transmission of 2D and 3D video over IPTV networks, presented in section 5.3, helped in the development of an architecture for real-time monitoring of the multimedia QoE [140]. In essence, a lightweight and scalable monitoring architecture called Qualitative Experience Monitoring (QuEM), schematized in Figure 5.30, was proposed, which makes the best use of the information available at network level to provide significant quality data to the service provider. The aim of the metric is detecting the events considered in the previously described subjective tests. In essence, the basic building block of the solution is the Qualitative Impairment Detector (QuID). A QuID is an element which monitors the stream (either before or after decoding) in order to detect a single kind of impairment event. Then, the most adequate model or metric could be applied for each kind of impairment, to map the values of the impairments to severity in the Severity Transfer Function (STF) blocks. Thus, the STFs provide a single severity value each T seconds (in principle, a pooling window of six seconds was proposed, so that the segments of the test sequences used in the subjective tests had also this duration). Finally, the outputs from the STFs are aggregated depending on the interests of the user of this solution (e.g., considering that the aggregated severity is the maximum of all the severities involved). It is worth noting that, although the output of the system could be MOS estimations, other features of the degradations could be reported, which may be more helpful for the service provider and could be compared with the customer opinions.

Figure 5.30: QuEM architecture design [140]

Figure 5.31: Diagram of the proposed rule-based system [62]

Furthermore, although it was originally proposed for monitoring conventional IPTV, it can be applied to any other network delivering multimedia services, such as HAS or 3DTV.

In this sense, another activity was carried out investigating the possibilities of detecting the type of degradation affecting the video quality (QuID in the QuEM architecture) using a rule-based system, prior to applying the most appropriate metric for that distortion (STF in the QuEM architecture). This way, it is possible to combine different reliable quality metrics to achieve robust evaluations of the video quality in scenarios with distinct degradations. Although the idea and final objective is to use NR metrics, a first implementation of the system (depicted in Figure 5.31) analyzed the use of standardized FR metrics [62]. Specifically, VQM [84] and VQuad [88] have been considered for coding and transmission distortions, respectively, and also some features of VQM were used in the detection step. The results showed a good performance of the approach compared with other state-of-the-art metrics, and that the overall performance is better than using the quality metrics individually [62]. Future work will be focused on the improvement of the stage for detecting and identifying impairments, and the study of using no-reference video quality metrics for quality monitoring in video distribution networks.

Furthermore, research work has been carried out studying objective metrics applicable to monitoring schemes, as QuEM, for estimating the impact on the end users' QoE of typical degradations in broadcasting video systems [30, 31], as well as working on a

perceptual model for 3D video applicable to monitoring systems (as QuID) [59] and unequal protection strategies [37].

5.7. Conclusions

After the latest advances in 3D video technology, the video distributors delivered 3D video content to the consumer market taking advantage of the existing infrastructure for conventional video services. Thus, with minor modifications it has been possible to use video distribution systems, such as IPTV and adaptive streaming, to introduce 3D video in households.

Therefore, in this chapter, typical degradations that appear in 3D video distribution systems have been studied through various subjective assessment tests, in order to understand their impact on the visual experience of the end users. In addition, a novel methodology, called CIETI, for subjective testing of the effects of transmission events has been proposed, in order to contribute to the current research activities towards methodologies that aim at obtaining more reliable and representative results of what end users perceive in their homes.

In essence, this methodology tries to set the test conditions simulating, as far as possible, real viewing situations. Thus, the use of a single-stimulus method for evaluating events in a nearly-continuous way is proposed, employing long sequences and contents usually watched by people at homes (e.g., movies, sports, documentaries, etc.). This methodology has been successfully used in various subjective assessment test to evaluate transmission events in IPTV, 3DTV, and adaptive streaming of 2D and 3D video. In addition, the validity of the CIETI methodology has been compared with the standard ACR methodology, obtaining a similar performance in the evaluation of the majority of the considered test conditions, and even providing more representative ratings of the end users' QoE for events in which the contextual reference plays a crucial role (e.g., bitrate variations), so using longer sequences is an advantage. In fact, CIETI methodology seems to be more adequate to evaluate visual comfort than ACR, since using a continuous sequence avoids severe changes in the accommodation of the visual system to correctly perceive the 3D effect. Nevertheless, the improvement of the evaluation methodology will be studied in order facilitate to the observers the way to provide their opinions regarding, not only video quality, but also other factors related to 3D content, such as visual comfort and depth perception.

Moreover, a subjective test was presented in this chapter investigating the effects of transmission errors in 3D video delivery through IPTV networks, such as video packet losses, audio losses, bitrate drops, and outage events. In addition, the impact of these degradations in comparison with 2D content was analyzed. The results mainly showed a similar influence of these impairments in the QoE of the observers, although it is worth noting that the effects of video losses may be more annoying in SbS 3D video, since they produce blocking artifacts which distort left and right views differently, which could cause binocular rivalry and visual discomfort to the viewers.

Furthermore, some strategies for adaptive streaming of 3D video were evaluated in two different experiments, and compared with their performance with 2D video. One of the main results that should be pointed out is that switching from 3D to 2D could be the best option to reduce the bitrate, while the inverse behavior does not provide a significant improvement to the viewers' QoE. In addition, reducing the coding quality could be a good option if the coding artifacts do not degrade so severely the structure of the stereo views, since this could cause visual discomfort. A similar conclusion could be extracted from the results obtained using asymmetric coding, since binocular rivalries could appear when the coding qualities of the left and right views are excessively distinct.

In addition, it is important to take into account in the test design the effects of these degradations in the visual comfort of the observers, as well as the fatigue that this type of subjective tests could cause to them. In essence, apart from the studied degradations that may cause visual difficulties, such as *macroblocking* effects or asymmetric coding, also the source content itself should be appropriately selected, since excessive disparities or high motion activity could cause visual discomfort.

Finally, it is worth noting that the subjective assessment tests that were carried out were designed to help in the development of objective metrics for automatically estimating the end users' QoE, especially in quality monitoring systems, and to provide useful conclusions in the applications of appropriate adaptive streaming strategies.

6 Display

6.1. Introduction

This chapter addresses the quality evaluation of 3D displays, which are the last components of the video processing chain. In the present section, firstly, a review of the state-of-the-art about quality evaluation of 3D displays, using subjective or objective methods, is presented. Then, two subjective experiments are exposed considering current 3D display technologies for consumer environments and future technologies.

On one side, the first study presented in this chapter was carried out to compare the performance of different consumer 3D displays currently available in the consumer market. Specifically, stereoscopic displays with active shutter glasses and polarized glasses were considered, in addition to an autostereoscopic display. The results of the test may provide some outcomes regarding the advantages and drawbacks of each technology and their capabilities to provide a satisfactory QoE to the users.

On the other side, the second subjective test was oriented towards the investigation of SMV displays, which seem to be the most promising glasses-free 3D system, since they offer less technical bottlenecks than other 3D display technologies [185]. Although nowadays only some prototypes are available, it is necessary to start evaluating different technological and perceptual factors that these systems will introduce. Thus, a novel methodology to evaluate these aspects is proposed, using displays currently accessible in the market, which is based on the simulation of the horizontal parallax that SMV provides. Actually, these preliminary approaches are crucial for an appropriate development of this technology. So, this study would help, within the MPEG standardization activities, in the development of models, evaluation methodologies, and coding approaches to be used with real SMV displays in the near future.

Lastly, some conclusions from these studies are given.

6.1.1. Quality evaluation

Several subjective experiments have been carried out over the last years to evaluate technological and physiological aspects related to the QoE of the users of 3D displays. Especially important are those studies concerning visual discomfort or fatigue [110], commonly experienced by the viewers of 3D video and caused by the use of glasses, motion artifacts [108], and accommodation-vergence conflicts [41]. Depth perception is another factor affecting 3D video visualization that is being deeply investigated [113],

as well as distortions introduced by 3D display technologies, like crosstalk between views [208, 228].

Another important issue regarding 3D displays is to subjectively evaluate the performance of the different technologies that are available to watch 3D content, and compare different systems to identify their advantages and drawbacks, to determine the best use cases for each solution. For example, Slanina *et al.* [168] compared the performance of different consumer 3D display technologies: a 32" stereoscopic LCD with passive glasses, a 42" stereoscopic plasma with active shutter glasses, and an 80" projection system also with shutter glasses. Various factors influencing the users' QoE (e.g., depth perception, image sharpness, visual comfort) were evaluated by means of a non-standard methodology designed by the authors and using SbS videos from a proprietary database. The results showed comparable performance among the considered systems regarding depth perception, but differences in image sharpness and visual comfort. Other example is the study carried out by Kaller *et al.* [100] comparing the angular characteristics of a polarization multiplex stereoscopic display and a time multiplex display, concluding that the time multiplex technology is better in terms of image quality and perceived depth. Also, Yun *et al.* [228] carried out a comparison between common stereoscopic displays with active and passive glasses in terms of perceptual resolution and crosstalk, finding out interesting results, such as the lower crosstalk in passive displays. Another interesting study was presented by Rerabek and Ebrahimi [155] comparing three different 3D portable displays that are commercially available, in terms of depth quality and overall quality provided by restitution techniques based on stereo and motion parallax. The results show that the motion parallax method offers a better overall quality, while stereo parallax provides better depth quality. Finally, it is also worth mentioning the study carried out by Coria *et al.* [26] evaluating the influence of the screen size in stereoscopic displays, testing three displays of 2.8", 22", and 55", and reporting that 83% of the observers would not watch 3D content on a mobile device, and that the majority of the observers preferred to watch 3D content on bigger screens.

In addition, some interesting studies have been presented evaluating through subjective tests novel approaches for 3D video visualization, such as HMD [14]. Also, in relation with display technologies still under research, there are also some preliminary studies that may help in the development of these systems, as well as in the understanding of perceptual factors related to the users' visual experiences. For instance, regarding SMV displays, although they are not yet widely available and only a few prototypes have been presented [193], there is a need of evaluating the impact on the viewers' QoE of certain factors, such as horizontal-parallax, smoothness, or visual comfort. In fact, although only a few studies have been already presented, there are some examples in which the smoothness has been evaluated using a head-up display [188], and the visual comfort effects have been simulated [102].

Similarly, as it happens with SMV displays, the existence of few prototypes of other displays, such as holographic systems, cause that only a reduced amount of subjective studies have been carried out evaluating their performance. One example is the study by Takaki and Yokouchi [189] evaluating accommodation issues in comparison with the

observation of real objects.

Moreover, it is worth noting that to obtain a reliable evaluation of the influence of the factors that new 3D displays present, a revision of the traditional test methodologies should be carried out [137], so an active research work is being carried out with this aim, as reflected by the activity led by the VQEG within the 3DTV project [205] concerning the exploration of new methodologies for subjective testing of 3D video QoE.

On the other hand, although the most reliable way to evaluate the performance of display systems is by means of subjective tests, some objective metrics have been already proposed to estimate the impact of some effects and degradations caused by 3D displays. For example, Konrad *et al.* [105] developed a model, dependent on the display system, to assess and reduce crosstalk and, therefore, enhance visual comfort. Another example is the algorithm proposed by Yamanoue [225] to objectively evaluate the puppet-theater effect, described in subsection 2.4.4.

Also, some metrics have been proposed to evaluate the quality of complex displays such as integral imaging systems. For example, Forman *et al.* [46] presented an angle-dependent PSNR metric to estimate image quality.

Finally, valuable studies have been carried out to objectively evaluate the visual discomfort produced by 3D displays based on physiological measures. A detailed overview was presented by Park and Mun [137], going through techniques based on ocular responses, brain responses, photoplethysmography, and nervous system measures (e.g., electrocardiography).

6.2. Comparison of consumer television technologies for 3D display

As previously stated, more studies about the quality perceived by the users of 3D video systems are needed, dealing with new approaches and technologies to achieve performance improvements, and to understand the users' expectations, especially in relation to the capabilities of consumer 3D display technologies. In fact, only a few works have been published regarding these issues, which are extremely important as it has been shown in similar fields, like in conventional television sets [201] or mobile devices [16]. Moreover, in the particular case of evaluating display systems, it is almost inevitable to turn to subjective assessment tests to obtain robust conclusions about the users' QoE.

Thus, taking this into account, a subjective assessment test was performed evaluating and comparing different consumer 3D display technologies, in order to understand people's acceptance, and to analyze possible factors and trends that could help in the improvement of 3D video development. With this aim, a deep comparison among the three most common 3D display technologies addressed to the consumer market (i.e., stereoscopic displays with polarized glasses, with active shutter glasses, and multiview

autostereoscopic displays) was carried out evaluating several perceptual factors (e.g., picture quality, depth quality, visual discomfort). To avoid the possible influence of the screen size in the comparison, displays of the same dimensions were used. In addition, standard recommendations for subjective tests with 3D video were considered, as well as realistic viewing situations, so the test environment was set mimicking typical living-room conditions to obtain more representative results, since these displays are usually oriented to home entertainment. Furthermore, videos from the NAMA3DS1 database, described in section 3.2 were used in the test, since it is a wide high-quality (with Full-HD for each view) collection of 3D videos, especially developed for subjective studies.

The following subsections present the details of the setup and results of the subjective assessment test.

6.2.1. Subjective test setup

6.2.1.1. Evaluation methodology

The use of an appropriate methodology is crucial to obtain reliable results in the subjective tests. Therefore, the recommendation ITU-R BT.2021 [92] was followed, as it provides the guidelines to carry out subjective tests of 3DTV systems, extending the traditional standard ITU-R BT.500 [89] for subjective assessments with conventional television. Since the goal of this work was to compare various display technologies (coding or transmission artifacts do not degrade the test sequences) and a simultaneous comparison is not feasible (e.g., due to the use of different glasses), a single stimulus method was selected. Thus, all the test sequences were evaluated by observers in each display at a time, following the ACR methodology (see subsection 2.5.1), and presenting a different random order of the displays in each test session. So, as depicted in Figure 2.33, after each test clip, a message in a gray background was displayed during ten seconds providing the time to the observers to write their scores in a paper questionnaire, and indicating the number of the corresponding box to write them, as depicted in Figure 2.33.

