DISSERTAÇÃO

Mestrado em Engenharia Eletrotécnica

Asymmetric 3D Video Coding Based on Regions of Perceptual Relevance

Luís Filipe Henriques Pinto

Leiria, novembro de 2013

MASTER DISSERTATION

Electrical Engineering

Asymmetric 3D Video Coding Based on Regions of Perceptual Relevance

Luís Filipe Henriques Pinto

Master dissertation performed under the guidance of Professor Pedro António Amado Assunção of Escola Superior de Tecnologia e Gestão of Instituto Politécnico de Leiria.

Leiria, November 2013

Learn from yesterday,
live for today,
hope for tomorrow.
The important thing
is not to stop questioning.

(Albert Einstein)

Acknowledgments

I would like to thank to everyone that in someway helped during this work and that made it possible to accomplish.

I am extremely grateful to professor Pedro Assunção for his excellent guidance not only during this work but also through the last 4 years working at Instituto de Telecomunicações. My thanks in this matter extend to professors Sérgio Faria and Nuno Rodrigues who were also always available to give valuable advice. These people introduced me to the research environment early during my graduation and have, since then, trusted me and guided me, helping me to improve as a student, as a researcher and also as a person.

I would like to thank Instituto de Telecomunicações and FCT for the opportunity of working as a researcher in the project "3D Video Adaptation based on Perceptual Metrics" (PTDC/EEA-TEL 114487/2009), which supported this dissertation. Also I am grateful to both Instituto de Telecomunicações and Escola Superior de Tecnologia e Gestão of the Instituto Politécnico de Leiria for providing the facilities and equipment required to accomplish this work.

To my colleagues at Instituto de Telecomunicações, João Carreira, Sylvain Marcelino, Luís Lucas, Lino Ferreira, Nelson Francisco, Anderson Oliveira, Bruno Feitor and Ricardo Monteiro I have to thank for the great work environment, ruled by friendship and mutual help. Also I am deeply thankful to everyone of my friends outside the research environment who have always supported me and provided memorable experiences.

A special and enormous thanks goes to my parents, Manuel Pinto and Isaura Pinto, for supporting me all me life and to whom I own everything I am today. I would also like to express my gratitude to my family, with a special mention to my uncles for providing me a second home while away from my first. I am also very lucky to have Ana Luísa Pedrosa by my side, to whom I must thank the love and support she has given me.

Finally I would like to thank to everyone that participated in the subjective assessment tests conducted during this work.

Abstract

This dissertation presents a study and experimental research on asymmetric coding of stereoscopic video. A review on 3D technologies, video formats and coding is first presented and then particular emphasis is given to asymmetric coding of 3D content and performance evaluation methods, based on subjective measures, of methods using asymmetric coding.

The research objective was defined to be an extension of the current concept of asymmetric coding for stereo video. To achieve this objective the first step consists in defining regions in the spatial dimension of auxiliary view with different perceptual relevance within the stereo pair, which are identified by a binary mask. Then these regions are encoded with better quality (lower quantisation) for the most relevant ones and worse quality (higher quantisation) for the those with lower perceptual relevance. The actual estimation of the relevance of a given region is based on a measure of disparity according to the absolute difference between views. To allow encoding of a stereo sequence using this method, a reference H.264/MVC encoder (JM) has been modified to allow additional configuration parameters and inputs. The final encoder is still standard compliant.

In order to show the viability of the method subjective assessment tests were performed over a wide range of objective qualities of the auxiliary view. The results of these tests allow us to prove 3 main goals. First, it is shown that the proposed method can be more efficient than traditional asymmetric coding when encoding stereo video at higher qualities/rates. The method can also be used to extend the threshold at which uniform asymmetric coding methods start to have an impact on the subjective quality perceived by the observers. Finally the issue of eye dominance is addressed. Results from stereo still images displayed over a short period of time showed it has little or no impact on the proposed method.

Keywords: Stereoscopic asymmetric coding, regions of perceptual interest, subjective assessment.

Resumo

Esta dissertação apresenta um estudo e investigação experimental sobre codificação assimétrica de vídeo estereoscópico. Uma revisão sobre tecnologias 3D, formatos de vídeo e codificação é apresentada primeiro e, de seguida, é dado um enfâse particular à codificação assimétrica de conteúdo 3D e métodos de avaliaçã do desempenho, baseado em medidas subjectivas, de métodos que usam codificação assimétrica.

O objectivo da investigação foi definido como sendo uma extensão dos conceitos actuais de codificação assimétrica de vídeo stereo. Para alcançar este objectivo o primeiro passo consiste em definir regiões na dimensão espacial da vista auxiliar com diferentes relevâncias perceptuais dentro do par stereo, que são identificadas através de uma máscara binária. Depois essas regiões são codificadas com melhor qualidade (menor quantização) nas mais relevantes e com pior qualidade (maior quantização) naquelas com menor relevância perceptual. A estimativa da relevância de uma dada região é baseada numa medida de disparidade de acordo com a diferença absoluta entre vistas. Para permitir a codificação de uma sequência stereo usando este método um codificador H.264/MVC de referência (JM) foi modificado para permitir a configuração de parâmetros e entradas adicionais. O codificador final mantém a compatibilidade com os standards.

De modo a demontrar a viabilidade do método foram realizados teste de avaliação subjectiva sobre uma gama alargada de qualidades objectivas da vista auxiliar. Os resultados destes testes permitem-nos provar 3 objectivos principais. Primeiro, mostra-se que o método proposto pode ser mais eficiente que codificação assimétrica tradicional quando se codifica vídeo stereo com maior qualidade/débitos. Este método pode ser também utilizado para extender o limite a partir do qual métodos de codificação assimétrica uniforme começam a ter impacto na qualidade subjectiva percebida pelos observadores. Por último o assunto do olho dominante é abordado. Resultados de imagens stereo apresentadas durante um curto período de tempo mostram que tem pouco ou nenhum impacto no método proposto.

Palavras chave: Codificação assimétrica de vídeo 3D, regiões de interesse perceptual, avaliação subjectiva.

Contents

List of Figures

List of Tables

List of Abbreviations

2D	Two-dimensional, p. 7
3DTV	Three-Dimensional Television, p. 1
3D	Three-Dimensional, p. 1
ACR	Absolute Category Rating, p. 40
AVC	Advanced Video Coding, p. 2
BD	Bjontegaard delta, p. 24
DIBR	Depth-image-based randering, p. 15
DSCQS	Double Stimulus Continuous Quality Scale, p. 43
DSIS	Double Stimulus Impairment Scale, p. 41
EPZS	Enhanced Predictive Zonal Search Algorithm, p. 62
FVP	Free-view point, p. 5
GOP	Group of pictures, p. 36
HDTV	High Definition Television, p. 1
HD	High Definition, p. 2
HEVC	High Efficiency Video Coding, p. 2
HVS	Human Visual System, p. 2
ITU	International Telecommunication Union, p. 2
LDV	Layered depth video, p. 17
MOS	Mean Opinion Score, p. 3
MPEG	Moving Picture Experts Group, p. 22

MVC	Multiview Video Coding, p. 2
MVD	Multiview-plus-depth, p. 17
MVP	Multi View Profile, p. 22
MVV	Multi-view video, p. 33
Mbps	Megabits per second, p. 15
NAL	Network Abstraction Layer, p. 32
POV	Point of view, p. 5
PSNR	Peak signal-to-noise ratio, p. 24
PVD	Preferred viewing distance, p. 41
QP	Quantisation Parameter, p. 30
R-D	Rate-distortion, p. 2
ROI	Region of interest, p. 36
RTP	Real-time Transport Protocol, p. 53
SAMVIQ	Subjective Assessment of Multimedia Video Quality, p. 45
SD	Standard definition, p. 41
SEI	supplemental enhancement information, p. 11
SNR	Signal-to-noise ratio, p. 33
SSCQE	Single Stimulus Continuous Quality Evaluation, p. 44
SSIM	Structural Similarity, p. 71
SVC	Scalable video coding, p. 31
TS	Transport stream, p. 27
V+D	Video plus depth, p. 15
dB	Decibel, p. 24

Chapter 1

Introduction

In this chapter an introduction about the research work conducted in this dissertation is given. The motivation for the work is presented as well as the objectives that were proposed to be achieved. In the outline section the structure of the dissertation is briefly described.

1.1 Motivation

The human perception of the real world is three-dimensional (3D) and necessarily includes information about volume and depth, which is not explicitly included in classic representations of natural scenes using digital multimedia signals. Since 3D video is the most common type of media content used to provide 3D immersion experience, this is also a driving factor for technology evolution in several domains, ranging from high resolution 3D cinema and 3D Television (3DTV) to small screen applications (e.g. games) using auto-stereoscopic displays.

Television is seen as an invaluable service in nowadays' society and has ever been subject to a fast-paced evolution. While high definition television (HDTV) broadcast is already a reality 3DTV is expected to shortly follow, taking its place as the natural next step in television services evolution. The consumer market is also being driven by the ever increasing diversity of devices and applications able of handling stereoscopic (two view content) and multiview content [1]. This recent evolution towards multiple view video coding and transmission imposes huge demand of both bandwidth and storage capacity in order to fulfil the requirements of forthcoming multimedia services and applications. Also, the diversity of content and devices with huge differences in characteristics (processing power, display, etc.) require new techniques that are able to adapt encoding of the content to better fit a wide variety of characteristics. All this leads to the need of standard representation formats in both uncompressed and compressed domains in order

to efficiently deal with the huge amounts of data required to represent 3D scenes.

Recent evolution of 3D media services along with increasing penetration of 3D-ready equipment in the consumer market lead to co-existence of emerging 3D systems with legacy ones. Several 3D representation formats are currently used to enable efficient coding for storage and transmission across interoperable systems also enabling operation with equipment in different technological evolution stages either in the segment of professional or consumer market.

The downside of stereoscopic and, especially, multiview 3D signal is the huge amount of bandwidth and storage capacity required to fulfil the demands of forthcoming multimedia services and applications. Also with the increasing diversity of devices, that differ in characteristics such as processing power and display, new techniques are required to be able to adapt encoding of the content to better fit that wide variety of characteristics.

The latest high-definition (HD) 3D video codecs such as the H.264/MPEG-4 advanced video coding (AVC) [2] [3] brought significant advances, increasing compression efficiency up to 50% compared to its predecessors. This advances were then extended to 3D (stereo and beyond) with the development of the multiview video coding (MVC) extension [4]. Once again compression efficiency was increased up to 50% (or 3dB) by jointly compressing the 2 (or more) views into a single bitstream instead of doing simulcast (separate bitstreams for each view) coding [5]. All of this is done while also achieving reduced decoding complexity in the receiver side.

On top of the previous advances, more recently a new standard for high efficiency video coding (H.265/HEVC) [6] has emerged, introducing some new coding tools, which further increase the compression efficiency in comparison to the previous H.264/AVC standard. Both objective and subjective tests [7] have shown that the compression efficiency can be doubled for the same quality, in comparison with H.264/AVC. Although decoding has similar complexity as the previous standard, encoding has become more demanding in processing power due to the higher complexity introduced by the new coding tools. HEVC has also been extended to cope with multiview video coding (MV-HEVC), as described in Annex F of Recommendation ITU-T H.265 [6]. Taking advantage of the existing inter-view redundancy it will enable 3D video coding at approximately half the rate of H.264/MVC.

However, despite all the reported advances, very high bitrates are still required which still motivates further research on methods to increase compression efficiency without penalising perceptual quality. This can be done by taking advantage of the limitations of the human visual system (HVS), in example, by exploiting the characteristics presented in the binocular suppression theory [8]. This has been studied in the past using mixed resolution in stereo images [9], where experimental results have shown that the HVS does

not notice the absence of high frequency information in one of the stereoscopic views. Moreover, when the left and right views have different sharpness, the perceived quality is near to that of the higher quality view [10]. This means that, within some quality bounds, one of the two stereoscopic views can be coded at a lower rate-distortion (R-D) point while still maintaining the overall 3D video quality perceived by the human brain.

In this dissertation we explore the hypothesis of extending the asymmetric coding presented in the literature by defining regions in a stereo pair of frames with different relevance to the quality perceived by observers. We aim to do this by encoding regions of the auxiliary view with greater relevance with higher fidelity and vice-versa. Meaning that regions of the frame considered to be less relevant to the end quality will be coarsely encoded. This method aims to extend traditional asymmetric coding by adding another spatial dimension to the asymmetry on the auxiliary view.