In particular, for each test sequence, the observers were asked to evaluate the picture quality, depth quality, and visual discomfort, which are the primary perceptual dimensions in 3D video [92]. A five-grade quality scale was used to evaluate picture quality and depth quality, while a five-grade comfort scale was used for visual discomfort, as shown in Table 2.1 [92].

After the evaluation of all the test sequences in a display, the observers were asked to provide a global score for the primary perceptual factors for that specific display, in addition to global evaluations for the additional perceptual dimensions: naturalness and sense of presence [92]. Also, an overall evaluation of the whole visual experience utilizing the quality scale was requested. Moreover, after the evaluation of all the displays, the observers were asked to rank them considering their global QoE.

The test sessions started with an exhaustive explanation of the experiment (focusing attention on the definitions of the perceptual dimensions extracted from the recommendation ITU-R BT.2021 [92]), a visual screening of the observers, and the visualization

of training sequences in all the displays, in order to familiarize the observers with the different TVs and provide a reference for the subsequent evaluations. After evaluating each display, the observers could have a break time while the following display was set up. The total duration of the whole assessment session was around fifty minutes.

To collect the opinions from the observers, questionnaires with boxes were used, where they were asked to write a mark for the corresponding evaluation. All the questionnaires were available in Spanish and English (for foreigners).

6.2.1.2. Environment and equipment

As the scope of this work was focused on consumer displays for 3D video visualization for home entertainment, the three most common technologies have been considered: stereoscopic displays with passive polarized glasses and active shutter glasses, and multiview autostereoscopic displays [32].

Four consumer displays were selected to carry out the tests, all of them having a screen size of 55", in order to avoid the influence of the screen dimensions in the visual experience. In particular, the displays used in the experiment are detailed in Table 6.1, where the values of luminance were measured without glasses and through them (in parenthesis). The complete technical specifications of the displays can be found in [119, 157, 171, 200].

The displays SO and LG are typical Full-HD LED consumer displays, while SA is a curved Ultra-High-Definition (UHD) LED display. According to manufacturers, this type of screen improves some aspects in comparison with typical flat screens, such as immersiveness, sense of depth, contrast, sharpness, field of view, and effective viewing angle. On the other hand, some drawbacks have been observed, like worse effects of onscreen reflections, geometry issues when watching from the sides, or a limited viewing area around the center of the screen to appreciate the advantages of the curved display. Some of these aspects were taken into account in the analysis of the results, except from those related to viewing positions, since they were not considered in the realization of the tests, where the observers were seated around the center of the screen.

The considered autostereoscopic display uses the lenticular technology to provide nine views (taking a stereo pair as input and using a proprietary algorithm to create seven more perspectives) with an effective resolution of 720p, thanks to a total panel resolution of 3840x2160 pixels, and a pixel layout to group each nine pixels. The display has an integrated camera that detects the observers' faces to adjust the viewing zones. In particular, it recognizes the person closest to the center of the TV and sets four more adjacent viewing zones. This process is especially important to assure the correct position of the observers within one of the viewing zones, which are approximately 50 cm wide, and allow the observers to properly perceive the nine views without blurring effects that appear when they are not well positioned.

To playback the test sequences a PC with a graphic card *GeForce GTX 560* and *Nvidia 3D Vision* drivers were used to run the *Stereoscopic Player* [177], providing the output to the displays in frame-packing 1080p/24Hz format (Full-HD each stereo

ID	Model	Technology	Luminance (cd/m ²)	Release Year
SO	Sony KDL-55HX920	Stereoscopic: Shutter glasses TDG-BR250/B	250 (100)	2011
LG	LG 55LW980S	Stereoscopic: Polarized glasses AG-F315	360 (140)	2011
SA	Samsung UE55HU8500L	Stereoscopic: Shutter glasses SSG-5100GB	260 (80)	2014
TO	Toshiba 55ZL2G	Autostereoscopic: Glasses-free	290	2012

Table 6.1: Characteristics of the displays

Figure 6.1: Test environment

view). However, due to severe flickering effects with the SO display, this TV was used in frame-packing format 720p/60Hz.

Since the considered displays are addressed to consumer market, the test conditions should be kept as far as possible similar to a typical real-life usage. Thus, the displays were configured in standard mode with default settings (e.g., medium brightness, maximum contrast, etc.) and disabling any image processing algorithm that they could implement (e.g., increase depth, edge enhancement, etc.), apart from the generation of the nine views of the autostereoscopic display, and the upscaling of SA and SO displays to fit the whole screen.

Furthermore, one of the main aspects in the design of a subjective test is the selection of the appropriate test environment. Therefore, keeping the idea of real-life usage, the test environment was set mimicking home conditions, as shown in Figure 6.1. Thus, a couch was placed at around 2.1 meters from the display position (which corresponds to a viewing distance of 3H and it is within the viewing distance ranges specified by the manufacturers of all the displays used), and the ambient lightning conditions were

ID	Sequence	Duration	SI	TI	DSI	DTI	D+	D-
V1	Barrier gate	16 s	59	21	20.4	15.4	6	9
V2	Basket	16 s	71	41	11.7	9.7	-14	26
V3	Hall	16 s	82	5	17	7	7	-3
V4	Lab	16 s	53	12	17.8	10	6	22
V5	Phone call	16 s	36	13	21.6	11.8	17	15
V6	Soccer	16 s	89	38	24.7	18.1	7	10
V7	Tree branches	16 s	101	14	23	13.6	3	9
V8	Umbrella	13 s	74	19	17	15.2	5	17
V9	Big Buck Bunny	16 s	60	30	26.1	35.2	21	3
T1	News report	16 s	53	4	21.6	8.7	16	6
T2	Boxers	16 s	50	19	24.4	18	9	3

Table 6.2: Sequences used in the tests

controlled to avoid disturbing reflections.

6.2.1.3. Test material

A total of nine stereoscopic Full-HD (1920x1080 each view) videos were used as source material for the test. The sequences were selected in order to cover a wide range of contents and spatial, temporal, and depth characteristics. All the sequences were selected from the freely available database NAMA3DS1, described in section 3.2, except *Big Buck Bunny*, that is a segment of an open source animation movie [10]. Although the properties of the majority of the sequences were provided in section 3.2, for ease of reading and analysis of the results of the current tests, the main features of the considered videos are described in Table 6.2. The sequences V1-V9 were used for the test sequence generation, while the sequences T1 and T2 were used in the training process of the observers (explained in subsection 6.2.1.1). The original sequences were converted using *Ffmpeg* [45] to *avi* files containing uncompressed video with 24 fps without audio, to avoid the introduction of possible coding degradations and synchronization errors, and the influence of audio in the observers' QoE. Since one of the objectives of the experiment was to compare the QoE perceived by the observers with stereoscopic and monoscopic content, the videos were also displayed in 2D using twice the left view as inputs, so the viewing conditions were the same than in 3D (frame packing format, display properties, observers wearing glasses, etc.). To equalize the conditions in which the videos were evaluated, the test sequences (according to ACR methodology depicted in Figure 2.33) were generated concatenating the 18 videos (9 in 3D and 9 in 2D), using different randomizations for each observer and display, with the condition of not watching the same source consecutively.

6.2.1.4. Observers

A total of 26 observers (9 females, 17 males) participated in the tests, all of them having normal or corrected vision as reported in a previous visual screening carried out before each test session, in which a Randot test were used for 3D vision. The ages of the participants ranged between 20 and 50, with an average age of 31. The observers were rewarded for their participation in the tests, and a maximum of two observers were allowed in each test session. Furthermore, a screening of the results scores provided by the observers was carried out according to the method recommended in ITU-R BT.500 [89], but no outliers were detected.

6.2.2. Results

In the following subsections the main results obtained from the experiment are exposed. The results shown in the figures were obtained computing the MOS from the evaluations provided by the observers in the questionnaires. In addition, the 95% CIs are represented according to the computation recommended by the standard ITU-R BT.500 [89].

6.2.2.1. Picture quality

The results obtained concerning picture quality are shown in Figure 6.2a for the 3D videos and Figure 6.2b for 2D sequences. Regarding 3D picture quality, the general trend is that SA and LG displays provide better image performance, while the autostereoscopic display (TO) is the weakest system in this aspect, due to the loss of sharpness caused by the reduction of the resolution by the number of views. An important point that could be extracted from these results is the difficulty in perceiving the resolution loss of the stereoscopic display with polarized glasses (LG). Also, the effects of visualizing the videos in 720p format in the SO display instead of 1080p are not easily discernible, since the picture quality of this display is similar to the other stereoscopic displays. Considering the results for the monoscopic test videos there is more similarity among the performances of each display.

To study the statistical significance of the differences among MOS values of picture quality for different displays and contents, analyses of variance have been performed. In particular, a two-way repeated measures ANOVA showed significant effects of the display ($p < 0.0001$ for 3D content and $p < 0.05$ for 2D), the content ($p < 0.0001$ for 3D and 2D), and a significant interaction between both factors ($p < 0.0001$ for 3D and 2D). Moreover, one-way repeated measures analyses were applied to study the simple main effects, revealing a significant effect of content for each display ($p < 0.0001$ for all cases in 2D and 3D) and a significant effect of display for each content (at least $p < 0.05$), except for V1, V5, and V9 for 3D, and V1, V3, V7, and V9 for 2D. For the 3D versions, this could be caused by the lower spatial detail of V1 and V5 in comparison with the rest of the clips, which could lessen the impact of resolution losses of LG and TO displays, and the typically better visualization of computer graphics movies in 3D for

(a) 3D sequences

(b) 2D sequences

Figure 6.2: Results for picture quality

V9. On the other hand, the less significant influence of the display in the 2D contents is in accordance with the high performances of the displays working in a conventional mode and without resolution losses. It is also worth noting that, in general, the poorest results were obtained for V2 and V4. In fact, V2 is a scene of a basketball match with a wide camera panning that could introduce some motion artifacts in 3D affecting the perceived quality. In the case of V4, poor results were obtained both in 2D and 3D, which could be caused by a strong lightning change contained in the sequence.

6.2.2.2. Depth quality

In relation with depth quality, Figure 6.3a and Figure 6.3b show the results for 3D and 2D content, respectively. As it can be observed, similar results have been obtained for the three stereoscopic displays (SA, LG, and SO) providing a good performance for all 3D contents. However, the results confirm the worse ability of autostereoscopic displays to provide depth sensation to the observers. Again, V2 is the content with lower scores for 3D depth quality, which, as aforementioned, could be caused by motion artifacts produced by the camera panning. On the other hand, V8 and V9 provide the best results. Concretely, the first sequence is a scene recorded from a close distance and taking care on avoiding the frame effect, in which the umbrella offers a good depth sensation as it is going out of the screen. Regarding V9, it is a cartoon video in which depth issues are easily controlled. Finally, as expected, the results show a poor depth quality of the test videos when displayed in 2D for all displays.

Similarly as for picture quality, ANOVA analyses have been performed with the MOS values obtained for depth quality. The two-way repeated measures ANOVA also showed a significant effect of the display and the content ($p < 0.0001$ for 3D and 2D for both effects), and a significant interaction between both factors ($p < 0.005$ for 3D and 2D). In addition, one-way repeated measures analyses revealed a significant effect of content for each display (at least $p < 0.05$ for all cases in 2D and 3D), except for the LG in 2D. Also, a significant effect of display was proved for all contents in 3D ($p < 0.0001$ for all cases), but only for V2, V4 and V8 in 2D at $p < 0.05$ level. Again, the less significant influence of the display in 2D reflects a uniform performance of the TVs in monoscopic mode, while LG display offers a more stable response regarding the depth sensation of content visualized in 2D. In opposition, the good results of SA obtained in some monoscopic videos, like in V2 based on a camera panning in a basketball match, could reflect an improvement in depth sensation thanks to the curved screen. However, this is not evidenced with 3D videos as exhibited by the similarity of the results for the three stereoscopic displays.

6.2.2.3. Visual discomfort

The results obtained from the evaluations of visual discomfort are shown in Figure 6.4a and Figure 6.4b, for 3D and 2D content, respectively. In general, the best performance in terms of visual comfort for 3D content is provided by the autostereoscopic display. On one hand, this fact could be related to the reduction of the depth sensation offered

(a) 3D sequences

(b) 2D sequences

Figure 6.3: Results for depth quality

(a) 3D sequences

(b) 2D sequences

Figure 6.4: Results for visual discomfort

by the autostereoscopic display, since it can lessen the impact of excessive disparity (as happens in V7) or frame effects (as appear in V4, where the disparity concordance between the content and the screen frame is violated). On the other hand, apart from the annoyance caused to the visual system, some observers considered the use of glasses as a source of discomfort in their viewing experience, so this could be another reason to the better performance of the TO display in this aspect. It is worth noting that for some contents the autostereoscopic display is significantly better than the stereoscopic technology with active shutter glasses, possibly due to the flickering effects that can be perceived with these systems in some scenes with high motion activity (V2 has the greatest TI) or bright colors (V5).

Regarding the visual comfort in 2D test videos, in general, better results were obtained in comparison with the corresponding 3D versions of the sequences, and similar performances are provided by all displays. Generally better results were reported by the autostereoscopic display, since no glasses are needed, and the stereoscopic with polarized glasses, caused by the absence of flickering effects.

Again, the two-way repeated measures ANOVA carried out with the MOS values obtained for visual discomfort reflected a significant effect of the display and the content ($p < 0.0001$ for 3D and 2D for both effects), and a significant interaction between both factors ($p < 0.0001$ for 3D and $p < 0.01$ for 2D). Furthermore, one-way repeated measures analyses revealed a significant effect of content for each display (at least $p < 0.005$ for all cases in 2D and 3D), and a significant effect of display for all contents (at least $p < 0.05$ for all cases) except for V8 in 3D, and V4, V7, and V8 in 2D. In the case of V8, it could be caused by the good properties of the content (slow motion, no frame effects, no excessive disparity, etc.), while for V4 and V7 the visual comfort with the monoscopic versions is considerably better than in 3D, probably due to frame effects and excessive disparity, respectively. A similar behavior could be observed with V2 that could be produced by the appearance of motion artifacts in 3D.