1.2 Objectives

As seen before, in 3D video it is of utmost importance to conduct research studies where the HVS characteristics are considered. The most feasible way to do this is to evaluate performance through subjective assessment, which means that the methods under study are subject to screening by the observers, in similar conditions to those of the real application. The quality evaluation from the observers results in a performance measure called mean opinion score (MOS). In this work, we exploit the previously stated asymmetric properties of the HVS by defining a new asymmetric coding scheme for stereoscopic video based on image regions with different perceptual relevance. Using mainly subjective assessment we aim to prove the following goals:

- i) using asymmetric coding based on regions with different perceptual relevance, for certain coding regions and for a given subjective quality it is possible to reach higher compression gains when compared to traditional uniform asymmetric coding methods (i.e., the whole image has roughly the same quality).
- ii) it is possible to efficiently extend the range of asymmetry obtained by uniform asymmetric coding using this technique and without compromising the perceptual quality delivered to the end-user.
- iii) observers with eye dominance corresponding to the lower quality view are not significantly more affected than those who have eye dominance corresponding to the higher quality view.

1.3 Outline

After this brief introduction, the next chapter presents a review of the state of the art where 3D video formats coding techniques and technologies are introduced and explained. Then we look into the literature to review asymmetric coding related works, by analyzing them in order to establish a good starting point for this work. Finally the main methods to perform subjective assessment of multimedia content, especially 3D video, are presented and compared.

In chapter 3 the implementation of both uniform and non-uniform asymmetric coding using regions of perceptual relevance is explained. The modified encoder and new functionalities are presented in detail, addressing the main functions, structures and routines that need to be changed in order to implement our method. The decoding side and also the overall context and possible interactions between the two ends of the codec are presented.

In order to validate the method developed in this work, subjective tests are performed using different content. In chapter 4 subjective quality tests using both our asymmetric encoding method and traditional uniform asymmetric coding are done in order to reach a comparison between the two. The assessment framework is explained and the results of both methods are compared. Discussion of the results and conclusions are presented at the end of the chapter.

In chapter 5 similar subjective tests are conducted but in this case using moving stereo video sequences. The results are presented and discussed, leading to new conclusions which show that our method can also be used to extend the current limits of asymmetric coding in stereoscopic video.

Global conclusions are presented in chapter 6, while revisiting the chapters and contributions of the dissertation. Finally proposals for future work are suggested and discussed.

Chapter 2

State of the Art

This chapter presents an overview of the state of the art regarding 3D video related technologies. The information and knowledge presented here is part of the theoretical study and provides the essential background for the research work developed in this dissertation. A brief overview of the technologies associated with 3D is given, followed by a presentation of 3D video formats and coding. Then the principles behind asymmetric coding are presented as well as the different methods used for achieving it. Finally, methods to assess the quality of the encoded 3D content are described, with particular focus on those which are more relevant for this research study.

2.1 Three-Dimensional Applications

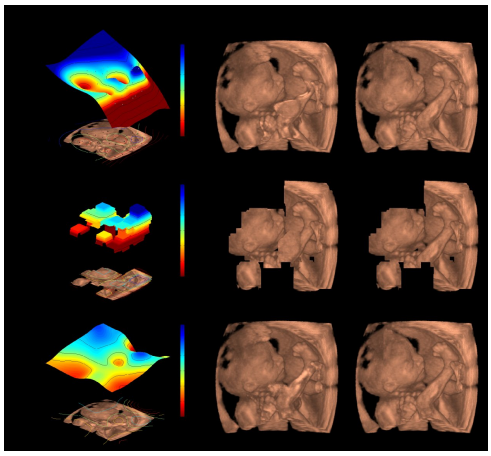
There is a panoply of applications that offer 3D immersion to the user. The most popular are 3D cinema and 3DTV but one can also find other examples both for entertainment or work/scientific purposes. The range of possible 3D applications is large and includes, among other, the following:

- Television (figure 2.1a);
- Cinema;
- Photography.
- Video games (figure 2.1b);
- Sports point of view (POV) capture (figure 2.1c);
- 3D microscopy - i.e. for quality inspection of very tiny pieces;
- Medical examination (figure 2.1d);
- Realistic simulations/training - i.e. pilots training;
- Free-view point (FVP) television;
- Holography.

(a)

(b)

(c)



(d)

Figure 2.1: Examples of 3D applications. 3DTV (a), gaming (b), sports (c) and medical exams/teaching (d).
Source: ready-up.net, selectgame.com.br, provideocoalition.com and cg.tuwien.ac.at

In Figure 2.1 some of these applications are shown. In Figure 2.1a a 3D-ready television is shown, using a technology where the user needs to wear appropriate glasses in order to experience 3D immersion. A portable gaming console is shown in Figure 2.1b, a new field (portable gaming) of entertainment that has been expanding following the success of 3D technologies in gaming. With the miniaturization of the technology, namely video cameras, the availability of coupled 2D video cameras to acquire stereoscopic video is becoming quite common in the consumer market. In Figure 2.1c two HD sports cameras are used to achieve a high quality stereo video and, finally, in Figure 2.1d a 3D model that can be used, i.e., for medical examinations or training is shown.

2.1.1 Technologies: Acquisition

The capture of 3D content can be done using several different types of equipment and data formats. The easiest way to capture stereoscopic video is to use a stereo camera such as the one in Figure 2.2a. By having two traditional 2D cameras side-by-side the capture of stereo video consists in two concatenated individual video streams, corresponding each one to the view of each eye. This is a user-friendly operation, as the user does not have a lot of control on the calibration mechanism and even a naive person in terms of 3D capture can operate such equipment. However, this can also be a drawback if the user wants to change the parameters of the camera pair. Also, this setup does not allow to capture more than two views.

In order to capture a larger amount of different views or to calibrate the system one should use a setup of 2D cameras with the possibility of being adjusted. This kind of setups can range from a simple 2-camera system on a rig, with adjustable baseline and other parameters, as seen in Figure 2.2b, to a more complex system as shown in Figure 2.2c. This requires a lot more technical expertise but also enables acquisition of new kinds of 3D representation such as plenoptic imaging.

One can also opt to only capture a traditional 2D video but using additional information such as the depth of the scene to be able to represent it in 3D later. In that case a simple commercial device such as the Kinect camera (Figure 2.2d) provides all that is needed for the capture through a 2D camera plus an infrared projector of structured lighting and an auxiliar camera for depth sensing.

2.1.2 Technologies: Display

Currently there is some diversity when choosing a 3D display capable of dealing with two views or a multitude of views. The traditional 3D stereoscopic displays (it may be a TV or a projector system) output a total of two stereoscopic views, each one corresponding

(a) (b)

(c) (d)

Figure 2.2: Examples of different 3D capture methods. Stereo camera (a), setup of 2D cameras (b), array of cameras (c) and video plus depth camera (d).
Source: ubergizmo.com, cheesycam.com, cardinalphoto.com and commons.wikimedia.org

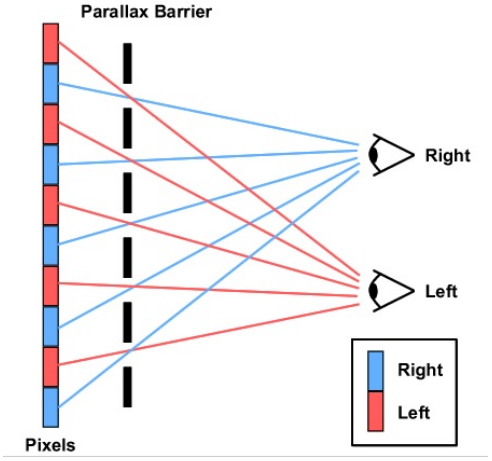
to an eye. This type of displays must rely on a system that must be capable of directing each view to the corresponding eye which is usually accomplished by the use of proper filtering glasses. These glasses can be passive, ranging from the old anaglyph models to wavelength selective lenses working in conjunction with filters on the display screen, or active (also called shutter glasses) using, i.e., infrared connections to synchronize the view displaying with the glasses lenses. These lenses are normally liquid crystals and when a view is being displayed the lens corresponding to the opposite eye receives an electric signal to close it and becomes opaque. Figure 2.3a shows the Nvidia branded example of this type of glasses that were used in the subjective quality assessments performed in this dissertation. Although this technology enables high degrees of parallax and a good level of 3D immersion, it has the drawback of requiring the use of glasses, which besides not being comfortable (especially active ones due to weight and batteries), they are expensive and may not be comfortable to use in a home environment. It works well on other environments such as cinema however. Also either spatial or temporal resolution need to be doubled in order to allow displaying both views at their original quality.

An alternative to glasses is the autostereoscopic display, which use techniques to direct each of the two or more views to their respective eye. Parallax barriers (Figure 2.3b) or lenticular lenslets (Figure 2.3c) are examples of how this can be achieved. Although users of this type of display do not have the hindrance of using glasses, they also need to find the sweet point to watch the 3D content. This means that the user must be able to place himself at a spot where each eye is correctly aligned with the respective view, as shown in the figures. There's also the problem of the convergence and accommodation that may lead to eye strain and, in some cases, greater discomfort. This happens because the eyes converge on a virtual object that is located on a plane other than the one of the screen but they accommodate at the screen plane in order to focus the object.

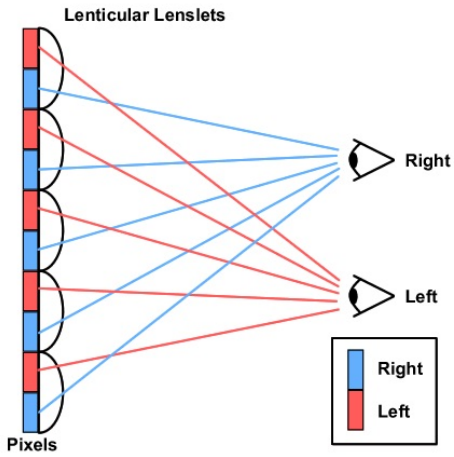
More recently displays having full parallax (horizontal and vertical) have appeared. These are autostereoscopic as well, not requiring the use of glasses, but rely on the display of an enormous amount of views, that may be achieved through synthesis. These can be integral imaging displays or volumetric displays (see Figure 2.3d). Integral imaging is based on the capture of the light spectrum thus providing high quality synthesis of views. An integral display has a 2D array of microlenses directs light from the pixels it covers in both horizontal and vertical directions. Although the viewing angle and depth of field are limited and the fabrication of the microlenses is still a challenge, the visual experience is a lot closer to the of watching the real world. In the case of volumetric displays light is emitted in order to occupy a volume in space, such as a spinning helix.

For a more complete overview of 3D displays and their applications one should refer to articles in the literature such as the one by Holliman *et al.* [11].

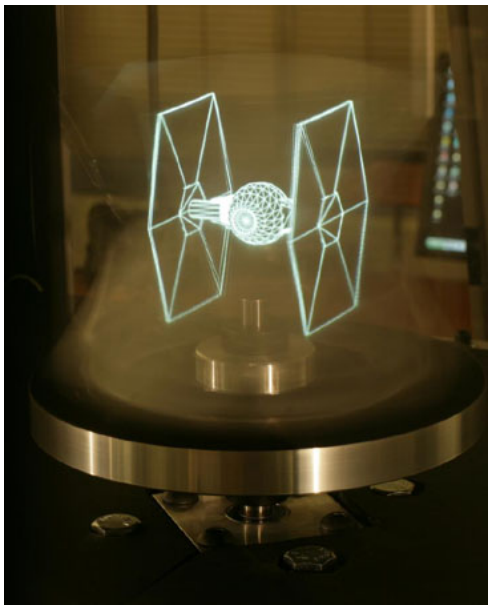
(a)



(b)



(c)



(d)

Figure 2.3: Examples of different 3D displays. Nvidia shutter glasses (a), autostereoscopic parallax barrier (b) and lenticular lenslet (c) display and volumetric display (d).
Source: 3d-forums.com and 3dcinecast.blogspot.pt

2.2 3D Video Representation Formats

2.2.1 Frame Compatible Formats

Several 3D representation formats are currently used to enable efficient coding for storage and transmission across interoperable systems, also enabling operation with equipment in different technological evolution stages either in the segment of professional or consumer market.

In the context of 3D multimedia services and applications, 3D video representation through frame compatible formats is a key factor to guarantee compatibility with existing 2D video networking technology and equipment. Successful deployment of 3D video delivery services and applications is enabled by making possible transmission of 2D-compatible formats over current networks and legacy decoders with 3D-ready displays already common in the user market. By using frame compatible formats, seamless compression of 3D video content is also accomplished with existing 2D encoders without the need to modify the coding algorithms.

In the case of stereoscopic video, representation in 2D compatible formats is achieved by multiplexing the two different views into a temporal sequence of 2D signals. This means merging two different views into a classic sequence of single frame representation. If the full resolution of the two views is maintained, then such representation format has twice the resolution of its equivalent 2D. However, taking into account that good perceived quality of 3D video does not necessarily require two high quality views, either one or both views may be subsampled in one dimension in order to fit two HD frames into only one HD frame slot [12]. Identification of left and right views is done via specific signaling, used to distinguish the data representing each one. Using H.264/AVC to encode 3D frame compatible formats, the recommended signaling method is the use of supplemental enhancement information (SEI) messages, as shown in Table 2.1, where the *frame_packing_arrangement_type* field of the SEI message is defined according to subclause D.2.25 in the standard [2]. SEI messages are used to convey information about how decoders should handle their output according to the frame-packaging scheme used. There is also an SEI value defining 2D format, which enables switching between stereo and non-stereo content. Additionally to the type of content, the SEI message includes other fields such as the arrangement id that can be used to identify which frame is the left or right view.

Side-by-side

The side-by-side format is shown in Figure 2.12 and the respective upsampling process in Figure 2.5. In this format, the two stereoscopic views are concatenated side by side,

Table 2.1: Standard frame compatible formats.

ID	Compatible format
0	Checkerboard
1	Column based interleaving
2	Row based interleaving
3	Side-by-side
4	Top-bottom
5	Temporal interleaving
6	2D
7	Tile format

giving rise to a single 2D matrix of pixels with the same resolution in the vertical direction while in the horizontal one there is twice the number of pixels of each single view.

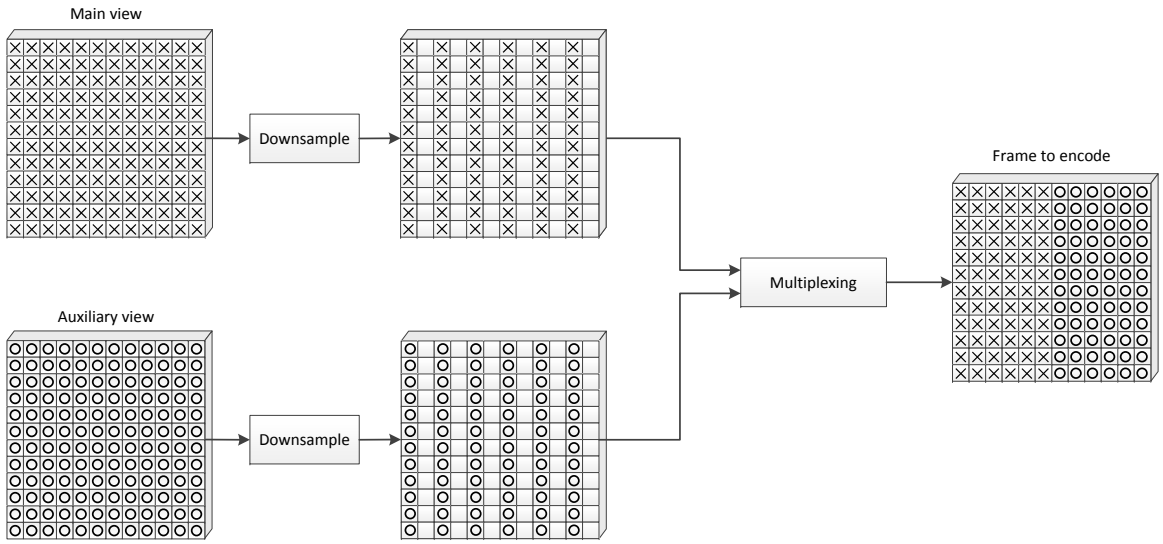


Figure 2.4: Side-by-side packing arrangement.

However, since doubling the spatial resolution has strong implications in compressed rates, the horizontal resolution of the original views might be halved through down-sampling before packing into this side-by-side arrangement for subsequent encoding and transmission. Correspondingly, the counterpart up-conversion process must be done after decoding to display the full resolution stereo images, as shown in Figure 2.5.

A different version of the side-by-side arrangement can be accomplished by sampling the stereoscopic views using a quincunx pattern. In this case, even though the horizontal resolution is also reduced to half of the original, still half of each view columns is maintained, as shown in Figure 2.6. Quincunx sampling relates better with the characteristics of the HVS in terms of frequency domain representation. Thus, this sampling pattern preserves more relevant information from the original signal, which has the potential to

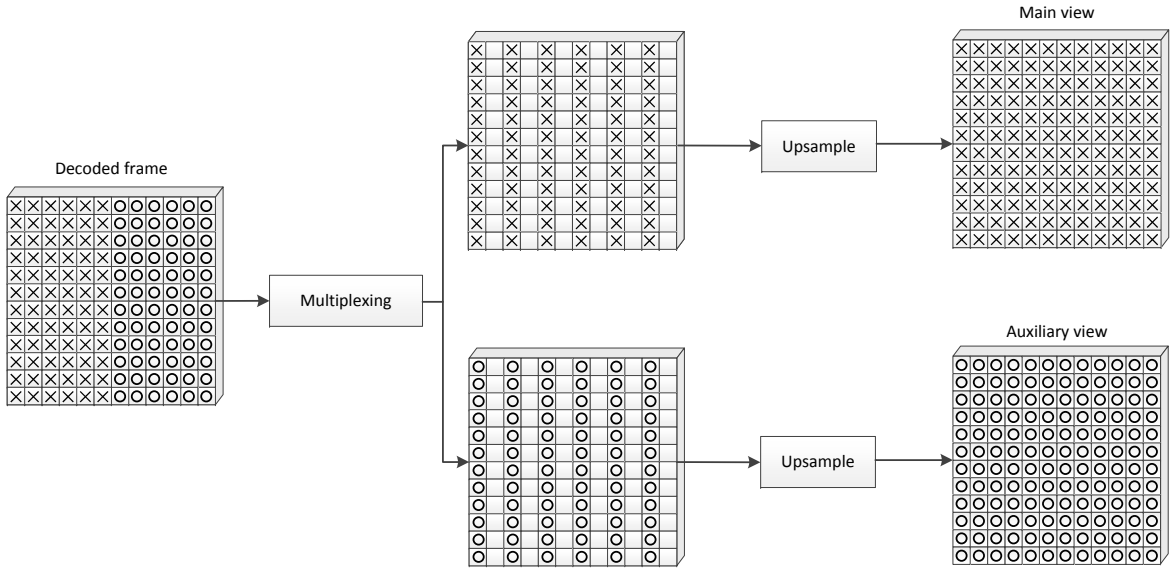


Figure 2.5: Side-by-side up-conversion process.

result in better perceived quality.

Top-Bottom

The top-bottom format is based on the same concept as side-by-side, but in this case downsampling is applied to the vertical resolution and the resulting frames concatenated as shown in Figure 2.7.

Unless otherwise specified by the SEI message, in standard top-bottom format the downsampled left view is concatenated into the first half (top) of a composite frame while the downsampled right view is concatenated into the bottom half. This 3D frame compatible format should not be used with interlaced source material because the vertical resolution of interlaced fields is already half of the full resolution frame and further downsampling in this dimension could incur in too much loss of spatial detail.

Both side-by-side and top-bottom formats are preferred for production and distribution of 3D content in comparison with the ones described next, based on spatial interleaving. This is because interleaving is prone to cross-talk artifacts and color bleeding.

Interleaved formats

Interleaving methods provide higher correlation in the composite frame by multiplexing both downsampled views, either vertically or horizontally according to the downsampled dimension. If both views are half sampled in the horizontal dimension, then a column interleaving arrangement is reached, as shown in Figure 2.8. Otherwise, if downsampling

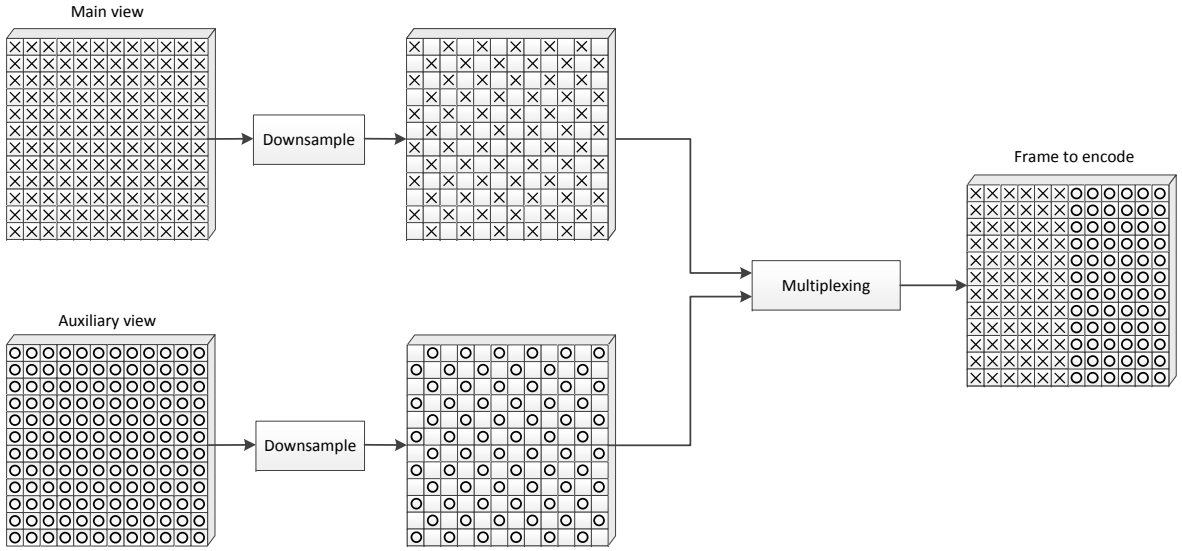


Figure 2.6: Side-by-side with quincunx arrangement.

is performed in the vertical dimension, then multiplexing is done row by row, attaining a row interleaving arrangement for the composite frame.

These interleaving methods can be further combined in order to create a multiplexed frame like checkerboard such as the one depicted in Figure 2.9. In this type of format, each view must be downsampled using checkerboard non-matching patterns. In the case of the left view, this means that in odd rows each other pixel should be kept starting from odd columns, while in even rows each other pixel should kept, starting from even columns. In the case of the right view the complementary pattern must be used.

Other frame compatible arrangement is based on interleaving in the temporal dimension. In this type of interleaving the frame rate of each view is reduced to half of its original rate and then the even frames from the left view are temporally multiplexed with the odd frames from the right view, as shown in Figure 2.10. In this type of format the spatial resolution of the original views is maintained. It can be particularly suitable to represent low motion 3D content, where frame rate is not a very relevant requirement.

Tile format

The last amendment of H.264/AVC in regard to the use of frame compatible formats introduced the tile format [13] [14]. The arrangement depicted in Figure 2.11 allows two HD frames (1280 per 720 pixels) to be packed into a Full HD frame (1920 per 1080 pixels) using a tiling method, where different regions of one view are tiled with the other view.