6.2.2.4. Global display comparison

As mentioned in subsection 6.2.1.1, after the evaluation of all the test videos in one display, the observers were asked to provide global evaluations of that display in terms of the five primary and additional perceptual dimensions [92], and a score for the overall QoE. From the results collected in the questionnaires, the MOS values were computed and they are depicted in Figure 6.5.

Concerning picture quality (PQ) and depth quality (DQ), the results confirm the aforementioned conclusions regarding the better performance of SA and LG in terms of picture quality, and the poor ability of the autostereoscopic display to provide depth sensation.

Looking into visual discomfort (VD), the best results are provided by the TO and LG displays, mainly thanks to the glasses-free technology in the first case, and the absence of flickering effects in the latter. The impact of flickering effects that appear using active shutter glasses are also reflected in the significantly worse results that the SO display obtains in this perceptual dimension.

Figure 6.5: Global evaluations

In relation with the additional perceptual dimensions, for naturalness (N) the LG display seems to offer a better perceptual realism. In the case of sense of presence (I), which is one of the aspects of the viewers' QoE that could be enhanced with curved displays (such as SA), similar results were obtained among the stereoscopic displays, while the autostereoscopic display offers the worst immersiveness.

Taking into account the evaluations of the overall QoE provided by each display, it can be observed that the SA and LG displays are the best options for 3D video visualization in a consumer environment. On one hand, this shows the good performance of stereoscopic displays with polarized glasses, in spite of the loss of resolution due to the spatial multiplex of the stereo views. These displays provide a high picture quality, good depth sensation, and the glasses are generally more comfortable than active shutter glasses and do not cause flickering effects. On the other hand, the evolution of the technology of stereoscopic displays with active shutter glasses, reducing the discomfort of the glasses, is evidenced, since the SA display provides one of the best performances. This fact, in addition to the increase of screen resolution offering a better picture quality, improves considerably the QoE provided by the older SO display. Finally, the results also reflect the low capabilities of the autostereoscopic displays to provide a satisfactory 3D sensation, in spite of the advantage of not needing glasses in terms of comfort.

To analyze the statistical significance of the differences among the MOS values obtained for each perceptual dimension and each display, a one-way repeated measures ANOVA was applied, showing a significant effect of the display in all dimensions (PQ: $p < 0.05$, DQ: $p < 0.0001$, VD: $p < 0.0001$, N: $p < 0.05$, I: $p < 0.0001$) and in the global QoE evaluation ($p < 0.0001$).

At the end of the test, the observers were asked to rank the displays according to their preferences. The results for the rankings concerning the overall QoE are shown in Figure 6.6. As it can be seen, the favorite alternative was the stereoscopic display

Figure 6.6: Displays ranked by the observers

with polarized glasses, selected as the best display by more than half the observers, followed by the newest display with active shutter glasses. On the other hand, the autostereoscopic display seems to be the worst option. These results confirm those aforementioned and depicted in Figure 6.5.

6.2.2.5. Other observations

After the evaluation of each display, the observers were allowed to write free-form comments about their visual experience. Mainly, these observations were focused on the discomfort caused by the active shutter glasses (due to flickering effects, reflections or lack of comfort, especially when the observer also wears progressive glasses), changes in the colors in the case of the SO display, and blurring effects in the autostereoscopic display.

In addition, the observers were asked whether they felt visual fatigue, and 54% of the observers reported symptoms like eye fatigue and slight dizziness or headache.

6.3. Evaluation of Super Multiview Video

Nowadays, SMV displays are still under research, so they are not yet widely available and only a few prototypes have been presented [193]. However, there is a need of evaluating the impact on the viewers' QoE of certain effects related to SMV visualization, such as horizontal parallax, smoothness, or visual comfort. Due to the unavailability of SMV displays, these parameters should be investigated using common 3D displays.

Thus, a preliminary study of the impact of various factors on the subjective evaluation of SMV content using two legacy displays was carried out, taking as reference a similar study that was proposed for the previous generation of multiview displays [173].

Figure 6.7: SMV scenarios

Concretely, the functionality of SMV displays has been simulated implementing a view sweeping of SMV sequences.

6.3.1. Subjective test setup

The objective of SMV displays is to provide a viewing experience without visual discomfort and a smooth horizontal parallax as the observers move their heads to see different viewpoints of the scene. To achieve this, the images projected by the display should not be so different between neighboring viewpoints that the observers could perceive discontinuities. This difference between neighboring views could be affected by the disparity of the content, the separation between the cameras when capturing the scene, and the rate of displaying new viewpoints for a certain head movement [173].

Therefore, this motivates the research on the influence of the view density and speed of head movements on the QoE with SMV. Due to the reduced availability of SMV displays, these parameters should be investigated using common 3D displays simulating the head movement of the observers to see different views of the scene.

Typical SMV scenarios are depicted in Figure 6.7, where the scene is captured by a set of n cameras (typically 80 or more), original or synthesized, spanning a certain total angle VR. As described in subsection 2.4.1, to acquire multiview sequences, the cameras could be set in linear and arc arrays, and also in linear convergent arrays [44, 80].

Thus, to evaluate SMV visualization in conventional 3D displays, the head movement of the observers to see different views of the scene has been simulated by displaying a sweep of those different views to static observers. This allows the evaluation of the effects of the View-Sweep Speed (VSS), the View Range (VR), and the View Density (VD). So, the user navigation changing the viewpoint within VR has been simulated composing video sequences by concatenating frames from consecutive views. Specifically, given a sweep time T_s , the parameters that define the view sweep (other than n and VR) are the following:

$$\text{ViewDensity} : VD = n/VR \quad (6.1)$$

$$\text{View - SweepSpeed} : VSS = VR/Ts \quad (6.2)$$

Thus, one sequence can be viewed for a given VR with a given VSS, and a given VD. In this first approach the view-sweep parameters are defined for cameras arranged in an arc, in which the angular change between cameras is constant. Nevertheless, a linear arrangement of cameras distant enough from the scene object could also be easily approximated by this parametrization.

6.3.1.1. Evaluation methodology

The objective of the test was to derive ranges of comfortable VSS and VD, and determine the influence of the view-sweep effect and the amplitude of the VR in the 3D perception of the scene.

Therefore, the subjects were asked to judge the following factors:

- Smoothness of the view transitions using a five-grade quality scale, as shown in Table 2.1 [89].
- 3D quality using the five-grade quality scale and taking into account factors such as depth perception, immersiveness, etc.
- Visual comfort using the five-grade comfort scale, as shown in Table 2.1 [92], and considering all the factors that could make uncomfortable the visual experience, such as difficulties in focusing and converging, visual discomfort caused by depth, comfort of the view sweeping speed, etc.

Thus, to evaluate the test sequences, the ACR method was used. So, after an initial message indicating the start of the test, the test sequences were shown, followed by a message indicating the observers to rate the corresponding video sequence (see Figure 2.33). In particular, the voting messages had a duration of ten seconds, to allow the subjects to comfortably provide their ratings after watching each test clip. Also, to collect the opinions of the observers, questionnaires with numbered boxes were used, where they were asked to write a mark for the corresponding evaluation.

The test sessions consisted of a previous visual screening of the subjects, followed by a training process in which some example sequences were shown to them explaining the test methodology and purpose. This also provided a reference to the observers of the 3D capabilities of the displays, so they could rate the factors to evaluate (especially 3D quality) taking into account the performance of each display and without directly comparing them (which was not the purpose of the test). Then, the test videos were shown to the observers in each display, with a rest period of approximately five minutes during the display change.

Also, different randomizations of the videos were used in each session and for each display to reduce contextual effects, with the condition of showing twice each test sequence but never consecutively. In addition, the display order was changed in each session. Two observers participated at a time in each session, which lasted one hour approximately.

SRC	Resolution	Framerate	Movement	Num. of cameras	Camera distance (mm)	Angle between adjacent cameras
Bee	1920x1088	-	Static	185	3.711	$\sim 0.226^\circ$
Shark	1920x1088	30 fps	Moving	185	7.004	$\sim 0.226^\circ$

Table 6.3: Characteristics of the SMV sequences

6.3.1.2. Environment and equipment

The test area was set according to international recommendations [89], with walls covered by mid-gray curtains, and the ambient lightning conditions were controlled to avoid disturbing reflections. The viewing distance was set at 2.1 m from the display position, which corresponds to a viewing distance of 3H and it is within the viewing distance ranges of the displays.

Two displays (one stereoscopic and one autostereoscopic) were used to carry out the subjective tests, both 55", in order to avoid the influence of screen size in the visual experience. In particular, two of the displays used in the test described in the previous section were used: the curved UHD LED consumer stereoscopic display from Samsung, and the autostereoscopic display from Toshiba. The main properties of these displays are shown in Table 6.1. It is worth noting that, in practice, the viewing cone of the autostereoscopic display is very narrow, providing the perception of a single viewpoint.

The videos were played using a PC with a graphic card *Nvidia GeForce GTX 760* and the *Media Player Classic Home Cinema* (MPC-HC) [128], which allowed a smooth playback of the sequences.

6.3.1.3. Test material

In this work, two multiview video sequences have been used, with the highest camera density that is currently available in the category SMV test material from MPEG [79]: *Bee* and *Shark*. Their characteristics are shown in Table 6.3.

Given the camera configuration in the test sequences [163], $VR = 41.62^\circ$. Thus, the maximum VD available using original cameras only is $VD_{\max} = 4.42^{\text{cam}/^\circ}$. Note that the cameras are not arranged in an arc to capture these contents, but in linear convergent configuration. Thus, the angular change between consecutive cameras is not constant, although it is very reduced (the minimum and maximum angle changes are 0.207° and 0.237°), so it has been approximated to the average angular change (0.226°).

For the subjective tests, the following test sequences with three different scene contents were prepared:

- *Bee* (static).
- *Shark static*: obtained using frame 68 of the sequence.

	VSS_0 0 °/s	VSS_1 3.39 °/s	VSS_2 6.78 °/s	VSS_3 13.56 °/s	VSS_4 27.12 °/s	VSS_5 54.24 °/s
VD_1 (cam/°)	-	4.42	4.42	4.42	2.21	1.11
VD_2 (cam/°)	-	2.21	2.21	1.11	0.74	-
VD_3 (cam/°)	-	1.47	1.11	0.74	0.44	-
VD_4 (cam/°)	-	1.11	0.74	0.44	0.28	-

Table 6.4: Combinations of VSS and VD tested

- *Shark motion*: using all 100 frames of the sequence.

Regarding the simulation of the SMV display through view sweep, the full VR (41.62°) and half VR (20.81°) were tested. Each sequence was displayed at five different VSSs: 3.39, 6.78, 13.56, 27.12 and 54.24 °/s. For each value of VSS four different values of VD were selected (see Table 6.4), which are coherent with previous similar works [173]. Additionally, also a fixed-viewpoint sequence (VSS_0) was shown for each case. Given that all test sequences have the same length, and are shown at different values of VSS, the number of times that the camera covers VR is different for each value of VSS. All tests sequences had a duration of ten seconds, and were displayed at a spatial resolution of 1920x1088 pixels (in SbS format) and a framerate of 60 fps. In the case of *shark motion*, the framerate was converted from 30 fps to 60 fps by frame repetition with a factor of $\times 2$. The stereo baseline for all test sequences was set to 5.57 cm.

6.3.1.4. Observers

A total of 22 observers (6 females, 16 males) participated in the tests, all of them having normal or corrected vision. The ages of the participants ranged between 22 and 49, with an average age of 30. Furthermore, a screening of the results scores provided by the observers was carried out analyzing the Pearson correlation between the data provided by each subject and the average of all the subjects [78], which lead to discard two observers.

The participants, all of them from UPM, were asked to fill a questionnaire where they indicated their experience in visualizing 3D video. All of them reported to have watched 3D content before the tests. In particular, 9% of the observers had watched 3D video once, 77% occasionally, and 14% frequently.

6.3.2. Experimental results

The results shown in the following figures were obtained computing the MOS and CIs from the evaluations provided by the observers in the questionnaires, according to the recommendation ITU-R BT.500 [89].

Display	Content	Highest comfortable VSS	Most comfortable VSS range
Stereo	Static	13.56 °/s	[6.78 – 13.56] °/s
	Moving	6.78 °/s	[6.78 – 13.56] °/s
Auto	Static	27.12 °/s	[3.39 – 6.78] °/s
	Moving	13.56 °/s	[3.39 – 6.78] °/s

Table 6.5: Comfortable values of VSS

6.3.2.1. Comfortable view-sweep speed

The first objective of the study was to establish a range of VSS values that are comfortable to view SMV content in stereoscopic and autostereoscopic displays. Therefore, the visual comfort results were analyzed, considering that a VSS is comfortable if the MOS score of the visual comfort is strictly above 3 (including the entire range of the CI). Figure 6.8a and Figure 6.8b show the MOS results for the visual comfort for all values of VSS for the stereoscopic and autostereoscopic displays, respectively. Concretely, for each value of VSS, the results for VD_1 are shown (i.e., the highest view density that has been tested). As a reference, the visual comfort results for the fixed-camera sequences ($VSS_0 = 0$) are also included. In addition, Table 6.5 shows the highest value of comfortable VSS, and the most comfortable VSS range for each type of content and display.

As shown in the figures, the visual comfort generally decreases with increasing VSS, except for the lowest VSS in static-content sequences, for which the highest VD values are not enough to provide a smooth camera navigation, affecting the observers' comfort.

Moreover, comparing both types of content, the static scenes allow higher values of VSS as the combination of scene motion and camera movement creates visual discomfort.

Finally, regarding the display, the autostereoscopic provides better comfort, even for high VSS, probably thanks to not needing glasses to watch 3D content.

6.3.2.2. Comfortable view density

The second objective was to find the range of values of VD that provide a perception of smooth transition through consecutive views. To find those values, the smoothness scores indicated by the observers were analyzed, considering again that the view sweep is smooth enough if the MOS score of the smoothness is strictly above 3 (including the entire range of the CI). Figure 6.9a and Figure 6.9b show the MOS results for the smoothness for all the combinations of VSS and VD values (see Table 6.4), in the stereoscopic display, for the static (averaging the results for *Bee* and *Shark static*) and moving content, respectively. Also, Figure 6.10a and Figure 6.10b show the same type of results for the autostereoscopic display. Note that for VSS_5 only one value of VD was tested.