As seen in the Figure, the left view is located at the top left corner of the Full HD frame without any type of pre-processing. The right view is then divided into three regions (R1,

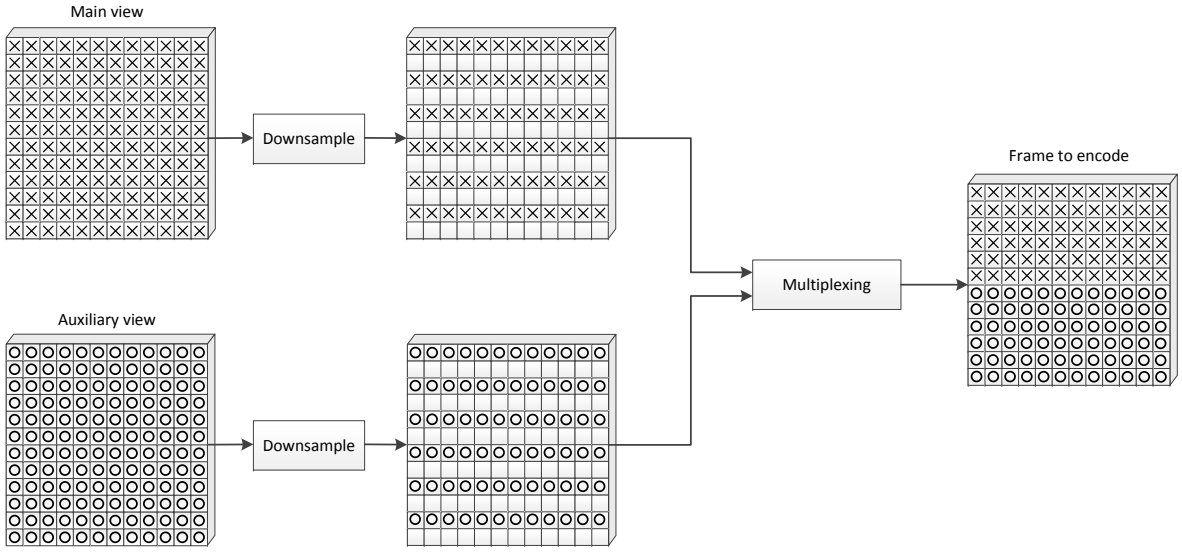


Figure 2.7: Top-bottom arrangement.

R2 and R3), which are placed in specific regions of the resulting Full HD frame.

The great advantage of this method is the backward compatibility with legacy 2D devices, as it requires only a 720p crop to obtain a 2D version of the content in HD resolution. Moreover there are no downsampling operations involved, which means that full resolution of the original frames is maintained in all dimensions.

A potential drawback introduced by this method is lower coding efficiency and annoying artifacts in the coded images due artificial edges created by tiling. However objective tests conducted shown that these are not problematic as impairments are not noticeable (comparing with simulcast), mainly above 1 Mbps [14].

2.2.2 Video plus Depth

An alternative to stereoscopic representation of 3D video consists in two separate 2D signals to convey color image information and the depth associated to each pixel, i.e., the distance to the camera [15]. Such information being available at display side can enable the generation of virtual views through depth-based image rendering techniques. Known as video plus depth (V+D), this format has implicit higher complexity than stereo views because it requires either additional computation to obtain the depth values from multiple views of the scene or specific image acquisition hardware to obtain the depth maps directly from the scene e.g., using hybrid camera systems [16].

Depth values are usually represented as integers in the range of $0 - 255$, thus using 8 bits per pixel which results in a gray-scale depth map, as shown in Figure 8. These values

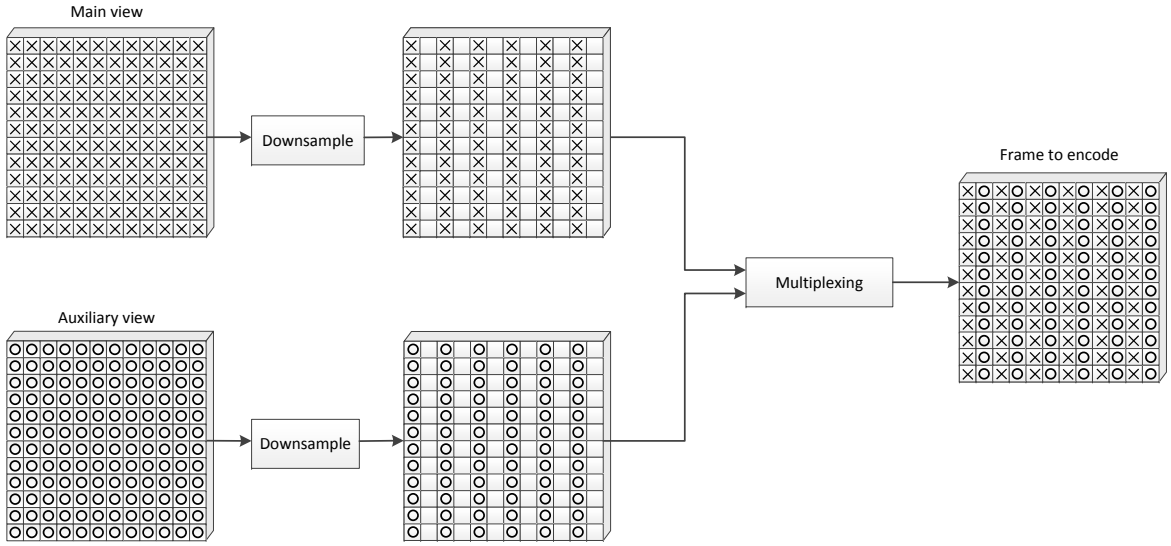


Figure 2.8: Column interleaving arrangement.

translate to the maximum and minimum distance of each point. A warping function should be used to reconstruct a stereoscopic sequence by synthesising the other view of a stereo pair from the color image and the corresponding depth map. Depth-image-based rendering (DIBR) is a method commonly used for this purpose [17]. In view synthesis using DIBR there are some problems that may result in image distortions, such as the possibility of occlusions due to the lack of unique texture data that may be needed to render the other stereo view through the depth map.

Separate encoding of each signal (video and depth) is possible by a standard monoscopic codec, such as H.264/AVC. In regard to encoding the depth map, the grey scale values can be given to the encoder as the luminance component of pseudo video signal where chrominances are set to a constant value. Since the color video is encoded as regular monocular video, this format has inherent backward compatibility with legacy decoders. This format allows extended possibilities at the receiving side compared to the traditional stereo video. For instance, it is possible to adjust the amount of depth perceived by viewers by adjusting view synthesis, or to render several different virtual views for multiview displays.

2.2.3 Multiview video plus Depth

As mentioned before, the V+D format is particularly suited to multiview displays because it enables generation of different virtual views of the same scene. However, if a wide range of views is required the V+D format is no longer suitable because not many different views can be rendered with enough quality from only one view and corresponding depth map.

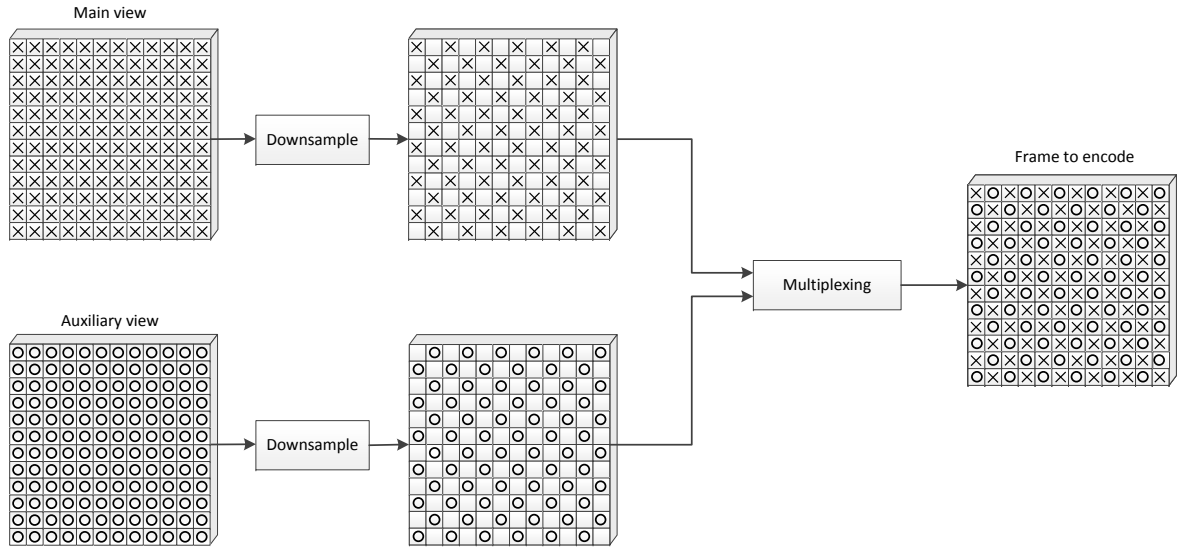


Figure 2.9: Checkerboard arrangement format.

This is because the original view may become farther away than the one to be synthesised producing visible artifacts due to occlusions, which cannot be properly handled in such cases. This is mainly relevant in wide range multiview (autostereoscopic) displays or free viewpoint video applications.

The multiview-plus-depth (MVD) format provides a solution for generating many virtual views by including several views from the same scene, each one with an associated depth map. Based on each pair (Vn,Dn), n=1..N, it is possible to render virtually any intermediate view, giving rise to free viewpoint video. Figure 2.13 shows an example with an autostereoscopic display, where 9 views are generated from only 3 views plus their associated depth maps. These 3 views are actually the only ones available to render the other 6 virtual views [18].

MVD has similar complexity issues as V+D since it also requires depth acquisition/estimation at the sender side and rendering stereo views at the decoder. On the one hand this format allows significant savings in storage and transmission requirements, but on the other hand higher processing complexity is required either in the display side only or in both sides, i.e., acquisition and display.

2.2.4 Layered Depth Video

The layered depth video (LDV) is another 3D video format comprising color images, depth maps and an additional layer providing occlusion information, which is used to fill up occluded regions in the rendering process. Such additional layer contains texture information from those regions in the scene that are occluded by foreground objects in

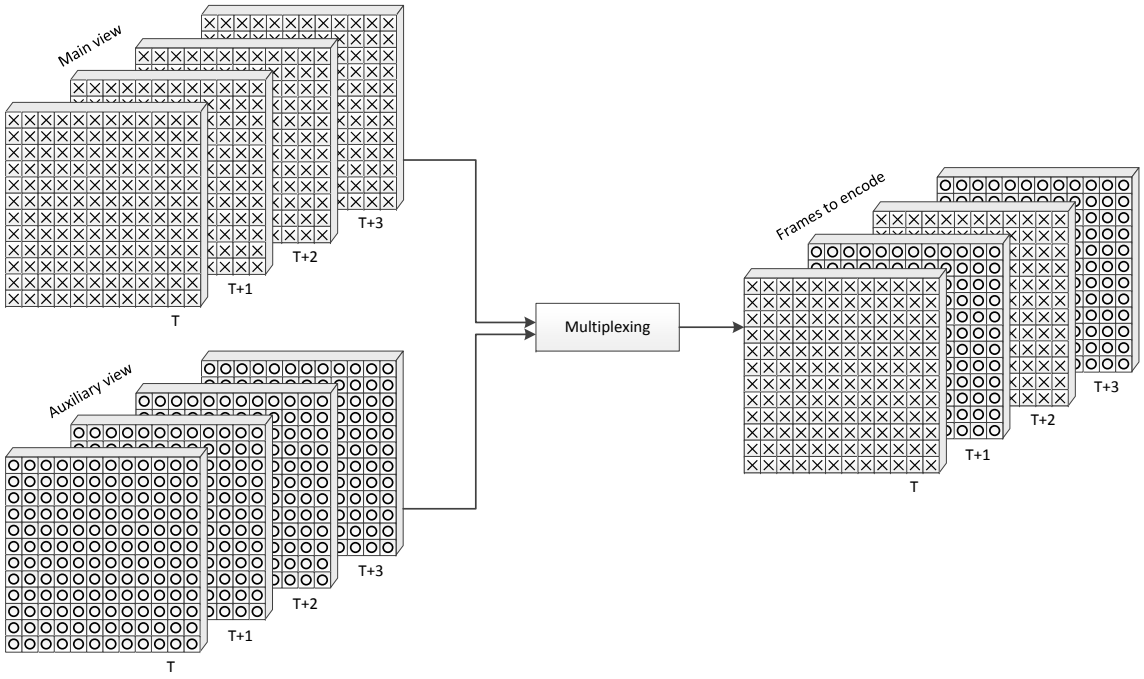


Figure 2.10: Temporal interleaving frame arrangement.

the available view. Using this format, rendering of virtual views benefits from layered information because it includes the necessary data to synthesise virtual views, which otherwise would be missing. In LDV it is also possible to represent residual layers to include visual data that is not available in the main view but visible from other viewing directions. Figure 2.14 shows an example of a color image and its depth map (top) along with an occlusion layer and depth map (bottom).