(a) Stereoscopic display

(b) Autostereoscopic display

Figure 6.8: Visual comfort results for different values of VSS

VSS	Stereo		Auto	
	Static	Moving	Static	Moving
VSS_0	-	-	-	-
VSS_1	X	$VD_2 = 2.21$	$VD_1 = 4.42$	$VD_1 = 4.42$
VSS_2	$VD_1 = 4.42$	$VD_3 = 1.11$	$VD_2 = 2.21$	$VD_2 = 2.21$
VSS_3	$VD_2 = 1.11$	$VD_2 = 1.11$	$VD_2 = 1.11$	$VD_1 = 4.42$
VSS_4	$VD_2 = 0.74$	X	$VD_2 = 0.74$	$VD_1 = 4.42$
VSS_5	-	-	-	-

Table 6.6: Minimum value of VD per value of VSS.

(a) Static content

(b) Moving content

Figure 6.9: Smoothness with the stereoscopic display

As expected, these ranges of “smooth” VD values depend on VSS. Thus, in Table 6.6 the minimum value of VD that is acceptable for a given value of VSS is provided. In fact, specially interesting are the VSS values that were identified as most comfortable for each type of content and display in the previous subsection 6.3.2.1. In particular, results marked in blue correspond to the most comfortable VSS for a given content type and display.

The main conclusions that can be drawn from the analysis of the smoothness results are the following:

- As stated in the previous subsection 6.3.2.1, high values of VSS in the moving sequence are uncomfortable, and this discomfort may have influenced the observers when ranking the smoothness of the view sweep, leading to irrelevant results (marked in red in Table 6.6).

(a) Static content

(b) Moving content

Figure 6.10: Smoothness with the autostereoscopic display

- Regarding the differences between both types of content for the stereoscopic display, the moving sequence can be shown with lower VD, and be still perceived as smooth.
- For the autostereoscopic display there are no remarkable difference in the values of minimum VD that are smooth for both types of content (for comfortable VSS values).

6.3.2.3. Effect of view sweep in 3D perception

Another objective of this experiment was to analyze the influence of the view-sweep effect in the 3D perception of the observers (i.e., perception of depth and immersiveness).

(a) Stereoscopic display

(b) Autostereoscopic display

Figure 6.11: 3D quality for the maximum VD and for each sequence

Thus, the 3D quality of the view-sweep sequences for different values of VSS against the fixed-camera sequences were compared. Figure 6.11a and Figure 6.11b show the MOS of the 3D quality results for each different content and VSS with the maximum VD (i.e., VD_1), to avoid the influence of the lack smoothness in the analysis.

The main conclusions that could be drawn from these results are the following:

- The 3D perception improves with view sweep for static sequences, but not for the moving sequences. This improvement is higher for autostereoscopic displays, in which the depth perception is lower initially.
- This 3D perception drops for sequences with motion at high VSSs that are not comfortable (see subsection 6.3.2.1).

6.3.2.4. Influence of the viewing range

Finally, the influence of the amplitude of VR in the visual comfort and the 3D quality perceived by the observers was also analyzed. The results showed that:

- There is no statistically significant difference in visual comfort between the MOS scores for full and half VR with the sequences considered in the test.
- Although the scores for 3D quality with full VR are slightly higher, the difference between the MOS scores for full and half VR are not statistically significant.

6.3.2.5. Other results

The participants were asked to fill a questionnaire after the test in which they were asked whether they felt visual fatigue or headache. Concretely, the fatigue was measured considering the five-grade scale defined in [226] (ranged from: 5-“My eyes are not tired” to 1-“My eyes are tired”), and a similar one was used for headache. It is worth noting that almost 55% of the observers expressed that they felt visual fatigue, and 23% reported a slight headache.

6.4. Conclusions

Although an important technological development has been accomplished during the last years, three dimensional displays have not yet achieved the total acceptance of the consumers. One of the main causes of this issue is the need of using specific glasses to watch 3D videos in stereoscopic systems, which are the most common 3D monitors nowadays. Moreover, autostereoscopic display technology, which is for the moment the most viable alternative to avoid glasses to watch 3D content, does not provide a satisfactory QoE for its price, because of the inherent loss of resolution that severely affects the image quality. Therefore, nowadays it seems that the hope of the 3D video industry is placed in the research on 3D display technologies, such as holographic, integral imaging systems, and especially SMV displays, to offer a high 3D QoE to the viewers.

In fact, the two experiments presented in this chapter could provide an idea of the drawbacks of current 3D displays and the benefits of promising technologies, like SMV systems. In particular, the tests were focused on two aspects: firstly, the evaluation of the performance of current consumer market display technologies, and secondly, on preliminary evaluations of the SMV technology, under the scope of the standardization work carried out by MPEG.

In particular, the first subjective test was carried out comparing the three most common technologies for 3D video visualization: stereoscopic displays with polarized glasses and with active shutter glasses, and multiview autostereoscopic displays. The subjective test was designed taking into account that these types of displays are addressed to the domestic use, so realistic viewing conditions were simulated to obtain representative

results. In addition, several factors influencing the users' QoE were evaluated using a diverse high-quality video database. In essence, some interesting conclusions could be extracted from the results. Firstly, the stereoscopic display technology with polarized glasses, which was the favorite alternative for more than half the observers, offers a notable performance, even though the resolution is reduced due to the spatial multiplex of both stereo views. Secondly, there has been a great improvement of the technology of stereoscopic displays with active shutter glasses, especially reducing discomforting aspects of the glasses, like flickering effects. Finally, autostereoscopic displays still provide a poor 3D visual experience, despite the advantage of not needing glasses to watch 3D content.

This low performance of current autostereoscopic displays is expected to be overcome with SMV displays, which were the focus of the second experiment presented in this chapter. Although nowadays the availability of these displays is still very limited, it is necessary to study the impact on the viewers' QoE of the new effects that arise with SMV displays. Therefore, a novel subjective approach was presented simulating the functionality of SMV displays, through view sweep using a stereoscopic and an autostereoscopic display. So, various parameters in relation with the horizontal parallax provided by this technology were investigated from the evaluation of perceptual factors involved in the 3D QoE, such as smoothness, visual comfort, and depth perception. Despite the limitations of this innovative experiment (especially in terms of content variety and limited to arc configurations of cameras), some conclusions could be extracted to serve as an initial guide for setting up subjective tests, concerning comfortable view-sweep speeds and satisfactory view densities for smooth horizontal parallax. For instance, this experiment allowed to investigate that the subjective perception of VSS and VD (in terms of speed comfort and smoothness) do not only depend on the angle or distance between cameras and view change rate, but also on other parameters not covered by the considered parametrization, such as focal distance, field of view, convergence, or distance to scene objects. As an example, the perceived VSS is radically different for two set of cameras in the same arrangement, when, for one of them the field of view is wide and for the other it is narrow. Therefore, more studies are needed, and the results of the presented work may provide relevant inputs subjective evaluations of SMV content [80] and for coding standardization purposes within the FTV group of MPEG.

7 Conclusions and future work

This chapter summarizes the main aspects of 3D video systems addressed in this thesis and presents the general conclusions extracted from the developed work. In addition, possible future research lines are pointed out.

7.1. Conclusions

During the last years, the audiovisual media industry has witness the last attempt to bring to the viewers a more complete visual experience with 3D video. In fact, although the breaking point was achieved with the arrival of 3D movies to the cinemas, the most decisive advance has been the introduction of 3D video technologies in domestic environments. So, nowadays it is possible to watch 3D content not only in consumer displays at households, but even in portable devices, such as smartphones.

However, after an initial period of growing interest and expectation from people towards 3D video content, the consumption of 3D video technologies suffered a critical decline finishing with the establishment of 3D video as a supplementary alternative, more than a substitute of conventional video. This has been the result of various critical factors related with all the elements through the video processing chain that have hindered the definitive success of 3D video systems.

Firstly, one of the major problems of this last attempt to bring 3D video to consumers has been the lack of high-quality content. The production of 3D content requires the proper use of more complex capture systems than with conventional video, as well as the application of new rules and guidelines to acquire 3D video that provides a satisfactory depth perception and a comfortable 3D visual experience [229]. In particular, the impact on the viewers' QoE of typical effects related to 3D content creation has been studied through a subjective test in chapter 3 in this thesis. For instance, aspects concerning the calibration of stereoscopic cameras have been analyzed, showing the critical importance of these issues in obtaining proper 3D content, such as the effects of geometrical distortions, which specially may appear in non-professional environments. In addition, some post-processing techniques were also considered in the study, such as the conversion from 2D video to 3D, which generally do not provide a comparable performance than directly recording 3D content. Actually, this has been a very important issue regarding the success of 3D video, since the appearance of various films converted to 3D from original 2D content may have caused that people related the real performance of 3D video with these poor quality-approaches, entailing skepticism towards the possibilities of 3D video. In addition, regarding the lack of easily-accessible

high-quality 3D video content, it is an imperative issue for the research community to carry out their investigations. Thus, the generation of a freely available database, called NAMA3DS1, has been also presented in chapter 3, which has been already widely used for research purposes.

Secondly, the processes of coding and transmission of 3D video have probably been those presenting the best development in comparison with the rest of steps of the video processing chain. In fact, the development of various representation formats and coding alternatives for 3D video, together with the adaptations of distribution scenarios to offer 3D content, have allowed the rapid introduction of 3D video technologies in home environments. In this sense, perceptual issues of different coding alternatives and configurations have been subjectively evaluated in chapter 4, while the impact of transmission events in different 3D video delivery scenarios has been investigated in chapter 5. On one hand, the high performance provided by frame-compatible representations, which allowed using the existing infrastructure to deliver 3D video, has been demonstrated with the comparison with 3D Full-HD video, expected to provide a better QoE. On the other hand, the impact of transmission events on the visual experience of the end users has been analyzed to get conclusions that may help in the development of objective metrics for video quality monitoring, error concealment techniques, and strategies for transmitting 3D video over current video delivery services (e.g., IPTV, HAS, etc.). For example, switching between 3D and 2D content has been reported to be a good alternative in these systems, while asymmetric coding is another good option when binocular rivalry is not provoked. Nevertheless, further research is needed to improve the performance of coding and transmission approaches (e.g., depth-based schemes) to make possible the enhancement of the 3D visual experience of the viewers with the new display technologies.

Actually, the development of new visualization systems for 3D video is one of the main challenges for the industry and research communities, since the current technologies have been another critical factor in the unachieved consolidation of 3D video in the media entertainment environment. Mainly, two factors have been decisive in this sense. Firstly, the need of using specific glasses to watch 3D content, which do not provide a natural and comfortable visual experience to the observers. Secondly, the visual discomfort experienced by the observers watching 3D content, due to different factors related to the capture, coding, and transmission of 3D video, and to the accommodation-vergence conflict that appear in current displays. In this sense, chapter 6 presented a subjective assessment test analyzing the performance of current consumer 3D displays (e.g., stereoscopic displays with active and passive glasses, and autostereoscopic displays), reporting conclusions about the advantages and drawbacks of each technology, such as resolution reductions in stereoscopic displays with polarized glasses, flickering effects in displays with active shutter glasses, or the low picture quality, in contrast to the higher visual comfort, provided by autostereoscopic displays. In addition, chapter 3 and 4 also provided some conclusions in relation to the screen size, since usually bigger displays provide a better 3D QoE to the observers [26]. Furthermore, also in chapter 6, another subjective experiment was presented to contribute in the work towards the development of near-future 3D displays, focused on investigating the possibilities of SMV systems, which seems to be the most promising technology

to provide an enhanced 3D visual experience to consumers without requiring special glasses, with better parallax, and without accommodation-vergence conflict. Actually, the research on emerging 3D displays systems, such as SMV and holographic displays, is essential in the challenge of providing a natural and high-quality 3D visual experience to viewers, and finally establishing 3D video technologies in the media entertainment market over conventional video.

Therefore, the studies presented in this thesis have allowed the understanding of the perceptual factors related to 3D video systems. In fact, the evaluation of these factors have been carried out taking into account the need of adequate testing methodologies to obtain reliable results. Thus, the activities in this sense carried out by the VQEG have been considered, as well as the development of novel methodologies that try to simulate more realistic viewing conditions to obtain more representative conclusions of what end users perceive at their homes. Actually, a novel methodology for subjective assessment of typical events that take place through the 3D video distribution chain has been proposed. This methodology, called CIETI, tries to mimic real-viewing conditions, considering the use of typical videos watched at households in terms of content and duration, as well as test environments emulating usual living room conditions. So, it is possible to carry out assessment tests in which the observers are more immersed in the visual experience than in detecting artifacts. Although it has been validated through various experiments, further research is needed to improve the methodology making it as less intrusive as possible.

Finally, the outcomes of the work in this thesis have been published in international journals and conferences. For instance, the NAMA3DS1 database work on content creation was published in *QoMEX 2012* [202], and has been widely used by the research community of 3D video. Secondly, the work on coding issues of 3D video resulted in collaborations with VQEG under the scope of the 3DTV project [209], and with MPEG with the participation in the validation tests for MPEG 3DV coding in 2011 [78]. Furthermore, the work carried out related to transmission events in 3D video distribution systems, resulted in various publications. For example, the studies concerning 3D video delivery over IPTV networks were published in various conferences, as well as in the *Bell Labs Technical Journal* [32, 57, 58]. In addition, the work considering adaptive streaming strategies for 3D video was published in *ICIP 2012* and in various international journals, such as *IEEE Journal on Selected Areas in Communications* [60, 194, 195, 196]. Also, the validation of the CIETI methodology was published in *QoMEX 2012* [61]. Moreover, the study comparing consumer display technologies for 3D video was published in the *Journal of Display Technology* [56], while the work concerning SMV was published in *QoMEX 2012* [15]. Finally, apart from the aforementioned contributions, the work carried out during the elaboration of this thesis in relation to the development of objective quality metrics resulted in publications in various conferences [30, 37, 59, 62, 140], and also in a contribution to the book *IPTV Multiservice QoE Management System* [31].

7.2. Future work

The work presented in this thesis may be extended with future research mainly envisioned in the three following directions:

- Continuing with the subjective studies of the effects of 3D video systems in the QoE perceived by the viewers in concordance with technological developments related to the 3D video processing chain. In particular, special interest will be focused on the advances in coding techniques (e.g., HEVC alternatives for 3D video coding), 3D video distribution (e.g., OTT services with adaptive streaming of 3D video), and visualization technologies (e.g., SMV displays). In this sense, the standardization activities of MPEG and VQEG will be followed-up and efforts will be made to contribute to these tasks.
- Investigating subjective assessment methodologies to obtain reliable results and representative conclusions of the visual experience of the end users of video services and applications. On one side, the work will be oriented towards the improvement of the CIETI methodology, trying to reflect, as far as possible, realistic viewing conditions and to obtain less intrusive test factors. For instance, the method for collecting the ratings from the observers should be improved, trying to obtain a more adequate and comfortable manner avoiding the need of averting the gaze from the screen to carry out the evaluations. In addition, the methodology should be adapted to the new video transmission scenarios, such as HAS, in which different events should be evaluated that may entail the revision of certain aspects, like the rating scale. On the other side, other methodologies will be also considered in the future works, such as the emerging approaches of crowdsourcing or the research on more appropriate methodologies carried out by the VQEG.
- Continuing the work towards the development of perceptual models to be applied in coding algorithms (e.g. V+D coding, HEVC, etc.), transmission techniques (e.g., protection schemes, 3D video-conferencing, etc.), and QoE monitoring tools. Special interest will be centered on this latter issue, in order to improve and complement the monitoring approaches that have been focus of work during the development of this thesis, considering the implementation of useful NR metrics and the application of theses approaches to emerging video services, such as HAS.

Bibliography

- [1] J. Apostolopoulos and A. Reibman, “The Challenge of Estimating Video Quality in Video Communication Applications,” *IEEE Signal Processing Magazine*, vol. 29, no. 2, pp. 160–158, Mar. 2012.
- [2] M. Asikuzzaman, M. J. Alam, A. J. Lambert, and M. R. Pickering, “A blind watermarking scheme for depth-image-based rendered 3D video using the dual-tree complex wavelet transform,” in *IEEE International Conference on Image Processing*, Paris, France, Oct. 2014, pp. 5497–5501.
- [3] S. Atkinson, “Stereoscopic-3D storytelling - Rethinking the conventions, grammar and aesthetics of a new medium,” *Journal of Media Practice*, vol. 12, no. 2, pp. 139–156, Jul. 2011.
- [4] T. Balogh, P. T. Kovacs, and A. Barsi, “Holovizio 3D Display System,” in *2007 3DTV Conference*, Kos, Greece, May 2007, pp. 1–4.
- [5] M. Barkowsky, M. H. Pinson, R. Ppion, and P. Le Callet, “Analysis of Freely Available Subjective Dataset for HDTV including Coding and Transmission Distortions,” in *International Workshop on Video Processing and Quality Metrics for Consumer Electronics*, Scottsdale, USA, Jan. 2010.
- [6] M. Barkowsky, K. Wang, R. Cousseau, K. Brunnstrm, R. Olsson, and P. Le Callet, “Subjective Quality Assessment of Error Concealment Strategies for 3DTV in the Presence of Asymmetric Transmission Errors,” in *International Packet Video Workshop*, Hong Kong, China, Dec. 2010.
- [7] V. Baroncini, J.-R. Ohm, and G. J. Sullivan, “Report of Subjective Test Results of Responses to the Joint Call for Proposals on Video Coding Technology for High Efficiency Video Coding (HEVC),” Document JCTVC-A204, Dresden, Germany, Apr. 2010.
- [8] M. Barreda-Angeles, R. Pepion, E. Bosc, P. Le Callet, and A. Pereda-Baos, “Exploring the effects of 3D visual discomfort on viewers’ emotions,” in *IEEE International Conference on Image Processing (ICIP)*, Oct. 2014, pp. 753–757.
- [9] S. A. Benton, *Selected Papers on Three-dimensional Displays*. Bellingham, WA: SPIE Optical Engineering Press, 2001.
- [10] Blender Foundation, “Big Buck Bunny Project Homepage.” [Online]. Available: <http://www.bigbuckbunny.org>.
- [11] J. L. Blin, “New Quality Evaluation Method Suited To Multimedia Context SAMVIQ,” in *International Workshop on Video Processing and Quality Metrics for Consumer Electronics*, Scottsdale, USA, Jan. 2006.

-
- [12] A. Boev, D. Hollosi, and A. Gotchev, "Classification of stereoscopic artefacts," Mobile3DTV Technical Report, Jul. 2008. [Online]. Available: http://sp.cs.tut.fi/mobile3dtv/results/tech/D5.1_Mobile3DTV_v1.0.pdf
 - [13] D. Brewster, *The stereoscope; its history, theory, and construction, with its application to the fine and useful arts and to education*. J. Murray London, 1856.
 - [14] G. Bruder, F. Steinicke, K. Rothaus, and K. Hinrichs, "Enhancing Presence in Head-Mounted Display Environments by Visual Body Feedback Using Head-Mounted Cameras," in *International Conference on CyberWorlds*, Bradford, UK, Sep. 2009, pp. 43–50.
 - [15] P. Carballeira, J. Gutiérrez, F. Morán, J. Cabrera, and N. García, "Subjective evaluation of super multiview video in consumer 3D displays," in *International Workshop on Quality of Multimedia Experience (QoMEX)*, Costa Navarino, Greece, May 2015, pp. 1–6.
 - [16] A. Catellier, M. Pinson, W. Ingram, and A. Webster, "Impact of mobile devices and usage location on perceived multimedia quality," in *International Workshop on Quality of Multimedia Experience*, Yarra Valley, Australia, Jul. 2012, pp. 39–44.
 - [17] G. W. Cermak, "Consumer Opinions About Frequency of Artifacts in Digital Video," *IEEE Journal of Selected Topics in Signal Processing*, vol. 3, no. 2, pp. 336–343, Apr. 2009.
 - [18] A. Chambolle and T. Pock, "A First-Order Primal-Dual Algorithm for Convex Problems with Applications to Imaging," *Journal of Mathematical Imaging and Vision*, vol. 40, no. 1, pp. 120–145, May 2011.
 - [19] W. Chen, J. Fournier, M. Barkowsky, and P. Le Callet, "New requirements of subjective video quality assessment methodologies for 3DTV," in *International Workshop on Video Processing and Quality Metrics*, Scottsdale, USA, Jan. 2010.
 - [20] —, "Quality of experience model for 3DTV," in *SPIE Stereoscopic Displays and Applications XXIII*, Feb. 2012, p. 82881P.
 - [21] Y. Chen, K. Wu, and Q. Zhang, "From QoS to QoE: A Tutorial on Video Quality Assessment," *IEEE Communications Surveys & Tutorials*, vol. 17, no. 2, pp. 1126–1165, Jan. 2015.
 - [22] Y. Chen, M. M. Hannuksela, T. Suzuki, and S. Hattori, "Overview of the MVC+D 3D video coding standard," *Journal of Visual Communication and Image Representation*, vol. 25, no. 4, pp. 679–688, May 2014.
 - [23] E. Cheng, P. Burton, J. Burton, A. Joseski, and I. Burnett, "RMIT3DV: Pre-announcement of a creative commons uncompressed HD 3D video database," in *International Workshop on Quality of Multimedia Experience*, Yarra Valley, Australia, Jul. 2012, pp. 212–217.
 - [24] S. Chikkerur, V. Sundaram, M. Reisslein, and L. J. Karam, "Objective Video Quality Assessment Methods: A Classification, Review, and Performance Comparison," *IEEE Transactions on Broadcasting*, vol. 57, no. 2, pp. 165–182, Jun. 2011.

- [25] H. Coolican, *Research methods and statistics in psychology*, 2nd ed. Hodder & Stoughton, 1994.
- [26] L. E. Coria, D. Xu, and P. Nasiopoulos, "Quality of experience of stereoscopic content on displays of different sizes: A comprehensive subjective evaluation," in *IEEE International Conference on Consumer Electronics (ICCE)*, Las Vegas, USA, Jan. 2011, pp. 755–756.
- [27] D. Corrigan, F. Pitie, V. Morris, A. Rankin, M. Linnane, G. Kearney, M. Gorzel, M. O'Dea, C. Lee, and A. Kokaram, "A Video Database for the Development of Stereo-3D Post-Production Algorithms," in *Conference on Visual Media Production*, London, UK, Nov. 2010, pp. 64–73.
- [28] P. Coverdale, S. Moller, A. Raake, and A. Takahashi, "Multimedia Quality Assessment Standards in ITU-T SG12," *IEEE Signal Processing Magazine*, vol. 28, no. 6, pp. 91–97, Nov. 2011.
- [29] N. Cranley and L. Murphy, "Incorporating User Perception in Adaptive Video Streaming Systems," in *Multimedia Technologies*. IGI Global, 2008, ch. 108, pp. 1491–1507.
- [30] A. Cuadra, M. Cutanda, A. Pérez, E. Rogles, J. Gutiérrez, and F. Jaureguizar, "SAVAGE: IPTV multiservice QoE management system," in *International Conference on Smart Communications in Network Technologies (SaCoNeT)*, Paris, France, Jun. 2013, pp. 1–5.
- [31] —, "IPTV Multiservice QoE Management System," in *Quality of Experience Engineering for Customer Added Value Services*. Hoboken, NJ, USA: John Wiley & Sons, Inc., Jul. 2014, ch. 5, pp. 73–92.
- [32] J. M. Cubero, J. Gutiérrez, P. Pérez, E. Estalayo, J. Cabrera, F. Jaureguizar, and N. García, "Providing 3D Video Services: The Challenge From 2D to 3DTV Quality of Experience," *Bell Labs Technical Journal*, vol. 16, no. 4, pp. 115–134, Mar. 2012.
- [33] S. J. Daly, "The Visible Differences Predictor: an algorithm for the assessment of image fidelity," in *SPIE Human Vision, Visual Processing, and Digital Display*, B. E. Rogowitz, Ed., Aug. 1992, pp. 2–15.
- [34] C. Danilo, M. Regis, J. V. De Miranda Cardoso, I. De Pontes Oliveira, and M. S. De Alencar, "Objective Estimation of 3D Video Quality : A Disparity-based Weighting Strategy," in *IEEE International Symposium on Broadband Multimedia Systems and Broadcasting*, vol. 2, no. 1, London, UK, Jun. 2013, pp. 1–6.
- [35] F. De Simone, M. Naccari, M. Tagliasacchi, F. Dufaux, S. Tubaro, and T. Ebrahimi, "Subjective Quality Assessment of H.264/AVC Video Streaming with Packet Losses," *EURASIP Journal on Image and Video Processing*, vol. 2011, pp. 1–12, Jan. 2011.
- [36] A. Dehili, "Evolution of Access Networks to support video Quality requirements," in *International Telecommunications Network Strategy and Planning Symposium*, New Dehli, India, Nov. 2006, pp. 1–3.

-
- [37] C. Diaz, J. Gutierrez, J. Cabrera, F. Jaureguizar, and N. Garcia, "Perceptual distortion modeling for side-by-side 3D video delivery," in *IEEE International Conference on Consumer Electronics (ICCE)*, vol. 20412, Las Vegas, USA, Jan. 2014, pp. 418–419.
 - [38] F. Dufaux, B. Pesquet-Popescu, and M. Cagnazzo, *Emerging Technologies for 3D Video*, F. Dufaux, B. Pesquet-Popescu, and M. Cagnazzo, Eds. Chichester, UK: John Wiley & Sons, Ltd, Jun. 2013.
 - [39] E. Dunic, S. Grgic, D. J. Bermejo, and L. Alberto da Silva Cruz, "Benchmark of state of the art objective measures for 3D stereoscopic video quality assessment on the Nantes database," in *Proceedings ELMAR-2014*, Zadar, Croatia, Sep. 2014, pp. 1–4.
 - [40] DVB Project, "Digital Video Broadcasting (DVB), Frame Compatible Plano-Stereoscopic 3DTV (DVB-3DTV)," DVB BlueBook doc. A154, Feb. 2011.
 - [41] M. Emoto, T. Niida, and F. Okano, "Repeated Vergence Adaptation Causes the Decline of Visual Functions in Watching Stereoscopic Television," *Journal of Display Technology*, vol. 1, no. 2, pp. 328–340, Dec. 2005.
 - [42] Y. Fang, J. Wang, J. Li, R. Pepion, and P. Le Callet, "An eye tracking database for stereoscopic video," in *International Workshop on Quality of Multimedia Experience (QoMEX)*, Singapore, Sep. 2014, pp. 51–52.
 - [43] O. Faugeras, "Digital color image processing within the framework of a human visual model," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 27, no. 4, pp. 380–393, Aug. 1979.
 - [44] C. Fehn and R. Pastoor, "Interactive 3-DTV-Concepts and Key Technologies," *Proceedings of the IEEE*, vol. 94, no. 3, pp. 524–538, Mar. 2006.
 - [45] Ffmpeg Project. [Online]. Available: <https://www.ffmpeg.org/>
 - [46] M. C. Forman, N. A. Davies, and M. McCormick, "Objective Quality Measurement of Integral 3D Images," in *Proc. SPIE Stereoscopic Displays and Virtual Reality Systems IX*, May 2002, pp. 155–162.
 - [47] M.-N. Garcia, F. De Simone, S. Tavakoli, N. Staelens, S. Egger, K. Brunnstrom, and A. Raake, "Quality of experience and HTTP adaptive streaming: A review of subjective studies," in *International Workshop on Quality of Multimedia Experience (QoMEX)*, Singapore, Sep. 2014, pp. 141–146.
 - [48] J. Geng, "Three-dimensional display technologies," *Advances in Optics and Photonics*, vol. 5, no. 4, p. 456, Dec. 2013.
 - [49] Y. Gitman, M. Erofeev, D. Vatolin, B. Andrey, and F. Alexey, "Semiautomatic visual-attention modeling and its application to video compression," in *IEEE International Conference on Image Processing*, Paris, France, Oct. 2014, pp. 1105–1109.
 - [50] L. Goldmann, F. De Simone, and T. Ebrahimi, "A Comprehensive Database and Subjective Evaluation Methodology for Quality of Experience in Stereoscopic