A variant of LDV is known as depth-enhanced stereo (DES), which basically includes two sets of LDV data [18]. Since LDV is a extension of MVD, its inherent computational complexity is higher than MVD due to the operations (warping, matrix subtraction/addition, etc) that are necessary to obtain the residual layers. On the receiver side, rendering the extra views using LDV has similar computational complexity as MVD. Besides the ability to better cope with occlusions, LDV has also the advantage of requiring a smaller amount of data than MVD for multiview 3D video representation. However, since LDV relies on residual data, the potential impact of transmission errors or data loss is greater than in MVD.

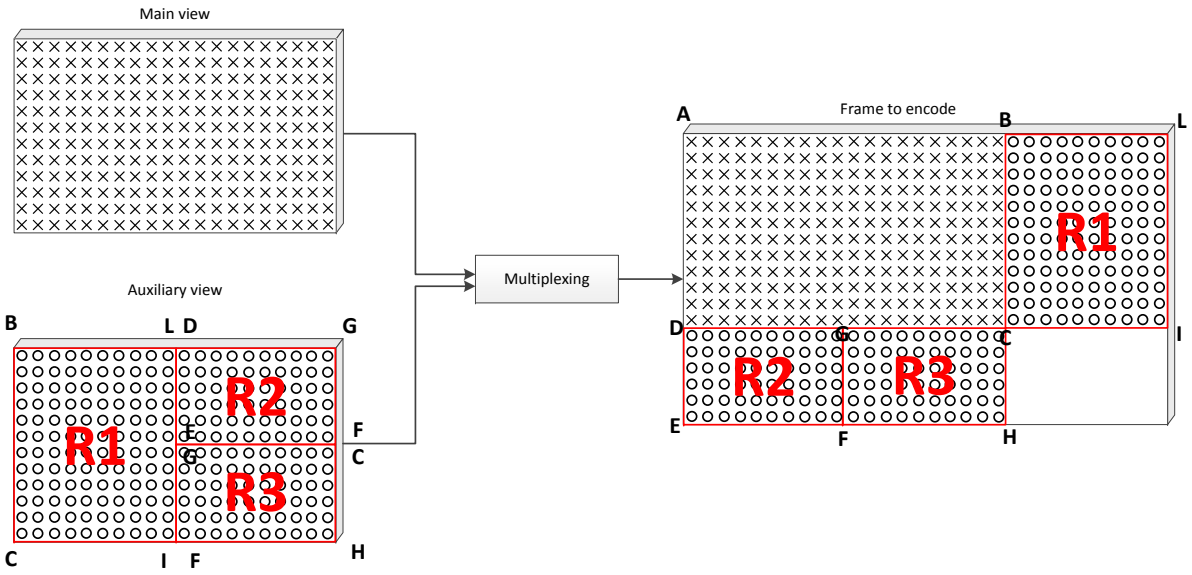


Figure 2.11: Tile frame arrangement.

Figure 2.12: Sequence Breakdance: video (left) plus depth (right).

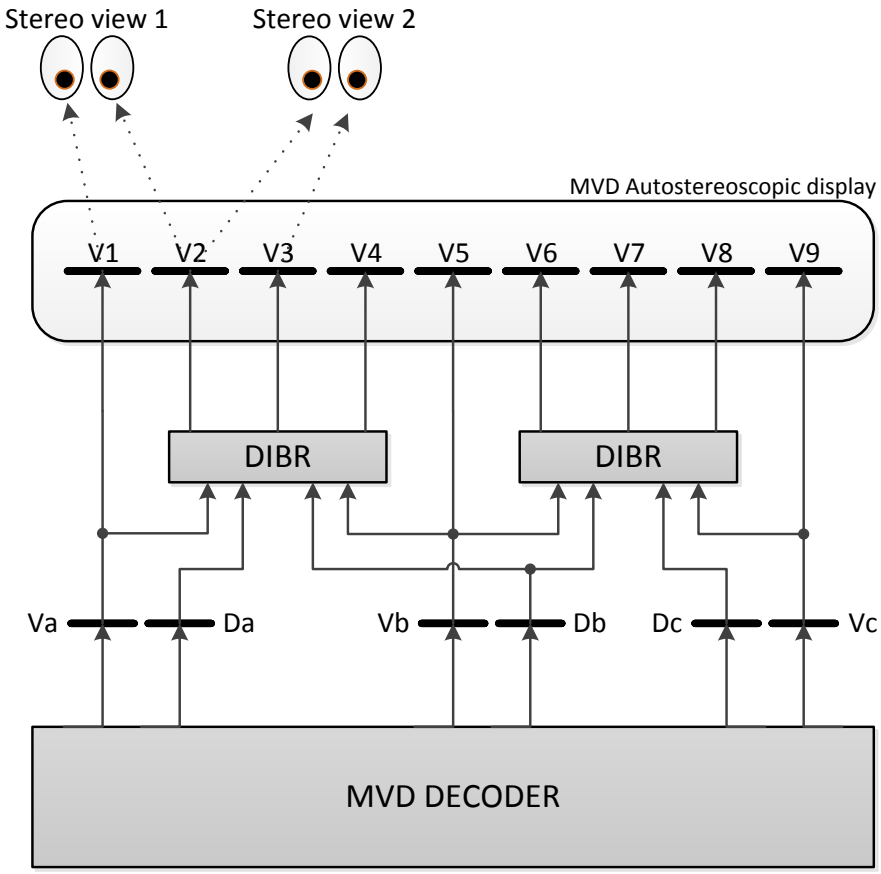


Figure 2.13: Use of MVD format.

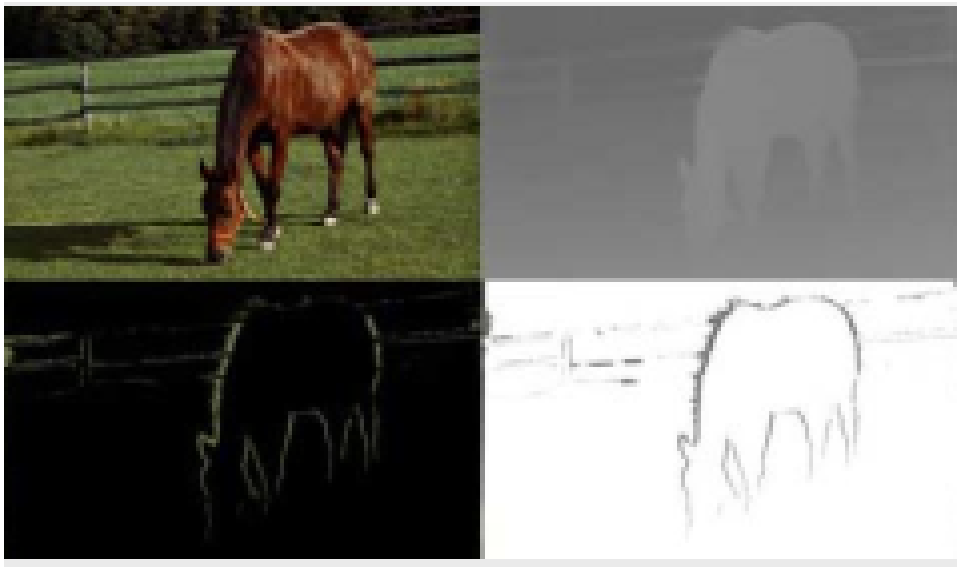


Figure 2.14: Example of LDV format.

2.3 3D Video Coding

Current 3D video technology is mostly based on stereo systems, but this is rapidly evolving to include more information, that may be either more views and/or depth information. These additional data is used to feed 3D systems that are able to provide richer immersive experiences to users. Regardless of the 3D video format used in such systems, these data have to be encoded, in order to fulfill the application and services requirements and to achieve useful compression ratios. Hence, due to the multi-dimensional nature of this content, it can be either jointly encoded as a whole, by exploiting their correlation, or separately as a set of independent sources.

2.3.1 Simulcast

Simulcast refers to independent encoding and transmission of several views and possibly their corresponding depth maps, using any encoder to encode each data sequence. To simulcast stereo or multiview video, each view (Left, Right) is independently encoded without using any type of interview prediction, as can be seen in Figure 2.15. Any standard video encoder such as H.264/AVC [2] or HEVC [6] can be used for this purpose. In this case, the complexity and processing delay is kept low, since dependencies between views are not exploited, and backward compatibility with 2D systems is maintained by decoding only one view.

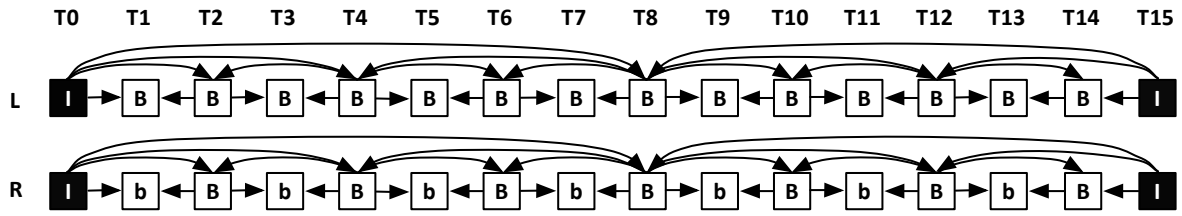


Figure 2.15: Simulcast structure.

However, a simulcast solution has a drawback in the coding efficiency, as it does not exploit the inter-view redundancy. In this sense, studies on asymmetric video coding suggest that one view may be encoded with less quality than the other, with significant bit rate savings. Such scheme may be implemented by means of coarse quantization or by reducing the spatial resolution [9]. This can be achieved without loss of the stereo perception, but when played for long periods, the unequal quality for each eye may cause eye fatigue. To overcome such effect the toggling of quality between both views has been suggested [19]. A more detailed description of asymmetric coding is presented in section

4 of this chapter.

2.3.2 Stereo and Multiview Video Coding

MVC comprises joint coding of two or more views of the same scene. When only two views are allowed this is named as stereo video coding. Figure 2.16 shows an example of a stereo video sequence and frame coding dependencies regarding the right (R) and left (L) views. The left view is independently encoded to ensure compatibility with 2D video systems, while the right view uses interview prediction from the left one achieving higher coding efficiency at the cost of greater coding complexity and dependency.

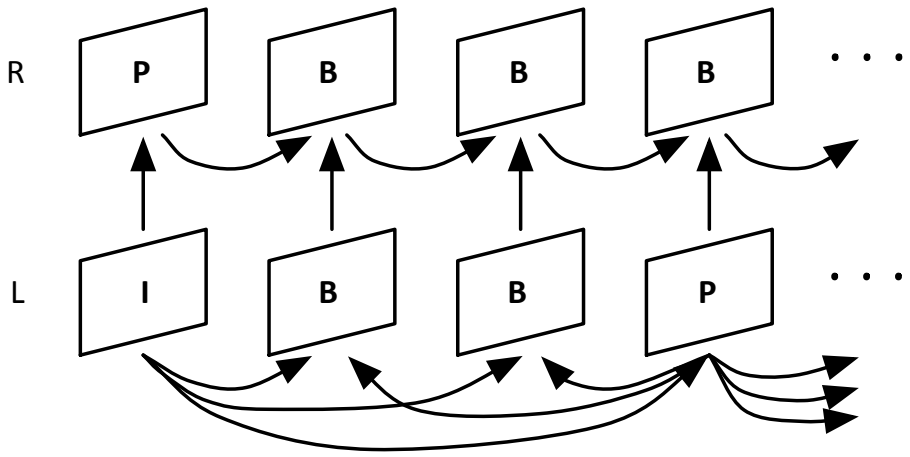


Figure 2.16: Prediction structure in stereo video.

The first standard for multiview applications was the extension of the MPEG-2 MVP [20] (Multi View Profile), which has been approved as International Standard in 1996 when it was envisioned to be a profile appropriate for applications requiring multiple viewpoints. The underlying coding principles used in this codec are mostly the same as those currently used in more advanced ones. The general architecture is depicted in Figure 2.17, where the base layer (left view) is encoded as monoscopic video to maintain compatibility with the MPEG-2 video decoders at the Main profile.

As shown in the functional diagram of Figure 2.17, the enhancement layer (right view) is encoded using hybrid prediction of motion and disparity and temporal scalability tools. By exploiting the similarity between the left and right views, higher compression of the right view was achieved. Both layers have the same spatial resolution at the same frame rate. For MVP, an extension has been introduced specifying the height of image device, the focal length, the F-number, the vertical angle of the field of view, the position and the direction of the camera and upper direction of the camera.