- Video,” *SPIE Electronic Imaging, 3D Image Processing and Applications*, vol. 7526, pp. 75 260S–75 260S–11, Feb. 2010.
- [51] E. B. Goldstein, *Sensation and Perception*. Wadsworth Publishing, 2009.
 - [52] P. Gorley and N. Holliman, “Stereoscopic image quality metrics and compression,” in *SPIE Stereoscopic Displays and Applications*, A. J. Woods, N. S. Holliman, and J. O. Merritt, Eds., Feb. 2008, pp. 680 305–680 305–12.
 - [53] O. Grau, M. Muller, and J. Kluger, “Tools for 3D-TV programme production,” in *3DTV Conference: The True Vision - Capture, Transmission and Display of 3D Video (3DTV-CON)*, Antalya, Turkey, May 2011, pp. 1–4.
 - [54] C. G. Gürler and A. M. Tekalp, “Adaptive Streaming of Multiview Video Over P2P Networks,” in *Emerging Technologies for 3D Video*, F. Dufaux, B. Pesquet-Popescu, and M. Cagnazzo, Eds. Chichester, UK: John Wiley & Sons, Ltd, Apr. 2013, ch. 10, pp. 187–204.
 - [55] C. G. Gurler, K. Bagci, and A. M. Tekalp, “Adaptive stereoscopic 3D video streaming,” in *IEEE International Conference on Image Processing*, Hong Kong, China, Sep. 2010, pp. 2409–2412.
 - [56] J. Gutiérrez, F. Jaureguizar, and N. García, “Subjective Comparison of Consumer Television Technologies for 3D Visualization,” *Journal of Display Technology*, vol. 11, no. 11, pp. 967–974, Nov. 2015.
 - [57] J. Gutiérrez, P. Pérez, F. Jaureguizar, J. Cabrera, and N. García, “Subjective assessment of the impact of transmission errors in 3DTV compared to HDTV,” in *2011 3DTV Conference: The True Vision - Capture, Transmission and Display of 3D Video (3DTV-CON)*, Antalya, Turkey, May 2011, pp. 1–4.
 - [58] —, “Subjective evaluation of transmission errors in IPTV and 3DTV,” in *Visual Communications and Image Processing (VCIP)*, Tainan, Taiwan, Nov. 2011, pp. 1–4.
 - [59] —, “Monitoring packet loss impact in IPTV and 3DTV receivers,” in *International Conference on Consumer Electronics (ICCE)*, Las Vegas, USA, Jan. 2012, pp. 172–173.
 - [60] —, “Subjective study of adaptive streaming strategies for 3DTV,” in *IEEE International Conference on Image Processing*, vol. 20412, Orlando, USA, Sep. 2012, pp. 2265–2268.
 - [61] —, “Validation of a novel approach to subjective quality evaluation of conventional and 3D broadcasted video services,” in *International Workshop on Quality of Multimedia Experience (QoMEX)*, Yarra Valley, Australia, Jul. 2012, pp. 230–235.
 - [62] J. Gutiérrez, A. Raake, S. Argyropoulos, and N. García, “Rule-based combination of video quality metrics,” in *International Workshop on Quality of Multimedia Experience (QoMEX)*, vol. 20412, Costa Navarino, Greece, May 2015, pp. 1–6.
 - [63] D. Hands, “A Basic Multimedia Quality Model,” *IEEE Transactions on Multimedia*, vol. 6, no. 6, pp. 806–816, Dec. 2004.

-
- [64] P. Hanhart and T. Ebrahimi, “EYEC3D: 3D video eye tracking dataset,” in *International Workshop on Quality of Multimedia Experience (QoMEX)*, Singapore, Sep. 2014, pp. 55–56.
 - [65] —, “Subjective evaluation of two stereoscopic imaging systems exploiting visual attention to improve 3D quality of experience,” in *SPIE Electronic Imaging, Stereoscopic Displays and Applications XXV*, Mar. 2014, p. 90110D.
 - [66] S. S. Hemami and A. R. Reibman, “No-reference image and video quality estimation: Applications and human-motivated design,” *Signal Processing: Image Communication*, vol. 25, no. 7, pp. 469–481, Aug. 2010.
 - [67] C. T. E. R. Hewage and M. G. Martini, “Reduced-reference quality metric for 3D depth map transmission,” in *3DTV-Conference: The True Vision - Capture, Transmission and Display of 3D Video*, Tampere, Finland, Jun. 2010, pp. 1–4.
 - [68] Y. Ho and Y. Song, “MPEG activities for 3D video coding,” in *Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA)*, Angkor Wat, Cambodia, Dec. 2014, pp. 1–4.
 - [69] N. S. Holliman, N. a. Dodgson, G. E. Favalora, and L. Pockett, “Three-Dimensional Displays: A Review and Applications Analysis,” *IEEE Transactions on Broadcasting*, vol. 57, no. 2, pp. 362–371, Jun. 2011.
 - [70] T. Hossfeld, C. Keimel, M. Hirth, B. Gardlo, J. Habigt, K. Diepold, and P. Tran-Gia, “Best Practices for QoE Crowdttesting: QoE Assessment With Crowdsourcing,” *IEEE Transactions on Multimedia*, vol. 16, no. 2, pp. 541–558, Feb. 2014.
 - [71] Q. Huynh-Thu, M. Barkowsky, and P. Le Callet, “The Importance of Visual Attention in Improving the 3D-TV Viewing Experience: Overview and New Perspectives,” *IEEE Transactions on Broadcasting*, vol. 57, no. 2, pp. 421–431, Jun. 2011.
 - [72] Q. Huynh-thu, M.-N. Garcia, F. Speranza, P. Corriveau, and A. Raake, “Study of Rating Scales for Subjective Quality Assessment of High-Definition Video,” *IEEE Transactions on Broadcasting*, vol. 57, no. 1, pp. 1–14, Mar. 2011.
 - [73] W. Ijsselsteijn, H. de Ridder, and J. Vliegen, “Subjective evaluation of stereoscopic images: effects of camera parameters and display duration,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 10, no. 2, pp. 225–233, Mar. 2000.
 - [74] IRCCyN-IVC, “Nantes-Madrid 3D Stereoscopic database,” Mar. 2012. [Online]. Available: <http://ivc.univ-nantes.fr/en/databases/NAMA3DS1/?article954>
 - [75] ISO/IEC JTC1/SC29/WG11, “Call for proposals on Scalable Video Coding technology,” Doc. N6193, Waikoloa, HI, Dec. 2003.
 - [76] —, “Updated call for proposal on Multiview Video Coding,” Doc. N7567, Nice, France, Oct. 2005.
 - [77] —, “Call for Proposals on 3D Video Coding Technology,” Doc. N12036, Geneva, Switzerland, Mar. 2011.

- [78] —, “Report of subjective test results from the call for proposals on 3D video coding technology,” Doc. N12347, Geneva, Switzerland, Nov. 2011.
- [79] —, “Experimental framework for FTV,” Doc. N15048, Geneva, Switzerland, Oct. 2014.
- [80] —, “Draft Call for Evidence on FTV,” Doc. N15095, Geneva, Switzerland, Feb. 2015.
- [81] ITU, “Subjective Assessment of Stereoscopic Television Pictures,” Recommendation ITU-R BT.1438, 2000.
- [82] —, “Method for specifying accuracy and cross-calibration of Video Quality Metrics (VQM),” Recommendation ITU-T J.149, Mar. 2004.
- [83] —, “Objective perceptual assessment of video quality: Full reference television,” ITU-T Tutorial, 2004.
- [84] —, “Objective perceptual video quality measurement techniques for digital cable television in the presence of a full reference,” Recommendation ITU-T J.144, 2004.
- [85] —, “Definitions of terms related to quality of service,” Recommendation ITU-T E.800, Sep. 2008.
- [86] —, “Subjective video quality assessment methods for multimedia applications,” *Recommendation ITU-T P.910*, Apr. 2008.
- [87] —, “Advanced Video Coding for Generic Audiovisual Services,” Recommendation ITU-T H.264, 2010.
- [88] —, “Objective perceptual multimedia video quality measurement of HDTV for digital cable television in the presence of a full reference,” Recommendation ITU-T J.341, Jan. 2011.
- [89] —, “Methodology for the subjective assessment of the quality of television pictures,” Recommendation ITU-R BT. 500-13, Jan. 2012.
- [90] —, “Parametric non-intrusive assessment of audiovisual media streaming quality,” Recommendation P.1201, Oct. 2012.
- [91] —, “Parametric non-intrusive bitstream assessment of video media streaming quality,” Recommendation P.1202, Oct. 2012.
- [92] —, “Subjective methods for the assessment of stereoscopic 3DTV systems,” Recommendation ITU-R BT.2021, Aug. 2012.
- [93] —, “Features of three-dimensional television video systems for broadcasting,” Report ITU-R BT.2160, Nov. 2013.
- [94] —, “Methods for the subjective assessment of video quality, audio quality and audiovisual quality of Internet video and distribution quality television in any environment,” Recommendation ITU-T P.913, Jan. 2014.
- [95] B. Javidi, J. Sola-Pikabea, and M. Martinez-Corral, “Breakthroughs in Photonics 2014: Recent Advances in 3-D Integral Imaging Sensing and Display,” *IEEE Photonics Journal*, vol. 7, no. 3, pp. 1–7, Jun. 2015.

-
- [96] JEDI (Just Explre Dimensions) ITEA2 Project. [Online]. Available: <http://www.jedi-itea2.org/>
- [97] L. Jin, A. Boev, A. Gotchev, and K. Egiazarian, "3D-DCT based perceptual quality assessment of stereo video," in *IEEE International Conference on Image Processing*, Brussels, Belgium, Sep. 2011, pp. 2521–2524.
- [98] H. Jorke and M. Fritz, "Infitec - A new stereoscopic visualisation tool by wavelength multiplex imaging," *Journal of three dimensional images*, vol. 19, no. 3, pp. 50–56, Sep. 2005.
- [99] S. Jumisko-Pyykkö and M. M. Hannuksela, "Does context matter in quality evaluation of mobile television?" in *Proc. International conference on Human computer interaction with mobile devices and services - (MobileHCI '08)*, New York, USA, Sep. 2008, p. 63.
- [100] O. Kaller, L. Boleček, M. Slanina, and T. Kratochvíl, "Comparison of Steroscopic Display Systems by Subjective Testing," in *International Conference on Digital Telecommunications*, Chamonix, France, May 2012, pp. 73–77.
- [101] T. Kawano, K. Yamagishi, and T. Hayashi, "Performance comparison of subjective assessment methods for 3D video quality," in *International Workshop on Quality of Multimedia Experience (QoMEX)*, Yarra Valley, Australia, Jul. 2012, pp. 218–223.
- [102] D. Kim and S. Kim, "Visual Comfort Range in the Super-Multi-View Display," in *International display research conference*, Sao Paulo, Brazil, Nov. 2010, pp. 1–4.
- [103] D. Kim and K. Sohn, "Depth adjustment for stereoscopic image using visual fatigue prediction and depth-based view synthesis," in *IEEE International Conference on Multimedia and Expo*, Singapore, Jul. 2010, pp. 956–961.
- [104] M. Klíma, K. Fliegel, P. Páta, S. Vitek, M. Blažek, P. Dostál, L. Krasula, T. Kratochvíl, V. Říčný, M. Slanina, L. Polák, O. Kaller, and L. Boleček, "DEIMOS - An Open Source Image Database," *Radioengineering*, vol. 20, no. 4, pp. 1016–1023, Dec. 2011.
- [105] J. Konrad, S. Member, B. Lacotte, and E. Dubois, "Cancellation of Image Crosstalk in Time-Sequential Displays of Stereoscopic Video," *IEEE Transactions on Image Processing*, vol. 9, no. 5, pp. 897–908, May 2000.
- [106] S. S. Krishnan and R. K. Sitaraman, "Video Stream Quality Impacts Viewer Behavior: Inferring Causality Using Quasi-Experimental Designs," *IEEE/ACM Transactions on Networking*, vol. 21, no. 6, pp. 2001–2014, Dec. 2013.
- [107] V. Kulyk, S. Tavakoli, M. Folkesson, K. Brunnstrom, K. Wang, and N. Garcia, "3D video quality assessment with multi-scale subjective method," in *International Workshop on Quality of Multimedia Experience (QoMEX)*, Klagenfurt, Austria, Jul. 2013, pp. 106–111.
- [108] J. Kuze and K. Ukai, "Subjective evaluation of visual fatigue caused by motion images," *Displays*, vol. 29, no. 2, pp. 159–166, Mar. 2008.