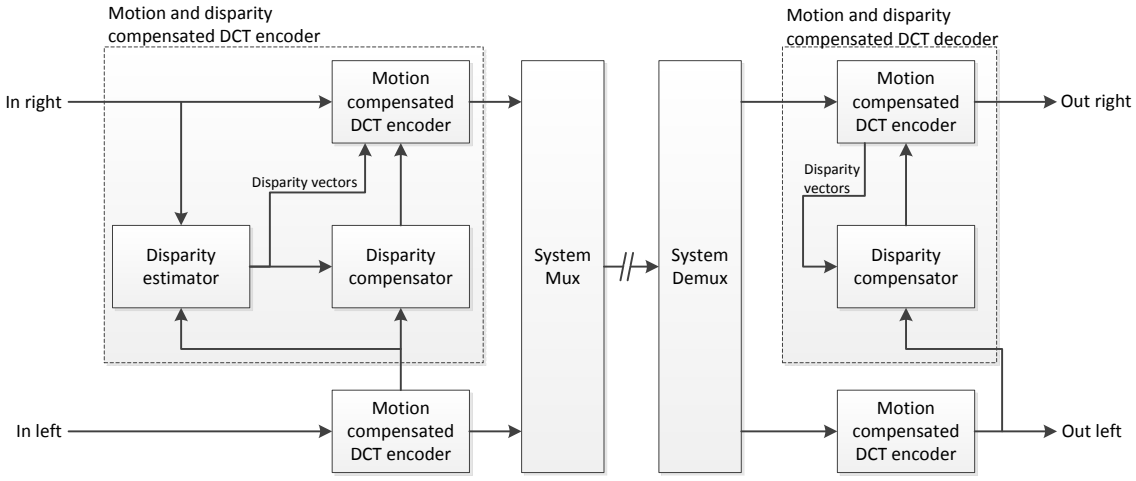


Figure 2.17: Codec reference model for the MVP [20].

A further step towards multiview compression has been done in 2009 with an amendment of H.264/AVC (Annex H) to support MVC with the Multiview High Profile [2]. A typical frame structure and interframe/view coding dependency is illustrated in Figure 2.18. Two types of predictions are explicitly used to achieve increased coding efficiency: intra-view and inter-view prediction. The prediction structure determines the decoding sequence according to the dependencies between frames. Depending on the acquisition arrangement, any two adjacent views may comprise a stereo pair to provide 3D viewing experience.

The standard extension for MVC supports a flexible reference picture management that is used by the inter-view prediction scheme, i.e. the decoded frames from other views are made available for prediction in the reference picture lists. This scheme allows a particular view to have some blocks predicted from temporal references while others can be predicted from inter-view references. MVC includes a mandatory base view in the compressed multiview H.264/AVC stream, which can be independently extracted and decoded for 2D viewing. To decode other views, information about view dependency is required. As in previous schemes, unequal rate allocation may be used across the views.

In 2010, H.264/AVC was extended with the Stereo High Profile for stereo video coding, with support for interlaced coding tools. In this case, the frame coding structure only stands for two views, as can be seen in Figure 2.19.

Figure 2.20 shows the two MVC profiles, Multiview High and Stereo High, highlighting the fact that a common set of coding tools is used in both of them. When coding a video sequence with two views using non-interlaced tools only, then the coded stream conforms to both profiles.

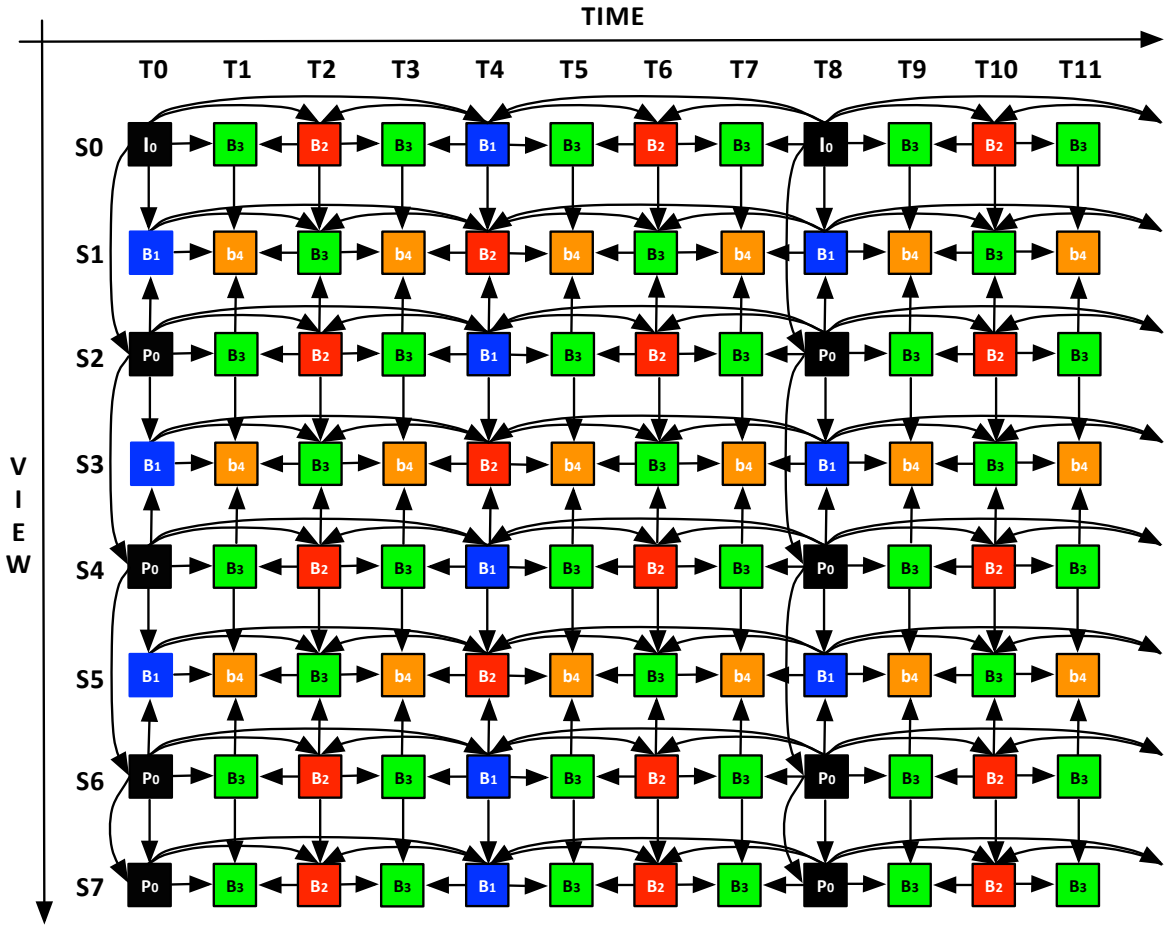


Figure 2.18: Typical MVC frame coding structure.

Coding Performance

A relevant performance metric of stereo video encoding is the amount of additional bit rate required to encode a second view using standard encoders. The authors in [21] carried out a performance study using MVC High Profile to encode 9 HD 3D video clips with various types of content, from animation and live action shots, including progressive and interlaced material. The results of subjective quality evaluation suggest that, as compared to 2D, 20 to 25% bitrate increase would provide satisfactory picture quality in 3D HD video applications. These results show that it is possible to trade-off between the bit-rates of the base and second views for a given total bandwidth. Higher bit rate for the dependent-view preserves better the 3D effect, but in this case fewer bits are left for the base-view. However, it may be more important to preserve the picture quality of the base-view, in order to guarantee good quality 2D viewing.

The performance of the MVC encoder using inter-view prediction tools, against simul-

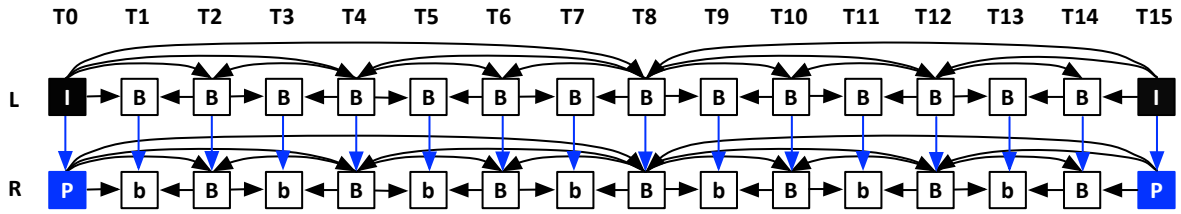


Figure 2.19: MVC stereo high profile frame coding structure.

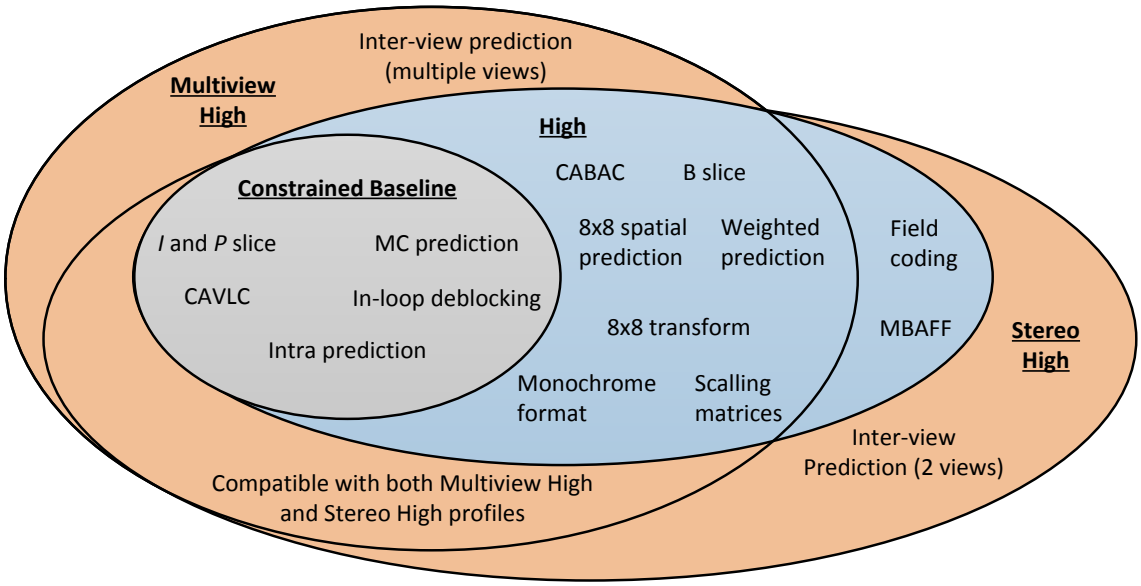


Figure 2.20: MVC profiles and tools associated.

cast has also been tested over a broad range of test material, using the common test conditions and test sequences specified in [22]. The results, given as Bjontegaard delta measurements (BD Bitrate and BD peak signal-to-noise ratio(PSNR)) [23], demonstrate that MVC with up to 8 views can save on average 24% of bitrate (1.2dB in BD PSNR gain) in comparison with the total simulcast bit rate, at the same equal quality in each view [24].

The MVC Stereo High profile also supports stereo coding for interlaced sequences. An experimental performance comparison between simulcast and MVC is presented in [25]. It was found that, for progressive stereo sequences, MVC achieved 9.36% and 11.8% of average bit rate savings for both views, respectively for the stereo MVC test sequences and LG stereo 1080p24 HD progressive sequences. For interlaced video sequences, an average of 6.74% bit rate saving was achieved, for both views. The estimated savings for the dependent view is twice the total saving, that is approximately 20% gain in the

progressive scan sequences and 15% in the interlaced scan sequences.

The coding performance of the MVC Stereo High profile was also compared against AVC High profile (Simulcast). The Stereo High profile achieved an average coding efficiency gain of about 15%, in both views, as compared to simulcast using AVC High profile. For some Hollywood 3D cinema clips, the coding efficiency gain can go up to 21% [26]. Regarding the right view only, the average BD bitrate gain is 30.50% (1.07 dB in BD PSNR). The coding efficiency gain is more pronounced in the animation sequences as the content of live shots very often presents different sharpness, brightness, color, contrast, etc. in the right view, as compared to the left view. This difference reduces the efficiency of the inter-view prediction.