- [109] M. Lambooi, M. Murdoch, W. IJsselsteijn, and I. Heynderickx, "The impact of video characteristics and subtitles on visual comfort of 3D TV," *Displays*, vol. 34, no. 1, pp. 8–16, Jan. 2013.
- [110] M. Lambooi, W. IJsselsteijn, M. Fortuin, and I. Heynderickx, "Visual Discomfort and Visual Fatigue of Stereoscopic Displays: A Review," *Journal of Imaging Science and Technology*, vol. 53, no. 3, p. 030201, May 2009.
- [111] P. Le Callet, S. Möller, and A. Perkis, "Qualinet white paper on definitions of quality of experience (2012)," European Network on QoE in Multimedia Systems and Services (COST Action IC 1003), Jun. 2012.
- [112] P. Lebreton, M. Barkowsky, A. Raake, and P. Le Callet, "3D Video," in *Quality of Experience - Advanced Concepts, Applications and Methods*, ser. T-Labs Series in Telecommunication Services, S. Möller and A. Raake, Eds. Springer International Publishing, 2014, ch. 20, pp. 339–349.
- [113] P. Lebreton, A. Raake, M. Barkowsky, and P. Le Callet, "Evaluating Depth Perception of 3D Stereoscopic Videos," *IEEE Journal of Selected Topics in Signal Processing*, vol. 6, no. 6, pp. 710–720, Oct. 2012.
- [114] —, "Perceptual preference of S3D over 2D for HDTV in dependence of video quality and depth," in *IEEE Image, Video, and Multidimensional Signal Processing Workshop*, Seoul, South Korea, Jun. 2013, pp. 3–6.
- [115] P. Lebreton, A. Raake, U. Wustenhagen, T. Buchholz, M. Barkowsky, and P. Le Callet, "A subjective and objective evaluation of a realistic 3D IPTV transmission chain," in *International Packet Video Workshop*, München, Germany, May 2012, pp. 179–183.
- [116] J.-S. Lee, L. Goldmann, and T. Ebrahimi, "Paired comparison-based subjective quality assessment of stereoscopic images," *Multimedia Tools and Applications*, vol. 67, no. 1, pp. 31–48, Nov. 2013.
- [117] W. J. M. Levelt, "On Binocular Rivalry," Ph.D. dissertation, Royal Van Gorcum, Assen, 1965.
- [118] T. Levola and P. Laakkonen, "Replicated slanted gratings with a high refractive index material for in and outcoupling of light," *Optics Express*, vol. 15, no. 5, p. 2067, Mar. 2007.
- [119] LG, "LG 55LW980S display technical specifications." [Online]. Available: <http://www.lg.com/es/television/lg-55LW980s-3d>
- [120] J. Li, "Methods for assessment and prediction of QoE, preference and visual discomfort in multimedia application with focus on S-3DTV," Ph.D. dissertation, Université de Nantes, 2013.
- [121] J. Li, M. Barkowsky, and P. Le Callet, "The influence of relative disparity and planar motion velocity on visual discomfort of stereoscopic videos," in *International Workshop on Quality of Multimedia Experience (QoMEX)*, Mechelen, Belgium, Sep. 2011, pp. 155–160.

-
- [122] J. Li, O. Kaller, F. De Simone, J. Hakala, D. Juszka, and P. Le Callet, "Cross-lab study on preference of experience in 3DTV: Influence from display technology and test environment," *International Workshop on Quality of Multimedia Experience (QoMEX)*, pp. 46–47, Jul. 2013.
- [123] M. Liao, J. Gao, R. Yang, and M. Gong, "Video Stereolization: Combining Motion Analysis with User Interaction," *IEEE Transactions on Visualization and Computer Graphics*, vol. 18, no. 7, pp. 1079–1088, Jul. 2012.
- [124] J. Lubin, "A human vision system model for objective picture quality measurements," in *International Broadcasting Conference (IBC)*, vol. 1997, Amsterdam, Netherlands, Sep. 1997, pp. 498–503.
- [125] J. Mannos and D. Sakrison, "The effects of a visual fidelity criterion of the encoding of images," *IEEE Transactions on Information Theory*, vol. 20, no. 4, pp. 525–536, Jul. 1974.
- [126] K. Masaoka, "Spatial distortion prediction system for stereoscopic images," *Journal of Electronic Imaging*, vol. 15, no. 1, p. 013002, Jan. 2006.
- [127] ISO/IEC SC29WG11, "Experimental Framework for FTV," Doc. N15048, Strasbourg, France, Oct. 2014.
- [128] Media Player Classic Home Cinema (MPC-HC). [Online]. Available: <http://mpc-hc.org/>
- [129] L. M. J. Meesters, W. A. Ijsselsteijn, and P. J. Seuntjens, "A Survey of Perceptual Evaluations and Requirements of Three-Dimensional TV," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 14, no. 3, pp. 381–391, Mar. 2004.
- [130] B. Mendibru, *3D Movie Making: Stereoscopic Digital Cinema from Script to Screen*. Elsevier, 2009.
- [131] MPEG, "Dynamic adaptive streaming over HTTP (DASH) – Part 1: Media presentation description and segment formats," 2015.
- [132] H. Mukawa, K. Akutsu, I. Matsumura, S. Nakano, T. Yoshida, M. Kuwahara, and K. Aiki, "A full-color eyewear display using planar waveguides with reflection volume holograms," *Journal of the Society for Information Display*, vol. 17, no. 3, p. 185, Jun. 2009.
- [133] T. Oelbaum and K. Diepold, "Building a Reduced Reference Video Quality Metric with Very Low Overhead using Multivariate Data Analysis," in *International Conference on Cybernetics and Information Technologies, Systems and Applications*, vol. 6, no. 5, Orlando, USA, Jul. 2007, pp. 81–86.
- [134] T. Oelbaum, H. Schwarz, M. Wien, and T. Wiegand, "Subjective Performance Evaluation of the SVC extension of H.264/AVC," in *IEEE International Conference on Image Processing*, San Diego, USA, Oct. 2008, pp. 2772–2775.
- [135] Y. Ohira, A. Chekhovskiy, T. Yamanoi, T. Endo, H. Fujita, and H. Toshiyoshi, "Hybrid mems optical scanner for volumetric 3D display," *Journal Soc. Inf. Display*, vol. 17, no. 5, pp. 419–422, May 2009.

- [136] Panasonic Corporation, “Panasonic AVCCAM AG-3DA1 integrated twins-lens 3D camera recorder.” [Online]. Available: http://panasonic.biz/sav/broch_bdf/AG-3DA1_e.pdf
- [137] M.-C. Park and S. Mun, “Overview of Measurement Methods for Factors Affecting the Human Visual System in 3D Displays,” *Journal of Display Technology*, vol. 9, pp. 1–1, Jan. 2015.
- [138] R. R. Pastrana-Vidal and J.-C. Gicquel, “Automatic Quality Assessment of Video Fluidity impairments using a No-Reference Metric,” in *International Workshop on Video Processing and Quality Metrics for Consumer Electronics*, Scottsdale, USA, Jan. 2006.
- [139] P. Pérez and N. García, “Lightweight multimedia packet prioritization model for unequal error protection,” *IEEE Transactions on Consumer Electronics*, vol. 57, no. 1, pp. 132–138, Feb. 2011.
- [140] P. Pérez, J. Gutiérrez, J. J. Ruiz, and N. García, “Qualitative Monitoring of Video Quality of Experience,” in *IEEE International Symposium on Multimedia*, Dana Point, USA, Dec. 2011, pp. 470–475.
- [141] M. Perkins, “Data compression of stereopairs,” *IEEE Transactions on Communications*, vol. 40, no. 4, pp. 684–696, Apr. 1992.
- [142] D. Piedade, M. Bernardo, P. Fiadeiro, A. Pinheiro, and M. Pereira, “Chromatic Variations on 3D Video and QoE,” in *European Signal Processing Conference*, Lisbon, Portugal, Sep. 2014, pp. 221 – 225.
- [143] M. H. Pinson, M. Barkowsky, and P. Le Callet, “Selecting scenes for 2D and 3D subjective video quality tests,” *EURASIP Journal on Image and Video Processing*, vol. 2013, no. 1, p. 50, Aug. 2013.
- [144] M. H. Pinson, M. Sullivan, and A. Catellier, “A New Method for Immersive Audiovisual Subjective Testing,” *International Workshop on Video Processing and Quality Metrics for Consumer Electronics*, Jan. 2014.
- [145] M. H. Pinson and S. Wolf, “Comparing subjective video quality testing methodologies,” in *SPIE Visual Communications and Image Processing*, Jun. 2003, pp. 573–582.
- [146] —, “A New Standardized Method for Objectively Measuring Video Quality,” *IEEE Transactions on Broadcasting*, vol. 50, no. 3, pp. 312–322, Sep. 2004.
- [147] I. Politis, V. De Silva, S. Dogan, C. De Alwis, T. Dagiuklas, and A. M. Kondoz, “Objective measurement of transmission losses and compression artefacts in stereoscopic 3D video,” in *IEEE International Conference on Communications Workshops (ICC)*, Budapest, Hungary, Jun. 2013, pp. 667–671.
- [148] M. T. Pourazad, P. Nasiopoulos, and R. K. Ward, “Generating the Depth Map from the Motion Information of H.264-Encoded 2D Video Sequence,” *EURASIP Journal on Image and Video Processing*, vol. 2010, pp. 1–13, Apr. 2010.
- [149] Qualinet, “Qualinet Databases.” [Online]. Available: <http://dbq.multimediatech.cz/>

-
- [150] A. Raake, J. Gustafsson, S. Argyropoulos, M.-n. Garcia, D. Lindegren, G. Heikkilä, M. Pettersson, P. List, and B. Feiten, “IP-Based Mobile and Fixed Network Audiovisual Media Services,” *IEEE Signal Processing Magazine*, vol. 28, no. 6, pp. 68–79, Nov. 2011.
 - [151] A. R. Reibman, “Monitoring video quality inside a network,” Distinguished Lecture for the IEEE Signal Processing Society, Santa Clara, USA, 2009.
 - [152] —, “Designing a biased specification-based subjective test of image quality,” in *SPIE-IS&T Electronic Imaging*, vol. 9394, Mar. 2015, p. 939409.
 - [153] S. Reichelt, R. Häussler, G. Fütterer, and N. Leister, “Depth cues in human visual perception and their realization in 3D displays,” in *SPIE Three-Dimensional Imaging, Visualization, and Display*, vol. 76900B, Orlando, Apr. 2010, p. 76900B.
 - [154] U. Reiter, K. Brunnström, K. De Moor, M.-C. Larabi, M. Pereira, A. Pinheiro, J. You, and A. Zgank, “Factors Influencing Quality of Experience,” in *Quality of Experience - Advanced Concepts, Applications and Methods*, ser. T-Labs Series in Telecommunication Services, S. Möller and A. Raake, Eds. Cham: Springer International Publishing, 2014, ch. 4, pp. 339–349.
 - [155] M. Rerabek and T. Ebrahimi, “Comparison of 3D portable display restitution techniques based on stereo and motion parallax,” in *International Workshop on Quality of Multimedia Experience (QoMEX)*, Yarra Valley, Australia, Jul. 2012, pp. 80–85.
 - [156] D. C. Robinson, Y. Jutras, and V. Craciun, “Subjective Video Quality Assessment of HTTP Adaptive Streaming Technologies,” *Bell Labs Technical Journal*, vol. 16, no. 4, pp. 5–23, Mar. 2012.
 - [157] Samsung, “Samsung UE55HU8500L display technical specifications.” [Online]. Available: <http://www.samsung.com/es/consumer/tv-av/tv/uhd/UE55HU8500LXXC>
 - [158] G. Saygili, C. G. Gurler, and A. M. Tekalp, “Evaluation of Asymmetric Stereo Video Coding and Rate Scaling for Adaptive 3D Video Streaming,” *IEEE Transactions on Broadcasting*, vol. 57, no. 2, pp. 593–601, Jun. 2011.
 - [159] Z. M. P. Sazzad, S. Yamanaka, and Y. Horita, “Spatio-temporal segmentation based continuous no-reference stereoscopic video quality prediction,” in *International Workshop on Quality of Multimedia Experience (QoMEX)*, Trondheim, Norway, Jun. 2010, pp. 106–111.
 - [160] D. Scharstein and R. Szeliski, “A Taxonomy and Evaluation of Dense Two-Frame Stereo Correspondence Algorithms,” *International Journal of Computer Vision*, vol. 47, no. 1/3, pp. 7–42, May 2002.
 - [161] H. Schwarz, D. Marpe, and T. Wiegand, “Overview of the Scalable Video Coding Extension of the H.264/AVC Standard,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 17, no. 9, pp. 1103–1120, Sep. 2007.
 - [162] I. Sedano, M. Kihl, K. Brunnstrom, and A. Aurelius, “Evaluation of video quality metrics on transmission distortions in H.264 coded video,” in *IEEE Interna-*

- tional Symposium on Broadband Multimedia Systems and Broadcasting (BMSB)*, Nuremberg, Germany, Jun. 2011, pp. 1–5.
- [163] T. Senoh, A. Ishikawa, M. Okui, K. Yamamoto, and N. Inoue, “FTV AHG: Super-Multiview Sequences of NICT,” *Contrib. M32201*, San Jose, USA, Jan. 2014.
 - [164] J. Seo, X. Liu, D. Kim, and K. Sohn, “An Objective Video Quality Metric for Compressed Stereoscopic Video,” *Circuits, Systems, and Signal Processing*, vol. 31, no. 3, pp. 1089–1107, Nov. 2011.
 - [165] P. J. Seuntiëns, “Visual Experience of 3D TV,” Ph.D. dissertation, Eindhoven University of Technology, 2006.
 - [166] P. J. Seuntiëns, L. M. J. Meesters, and W. A. Ijsselsteijn, “Perceptual attributes of crosstalk in 3D images,” *Displays*, vol. 26, no. 4-5, pp. 177–183, Oct. 2005.
 - [167] J. Shade, S. Gortler, L.-W. He, and R. Szeliski, “Layered depth images,” in *Proc. Conference on Computer graphics and interactive techniques (SIGGRAPH)*, New York, USA, Jul. 1998, pp. 231–242.
 - [168] M. Slanina, T. Kratochvíl, V. Říčný, L. Boleček, O. Kaller, and L. Polák, “Testing QoE in Different 3D HDTV Technologies,” *Radioengineering*, vol. 21, no. 1, pp. 445–454, Apr. 2012.
 - [169] A. Smolic, G. Tech, and H. Brust, “Report on generation of stereo video data base,” *Mobile3DTV Technical Report*, Jul. 2010. [Online]. Available: http://sp.cs.tut.fi/mobile3dtv/results/tech/D2.1_Mobile3DTV_v1.pdf
 - [170] J.-Y. Son, B. Javidi, S. Yano, and K.-H. Choi, “Recent Developments in 3-D Imaging Technologies,” *Journal of Display Technology*, vol. 6, no. 10, pp. 394–403, Oct. 2010.
 - [171] Sony, “Sony KDL-55HX920 display technical specifications.” [Online]. Available: <http://www.sony.co.uk/support/en/content/cnt-specs/KDL-55HX920/list>
 - [172] F. Speranza, W. J. Tam, R. Renaud, and N. Hur, “Effect of disparity and motion on visual comfort of stereoscopic images,” in *Proc. SPIE Stereoscopic Displays and Virtual Reality Systems XIII*, Feb. 2006, pp. 60 550B–60 550B–10.
 - [173] F. Speranza and W. Tam, “Perceived smoothness of viewpoint transition in multi-viewpoint stereoscopic displays,” in *Proc. SPIE Stereoscopic Displays and Virtual Reality Systems XII*, vol. 5664, Mar. 2005, pp. 72–82.
 - [174] N. Staelens, S. Moens, W. Van den Broeck, I. Marien, B. Vermeulen, P. Lambert, R. Van de Walle, and P. Demeester, “Assessing Quality of Experience of IPTV and Video on Demand Services in Real-Life Environments,” *IEEE Transactions on Broadcasting*, vol. 56, no. 4, pp. 458–466, Dec. 2010.
 - [175] M. Stanley, M. A. Smith, A. P. Smith, P. J. Watson, S. D. Coomber, C. D. Cameron, C. W. Slinger, and A. Wood, “3D electronic holography display system using a 100-megapixel spatial light modulator,” in *SPIE Optical Design Engineering*, Feb. 2004, pp. 297–308.