The recently approved standard for high efficiency video coding (H.265/HEVC) [6] has emerged with a new set of coding tools, which have significantly increased the compression efficiency in comparison to the previous video compression standards. Thus, quality evaluation tests demonstrate that HEVC is able to increase the compression ration about 50% at the same quality, in comparison with H.264/AVC. This has also been reported upon subjective tests [7]. However, such increased efficiency is obtained at the cost of higher computational complexity, as HEVC requires 2 to 10 times more computation in the encoder when compared with H.264/AVC. At the decoder side, HEVC presents similar complexity to that of the H.264/AVC [27].

The extension of HEVC to multiview video coding (MV-HEVC) is described in Annex F of Recommendation ITU-T H.265, supporting 3D applications such as stereoscopic television. This extension will enable HEVC-based high quality 3D video coding, at approximately half of the bit rate required by previous services and applications like 3D television and 3D Blu-ray discs. Similarly to the MVC extension of H.264/AVC, MV-HEVC takes advantage of the inter-view predictions tools to exploit the redundancy between views.

Due to its powerful encoding tools, simulcast with HEVC, i.e., each view independently encoded, outperforms H.264/MVC. In the case of H.264, where MVC provides significant bitrate reduction when compared to AVC simulcast, MV-HEVC also outperforms HEVC simulcast [28]. Regarding multiview, MV-HEVC halves the bitrate required by MVC to encode a multiview sequence, in the same proportion as HEVC improves the coding efficiency when compared to H.264/AVC coding of a single view video. Similarly to MVC, MV-HEVC provides backwards compatibility to allow single view decoding. In MV-HEVC, inter-view prediction allows the inclusion of inter-view reference pictures in the reference picture lists that are used for prediction. This is also similar to H.264/MVC.

Video plus depth coding

As previously described, depth-based representations are emerging as an important class of 3D formats, enabling the generation of virtual views through DIB) techniques. Thereby, this format enables display-independent solutions, as different displays may generate more views as required. Although the depth data is not directly output to a display and viewed, maintaining the fidelity of depth information is very important while encoding because it has great influence in the view synthesis quality, due to the geometric information provided by depth. Thus, reaching a good balance between compression ratio and quality of coded depth data is of utmost importance. Note that depth information can be used by encoders to attain more efficient compression, through view synthesis prediction schemes.

The ISO/IEC 23002-3 standard, also referred to as MPEG-C Part 3, specifies the representation of auxiliary video and supplemental information [29]. This is the first standard where signaling of coded depth map is explicitly allowed to support the coded format of V+D. It is worthwhile to notice that this standard does not specify the coding standard that should be used for depth and video information, which allows compatibility with any legacy receiver. The use of MPEG-C Part 3 is illustrated in Figure 2.21a with two H.264/AVC encoders generating two streams (BS), one for video and another for depth. Then these streams are multiplexed using frame-by-frame interleaving before encapsulation into a single Transport Stream (TS). At the receiving side these streams are demultiplexed and independently decoded to produce the output video and depth signals, which in turn are used to generate the second stereo view.

Another possibility for encoding and transmission of video plus depth is to use the "Auxiliary Picture Syntax" defined in the H.264/AVC standard, which defines that auxiliary monochrome pictures can be sent with the video stream, i.e., the primary coded pictures may be associated with other types of data, jointly encoded but not used for display. This is illustrated in Figure 2.21b where depth is treated as the auxiliary picture of the color view. The H.264/AVC encoder is given both sequences to be jointly encoded, producing one single coded stream (BS/TS). The color images and corresponding depth maps are seamlessly decoded as single pictures and then separated into two different signals for viewing and synthesis of the second stereo view.

As mentioned before, in order to overcome the drawbacks of the 2D plus depth format in generating diverse virtual views, the MVD format allows enhancing the 3D rendering capabilities at a reduced number of transmitted views plus corresponding depth.

In Figure 2.22, a general coding and transmission system for MVD is depicted. A few cameras (2 in the case of Figure 21) are necessary to acquire multiple views of the scene while the depth information can be estimated from the video signal itself by solving for stereo correspondences, or directly provided by special range cameras. Depth may

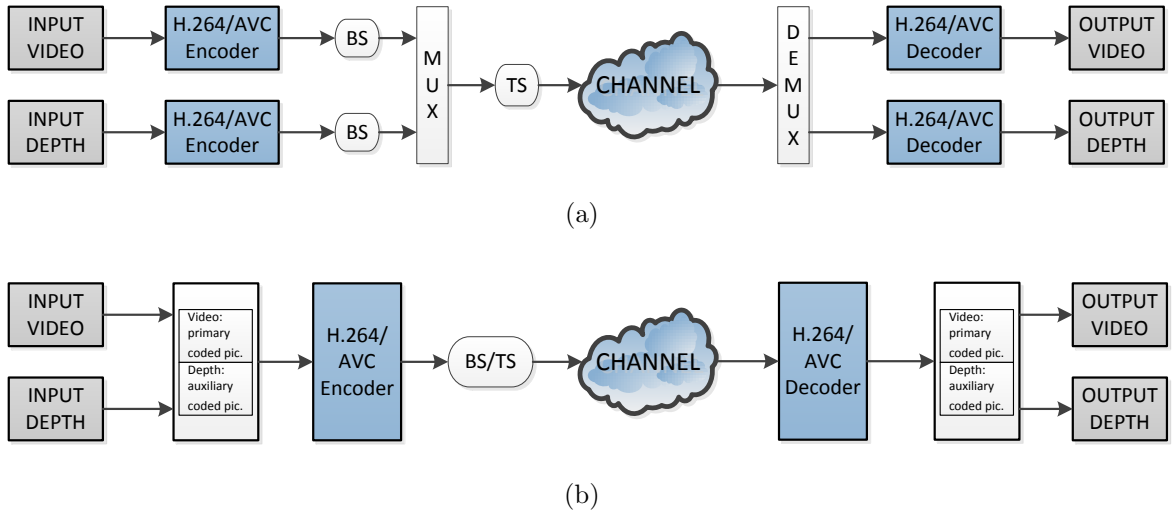


Figure 2.21: V+D coding and transmission system.

also be an inherent part of the content, such as with computer-generated imagery. Both types of signals can be either independently or jointly encoded using any coding schemed previously described. At the receiver, the few decoded views and their corresponding depth maps are used to generate a higher number of virtual views as necessary for each particular multiview service or application.

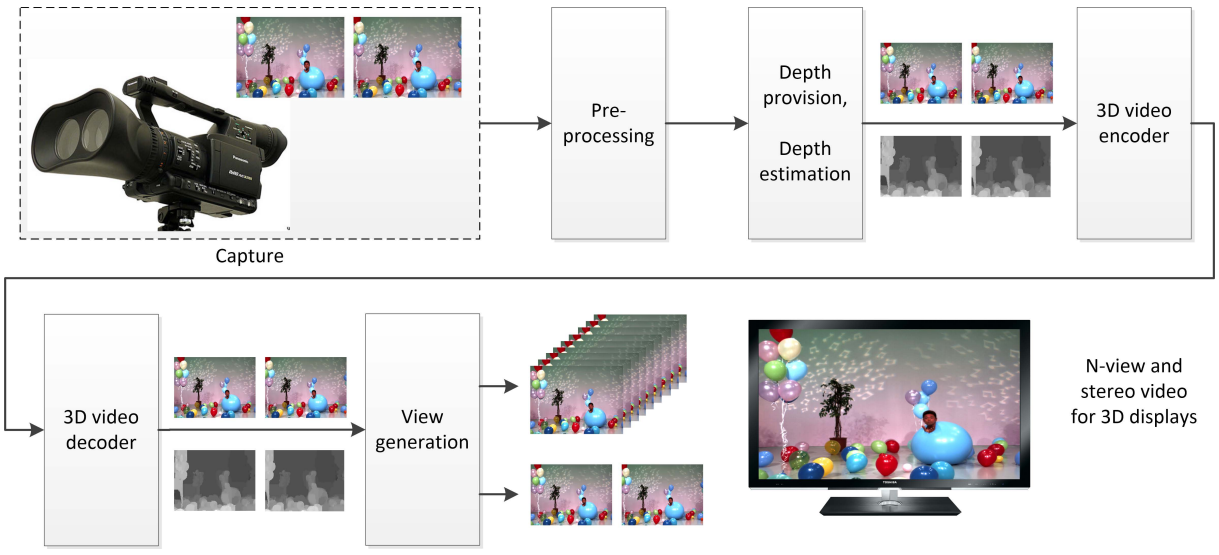


Figure 2.22: Multiview plus depth coding and transmission.

Since depth information mainly consists of larger homogeneous areas and sharp transitions along object boundaries the frequency spectrum of a depth map mostly comprises low and very high frequencies. As a depth sample represents a spatial shift in the color samples of the original views, coding errors result in wrong pixel shifts in synthesized

views, which may lead to annoying artifacts, especially along object boundaries.

Joint video and depth coding is the current path for finding a scheme able to achieve high compression efficiency. For instance, it has been shown that coding efficiency can be significantly increased using scene geometry information such as depth maps [15]. Beyond the use of inter-view prediction techniques, which can be applied to both video and depth independently, there is some block-level information, such as motion vectors, that may be similar for both data types and thus can be shared. In the context of multiview coding, adjacent views may be warped towards reference views in order to reduce the residual error between views. However, since video compression algorithms are typically designed to preserve low frequencies, to maintain the fidelity of edge information in depth maps special coding techniques are required to deal with such particular characteristics.

To evaluate the compression efficiency achieved by the new coding tools implemented in HEVC, several performance studies have been carried out [30]. Different schemes have been compared, namely HEVC simulcast (based on HM 6.0), MV-HEVC (multiview HEVC) and 3D-HEVC (both based on HTM 3.1). MV-HEVC is a simple extension of HEVC, using the same principles of H.264/MVC framework, providing backwards compatibility for 2D video decoding and utilizing inter-view prediction. The 3D-HTM encoder is an extension of HEVC, where the base view is fully compatible with HEVC and the dependent views use additional tools to exploit motion correlation and mode parameters between base and dependent views. Using the common test conditions for 3D video coding (Doc. JCT3V-A1100), for an average of 7 sequences, HTM is able to achieve gains up to 22,7% over MV-HEVC and 47,6% over HEVC simulcast, measured as Bjontegaard delta bit rates.

The 3DV standardization is expected to be finalized by early 2014, which will probably include 3D AVC (based on H.264/AVC) and 3D-HTM (based on HEVC). These schemes allow the choice of MVD, reducing the number of transmitted views and enabling joint encoding of view-depth. It is also possible that single depth and asymmetric frame sizes for view and depth will be supported.

Depth Coding

Extensive research has been done on efficient coded algorithms for depth, as well as in the use of depth for video coding. Correlation between depth maps from different views has been exploited and decoded information from texture components was found useful for depth decoding, e.g. the motion prediction modes. In the case where video is decoded independent from depth, the decoder maintains compatibility with stereo decoders that do not support decoding of depth component. Otherwise, if view synthesis prediction is utilized, decoding of depth is required prior to decode the video [31]. Such tools have the

potential to provide interesting compression gains at the cost of reducing compatibility.

An important issue in the design of joint video and depth coding is the quality optimization of synthesized views. Instead of evaluating the decoding quality in comparison with an uncoded reference, the MVD format, besides good video and depth quality, also requires good quality for the intermediate synthesized views. As often the original intermediate view is not present, comprehensive subjective evaluation is required. Such evaluation takes into account new types of errors, like pixel shifts, regions appearing with wrong depths or outworn object boundaries at depth edges.

In experimental results [1], PSNR measurement is used to compare the synthesized views with uncoded reference views. Tests increased the video bitrate at the expense of the depth bitrate, thus increasing the video quality and reducing the depth map quality. At the original camera positions, the configuration that uses a lower quantisation parameter (QP) for video achieves better reconstruction results than using higher QP (30). However, for the intermediate positions using a higher QP to encode the video and allowing the depth maps to be encoded with a lower QP, increases the depth signal quality and reduces the displacement errors from view synthesis, resulting in a better overall quality for the synthesized view. Note that by using the depth information, the view synthesis scheme warps the original views to an intermediate position and applies a view-dependent weighting to perform the view interpolation. The furthest distance from any original view presents the lower quality values. Hence, these results show how important is to preserve the depth maps quality for the synthesis process, mainly at middle positions, far away from the original views. Besides these results, all dependencies between video and depth information are currently under evaluation in terms of their compression and rendering capabilities, as well as the repercussions in the compatibility and complexity of future coding algorithms.