-
- [176] L. Stelmach, Wa James Tam, D. Meegan, and A. Vincent, "Stereo image quality: effects of mixed spatio-temporal resolution," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 10, no. 2, pp. 188–193, Mar. 2000.
- [177] Stereoscopic Player. [Online]. Available: <http://www.3dtv.at/>
- [178] T. Stockhammer, "Dynamic adaptive streaming over HTTP - Standards and Design Principles," in *Proc. ACM Conference on Multimedia Systems (MMSys)*, no. i, New York, USA, Feb. 2011, pp. 133–143.
- [179] R. C. Streijl, S. Winkler, and D. S. Hands, "Mean opinion score (MOS) revisited: methods and applications, limitations and alternatives," *Multimedia Systems*, pp. 1–20, Dec. 2014.
- [180] D. Strohmeier, S. Jumisko-Pyykkö, K. Kunze, G. Tech, D. Buğdayci, and M. OguzBici, "D4.3 Results of Quality Attributes of Coding, Transmission, and their Combinations," *Mobile3DTV Technical Report*, Feb. 2010, available at <http://sp.cs.tut.fi/mobile3dtv/>. [Online]. Available: http://sp.cs.tut.fi/mobile3dtv/results/tech/D4.3_Mobile3DTV_v1.0.pdf
- [181] D. Strohmeier, S. Jumisko-Pyykkö, and K. Eulenberg, "Open Profiling of Quality: Probing the method in the context of use," in *International Workshop on Quality of Multimedia Experience (QoMEX)*, Mechelen, Belgium, Sep. 2011, pp. 7–12.
- [182] A. Sullivan, "DepthCube solid-state 3D volumetric display," in *SPIE Conference Series*, vol. 5291, May 2004, pp. 279–284.
- [183] G. J. Sullivan, J. M. Boyce, Y. Chen, J.-r. Ohm, C. A. Segall, and A. Vetro, "Standardized Extensions of High Efficiency Video Coding (HEVC)," *IEEE Journal of Selected Topics in Signal Processing*, vol. 7, no. 6, pp. 1001–1016, Dec. 2013.
- [184] M. Sullivan, J. Pratt, and P. Kortum, "Practical issues in subjective video quality evaluation," in *Proc. International Conference on Designing Interactive User Experiences for TV and Video (UXTV)*, New York, USA, Oct. 2008, p. 1.
- [185] P. Surman and X. W. Sun, "Towards the reality of 3D imaging and display," in *3DTV-Conference: The True Vision - Capture, Transmission and Display of 3D Video (3DTV-CON)*, Budapest, Hungary, Jul. 2014, pp. 1–4.
- [186] R. Suryakumar, "Study of the dynamic interactions between vergence and accommodation," Ph.D. dissertation, Faculty of Sciences, University of Waterloo, Ontario, 2005.
- [187] Y. Takaki, "Development of Super Multi-View Displays," *ITE Transactions on Media Technology and Applications*, vol. 2, no. 1, pp. 8–14, Jan. 2014.
- [188] Y. Takaki, Y. Urano, and H. Nishio, "Motion-parallax smoothness of short-, medium-, and long-distance 3D image presentation using multi-view displays," *Optics Express*, vol. 20, no. 24, p. 27180, Nov. 2012.
- [189] Y. Takaki and M. Yokouchi, "Accommodation measurements of horizontally scanning holographic display," *Optics Express*, vol. 20, no. 4, pp. 3918–3931, Feb. 2012.

- [190] W. J. Tam, F. Speranza, S. Yano, K. Shimono, and H. Ono, "Stereoscopic 3D-TV: Visual Comfort," *IEEE Transactions on Broadcasting*, vol. 57, no. 2, pp. 335–346, Jun. 2011.
- [191] X. Tang, M. Yu, X. Jin, Z. Peng, F. Shao, and G. Jiang, "New Stereo Video Quality Assessment Metric for Three Dimensional Video Systems," *Journal of Computers*, vol. 9, no. 11, pp. 2609–2615, Nov. 2014.
- [192] M. Tanimoto, "FTV: Free-viewpoint Television," *Signal Processing: Image Communication*, vol. 27, no. 6, pp. 555–570, Jul. 2012.
- [193] —, "FTV standardization in MPEG," in *3DTV-Conference: The True Vision - Capture, Transmission and Display of 3D Video (3DTV-CON)*, Budapest, Hungary, Jul. 2014, pp. 1–4.
- [194] S. Tavakoli, K. Brunnström, J. Gutiérrez, and N. García, "Quality of Experience of adaptive video streaming: Investigation in service parameters and subjective quality assessment methodology," *Signal Processing: Image Communication*, vol. 39, pp. 432–443, Nov. 2015.
- [195] S. Tavakoli, J. Gutiérrez, and N. García, "Quality assessment of adaptive 3D video streaming," in *Three-Dimensional Image Processing and Applications*, A. M. Baskurt and R. Sitnik, Eds., vol. 8650, Mar. 2013, p. 86500O.
- [196] —, "Subjective Quality Study of Adaptive Streaming of Monoscopic and Stereoscopic Video," *IEEE Journal on Selected Areas in Communications*, vol. 32, no. 4, pp. 684–692, Apr. 2014.
- [197] T. C. Thang, H. X. Nguyen, A. T. Pham, and Nam Pham Ngoc, "Perceptual difference evaluation of video alternatives in adaptive streaming," in *International Conference on Communications and Electronics (ICCE)*, Hue, Vietnam, Aug. 2012, pp. 322–326.
- [198] A. Tikanmaki, A. Gotchev, A. Smolic, and K. Miller, "Quality assessment of 3D video in rate allocation experiments," *IEEE International Symposium on Consumer Electronics*, pp. 1–4, Apr. 2008.
- [199] Ting-Lan Lin, S. Kanumuri, Yuan Zhi, D. Poole, P. Cosman, and A. Reibman, "A Versatile Model for Packet Loss Visibility and its Application to Packet Prioritization," *IEEE Transactions on Image Processing*, vol. 19, no. 3, pp. 722–735, Mar. 2010.
- [200] Toshiba, "Toshiba 55ZL2G display technical specifications." [Online]. Available: <http://www.toshiba.eu/discontinued-products/55zl2/>
- [201] S. Tourancheau, P. Le Callet, and D. Barba, "Image and Video Quality Assessment Using LCD: Comparisons with CRT Conditions," *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences*, vol. E91-A, no. 6, pp. 1383–1391, Jun. 2008.
- [202] M. Urvoy, M. Barkowsky, R. Cousseau, Y. Koudota, V. Ricorde, P. Le Callet, J. Gutiérrez, and N. García, "NAMA3DS1-COSPAD1: Subjective video quality

- assessment database on coding conditions introducing freely available high quality 3D stereoscopic sequences,” in *International Workshop on Quality of Multimedia Experience*, Yarra Valley, Australia, Jul. 2012, pp. 109–114.
- [203] A. Vetro, A. M. Tourapis, K. Müller, and T. Chen, “3D-TV Content Storage and Transmission,” *IEEE Transactions on Broadcasting*, vol. 57, no. 2, pp. 384–394, Jun. 2011.
- [204] A. Vetro, T. Wiegand, and G. J. Sullivan, “Overview of the Stereo and Multiview Video Coding Extensions of the H.264/MPEG-4 AVC Standard,” *Proceedings of the IEEE*, vol. 99, no. 4, pp. 626–642, Apr. 2011.
- [205] VQEG, “3DTV project.” [Online]. Available: <http://www.its.bldrdoc.gov/vqeg/>
- [206] ———, “Validation of Reduced-Reference and No-Reference Objective models for standard definition television, Phase I,” 2009.
- [207] ———, “Report on the Validation of Video Quality Models for High Definition Video Content,” Jun. 2010.
- [208] ———, “Test Plan for Crosstalk Influences on User Quality of Experience,” Oct. 2010.
- [209] VQEG, “Test Plan for establishing a Ground Truth for Quality of Experience in 3D for assessment methodologies in 3D Video Quality Assessment GroTruQoE3D1,” 2014.
- [210] G. V. Wallendaël, S. V. Leuven, J. D. Cock, P. Lambert, R. V. D. Walle, N. Staelens, and P. Demeester, “Evaluation of Full-Reference Objective Video Quality Metrics on High Efficiency Video Coding,” *IFIP/IEEE International Symposium on Integrated Network Management*, pp. 1294–1299, May 2013.
- [211] J. Wang, M. P. DaSilva, P. Le Callet, and V. Ricordel, “Computational Model of Stereoscopic 3D Visual Saliency,” *IEEE Transactions on Image Processing*, vol. 22, no. 6, pp. 2151–2165, Jun. 2013.
- [212] K. Wang, M. Barkowsky, K. Brunnstrom, M. Sjöström, R. Cousseau, and P. Le Callet, “Perceived 3D TV Transmission Quality Assessment: Multi-Laboratory Results Using Absolute Category Rating on Quality of Experience Scale,” *IEEE Transactions on Broadcasting*, vol. 58, no. 4, pp. 544–557, Dec. 2012.
- [213] K. Wang, M. Barkowsky, R. Cousseau, K. Brunnström, R. Olsson, P. Le Callet, and M. r. Sjöström, “Subjective evaluation of HDTV stereoscopic videos in IPTV scenarios using absolute category rating,” in *SPIE-IS&T Stereoscopic Displays and Applications*, vol. 7863, Feb. 2011, pp. 78 631T–78 631T–12.
- [214] S. Wang, D. Zheng, J. Zhao, W. J. Tam, and F. Speranza, “An Image Quality Evaluation Method Based on Digital Watermarking,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 17, no. 1, pp. 98–105, Jan. 2007.
- [215] Z. Wang, L. Lu, and A. C. Bovik, “Video Quality Assessment based on Structural Distortion Measurement,” *Signal Processing: Image Communication*, vol. 19, no. 2, pp. 121–132, Feb. 2004.

- [216] A. B. Watson, J. Hu, and J. F. McGowan III, "Digital video quality metric based on human vision," *Journal of Electronic Imaging*, vol. 10, no. 1, pp. 20–29, Jan. 2001.
- [217] J. Welch and J. Clark, "A Proposed Media Delivery Index (MDI)," Internet Engineering Task Force (IETF) RFC 4445, 2006.
- [218] C. Wheatstone, "Contributions to the physiology of vision I: On some remarkable, and hitherto unobserved, phaenomena of binocular vision," *Proceedings of the Royal Society of London*, vol. 18, no. 13, pp. 371–395, Jan. 1838.
- [219] S. Winkler, *Digital Video Quality. Vision models and metrics*, Wiley, Ed. Wiley, 2005.
- [220] —, "Analysis of Public Image and Video Databases for Quality Assessment," *IEEE Journal of Selected Topics in Signal Processing*, vol. 6, no. 6, pp. 616–625, Oct. 2012.
- [221] S. Winkler and D. Min, "Stereo/multiview picture quality: Overview and recent advances," *Signal Processing: Image Communication*, vol. 28, no. 10, pp. 1358–1373, Nov. 2013.
- [222] S. Winkler and P. Mohandas, "The Evolution of Video Quality Measurement: From PSNR to Hybrid Metrics," *IEEE Transactions on Broadcasting*, vol. 54, no. 3, pp. 660–668, Sep. 2008.
- [223] A. J. Woods, T. Docherty, and R. Koch, "Image distortions in stereoscopic video systems," in *SPIE Stereoscopic Displays and Applications*, vol. 1915, Sep. 1993, pp. 36–48.
- [224] L. Xing, J. You, T. Ebrahimi, and A. Perkis, "Stereoscopic quality datasets under various test conditions," in *International Workshop on Quality of Multimedia Experience (QoMEX)*, Klagenfurt, Austria, Jul. 2013, pp. 136–141.
- [225] H. Yamanoue, M. Okui, and F. Okano, "Geometrical analysis of puppet-theater and cardboard effects in stereoscopic HDTV images," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 16, no. 6, pp. 744–752, Jun. 2006.
- [226] S. Yano, S. Ide, T. Mitsunashi, and H. Thwaites, "A study of visual fatigue and visual comfort for 3D HDTV/HDTV images," *Displays*, vol. 23, no. 4, pp. 191–201, Sep. 2002.
- [227] S. Yasakethu, C. Hewage, W. Fernando, and A. Kondo, "Quality analysis for 3D video using 2D video quality models," *IEEE Transactions on Consumer Electronics*, vol. 54, no. 4, pp. 1969–1976, Nov. 2008.
- [228] J. D. Yun, Y. Kwak, and S. Yang, "Evaluation of Perceptual Resolution and Crosstalk in Stereoscopic Displays," *Journal of Display Technology*, vol. 9, no. 2, pp. 106–111, Feb. 2013.
- [229] F. Zilly, J. Kluger, and P. Kauff, "Production Rules for Stereo Acquisition," *Proceedings of the IEEE*, vol. 99, no. 4, pp. 590–606, Apr. 2011.

- [230] V. Zota and J.-K. Janssen, “Tiefenbehandlung, 2D-Videos in 3D abspielen (in german),” *c’t Magazin fuer Computer Technik*, vol. 6, p. 116, 2010.