2.4 Asymmetric 3D Video Coding

Asymmetric 3D video coding relies on the binocular suppression theory of the human visual system (HVS), which states that a given stereoscopic content with different quality between the two views can be perceived with the same quality as that of the higher quality view [8]. One of the reasons for this consists in the HVS response, which does not notice as relevant the absence of high frequency information in one of the views. Such characteristic of the HVS in regard to stereoscopic viewing can be used to achieve extra coding gains in comparison with classic single-view encoders where this is not exploited.

To benefit from the suppression theory, 3D video encoders may coarsely encode one component of the stereo video signal, in order to obtain increased coding efficiency gains

without harming the quality experienced by users. Several methods can be used to achieve unbalanced quality between the two views. For instance, either the spatial or temporal resolution of one view might be reduced through pre-processing. The texture quality of one view can also be reduced by coarse quantisation, resulting in asymmetric video encoding. Another method that can be used to reduce the amount of compressed data with little or no impact in subjective quality consists in dropping the chroma information of one view. Since chrominance degradation is less likely to be perceived, the chroma components of one stereo view can be dropped with no influence in the subjective depth perception of viewers [32]. Moreover if the lack of chroma information in one view is compensated through disparity or another reconstruction mechanism, then the color information in both views remains close to the original. This is because the image fusion process in the HVS superposes both images, resulting in a single chromatic content, rather than two slightly different ones from each view.

2.4.1 Mixed resolution for asymmetric coding

Asymmetric spatial resolution can be used in stereoscopic video coding by mixing different view resolutions [9]. Using mixed spatial and temporal resolution to a certain extent, the high frequency information removed from one view by low-pass spatial filtering is not detected by the HVS. In [9], spatial downsampling was implemented with filtering at 1/2 and 1/4 of the image resolution and temporal filtering was also done in two different ways. i.e., either by averaging the pixels from adjacent fields or dropping and repeating each other frame. The conclusions were drawn upon subjective testing carried out according to the ITU-R Recommendation 500 [33]. Other studies have also shown that mixed-resolution stereo video can achieve an overall perceptual quality and sharpness close to that of the higher quality view [10]. Therefore this is a possible method to reduce the coding rate of 3D video, providing that view asymmetry lies within an acceptable range. Figure 2.23 shows an example of asymmetric views obtained from low-pass filtering of one of them [34].

Mixed resolution coding is also a valid option to reduce the bandwidth required by mobile 3D multimedia applications. In [34], the results of subjective tests found that for an overall bit rate of 400kbps, the lower quality view can be encoded at 30% to 45% of the total bit rate allocated to the stereo pair.

Although spatial resolution asymmetry is not supported in standard multiview encoders, this is still possible to implement using the scaling properties of scalable video coding (SVC). A possible coding strategy is to use non-scalable H.264/AVC for the base view while the auxiliary view is encoded with a modified SVC encoder [35]. Objective quality results show improved coding efficiency compared to simulcast (-52.05 BD bitrate

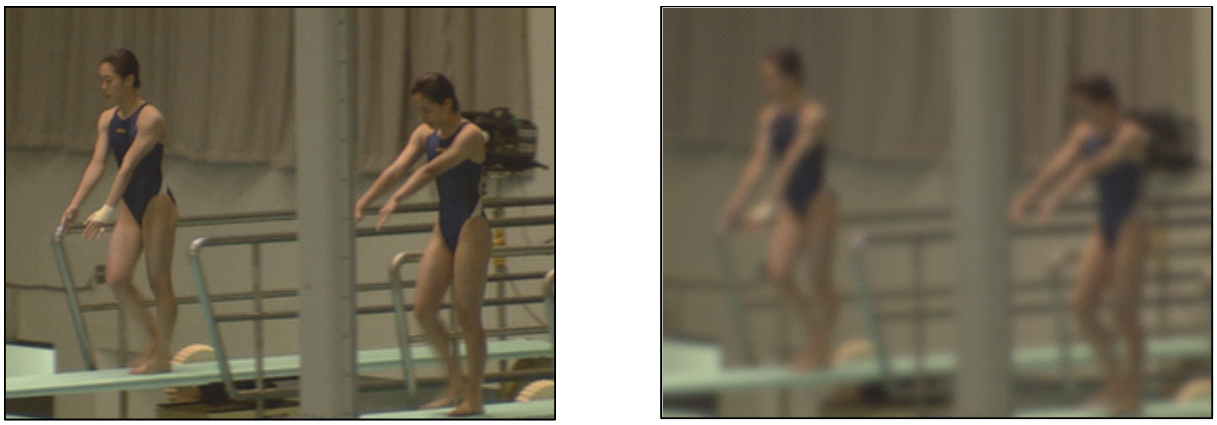


Figure 2.23: Asymmetric views obtained with low-pass filtering of one view [34].

or 4.66 BD-PSNR) or using inter-view predictions (-12.68 BD bitrate or 0.85 BD-PSNR). However since no subjective data is provided in [35], the actual perceptual performance of this method is not evaluated. To mitigate the additional computational complexity required for downsampling, spatial asymmetric multiview coding using lower complexity motion compensation can also be used [36].

To achieve spatial resolution asymmetry, Fehn et al. [37] proposed using additional downsampling filters in the encoding process. It maintains backward compatibility as the primary view is encoded according to the MVC standard and then downscaled to provide interview references to the secondary view. This way, spatially reduced frames of the second view can be predicted from the primary view at full resolution. However this results in a non-normative MVC stream.

Although mixed resolution asymmetric methodologies proved to be an efficient, usable and quality friendly method, similar studies on temporal-only asymmetric methods demonstrated low quality levels. Also the quality tends to be lower the higher the amount of (fast) motion in the encoded content. Some possible options to achieve temporal scaling of a stereoscopic video, such as dropping and repeating frames in one of the views or averaging frames to produce intermediate ones yield unacceptable results [10]. Also, in order to achieve temporal asymmetry, Aksay et al. [38] proposed a frame skipping method to reduce secondary view bit rate, using post processing mechanisms to replace odd secondary frames network abstraction layer (NAL) units with others signaling skip mode, reducing the overall bitrate of the stream. To decode these frames macroblocks from the previous frames were copied, repeating the reference frame. Else these frames could be discarded and an approximate version could be obtained by error concealment at the decoder.

2.4.2 Asymmetric quality

Using asymmetric quality in stereoscopic video provides an ease method to fine tune the quality of rate-constrained video [10] [39]. Figure 2.24 shows an example of a stereo pair obtained from asymmetric coding where blocking artifacts can be seen in the right view but not so much in the left one. This type of signal-to-noise ratio (SNR) asymmetric coding is one of the most suitable to be used in current standard MVC encoders because it does not require coping with different resolutions and downsampling/upsampling operations.

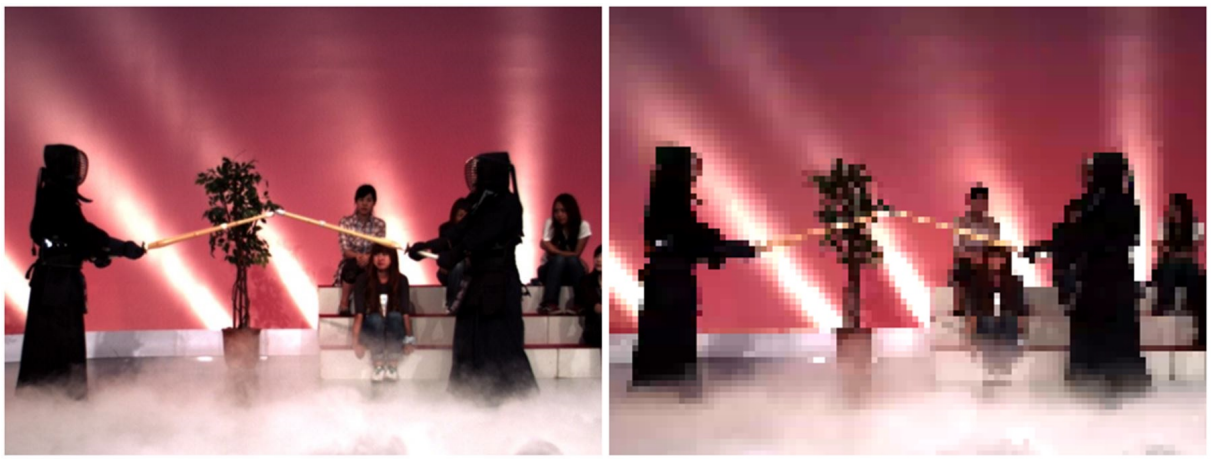


Figure 2.24: Result of asymmetric quality coding.

Those views not used as reference for others can be encoded with higher QP or allocated a lower bitrate than other views. This results in asymmetric quality amongst primary (reference) and secondary (non-reference) views. It is also possible to interleave lower quality frames with high quality ones in different views, e.g., odd frames in one view and even frames in another view with lower quality than the others [39]. This method has the advantage of eliminating viewer's asymmetric visual acuity (i.e., the influence of a dominant eye) and it is preferable to drop and repeat frames in the sequence.

When talking about 3D video one must not forget that free-view video is also regarded as a future big step in 3D technology and, as so, multi-view video (MVV) must be considered. Using a large large number of views has huge bandwidth requirements and it is also required to have an increasing number of views available for a more realistic experience. To solve the bandwidth problem asymmetry may be a solution if the number of views is relatively small, as shown in [40]. Asymmetric streaming of MVV is reasonable for no more than five views, alternating high and low quality views in order to exploit the HVS. Views can also have different rates, according to their importance in the rendering process. A complete framework, also including depth information, consisting of an MVD encoder and a bit allocation mechanism with chrominance reconstruction can be found

in [40]. This approach consists in asymmetric coding of the MVD-based video using the JMVM codec with reconstruction at the decoder of those chrominance components that are discarded at the encoder in the lower quality views. Considering the total bitrate, this approach significantly improves the coding efficiency while maintaining the overall quality experienced by end users.

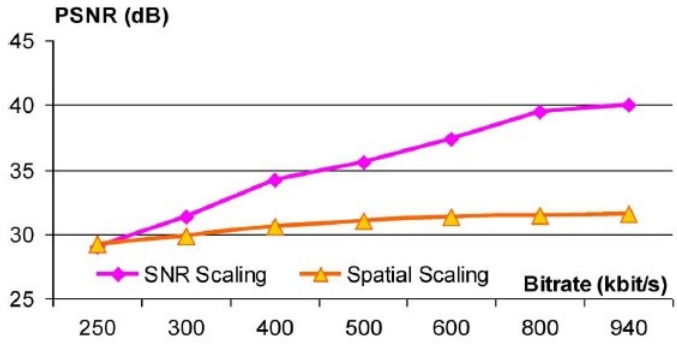
Another possible approach to encode asymmetric 3D video is to encode one view with a standard codec H.264/AVC and the other one with its scalable extension SVC [41]. Such scheme allows exploitation of a wide range of asymmetry with only one encoding channel. While the quality of one view is fixed, the one coded with SVC has high quality scalability range and it can be easily extracted with either lower or higher quality than the H.264/AVC view.

2.4.3 Perceptual quality thresholds

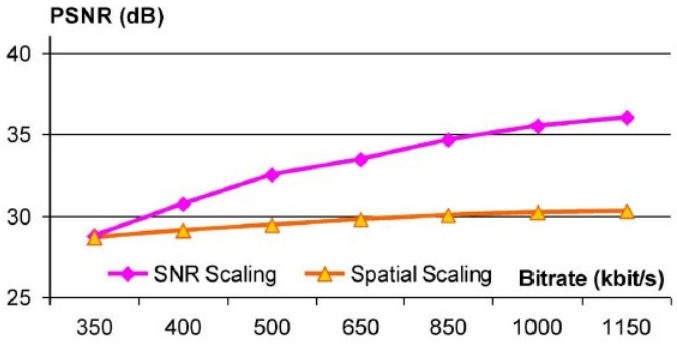
A relevant issue in asymmetric coding is to find perceptual thresholds beyond which quality degradation is noticeable by viewers. Relevant thresholds were experimentally found in [42] through subjective testing, where users started evaluating high quality symmetric coding (i.e., PSNR=40dB) in both views and reducing the quality of the auxiliary view down to 25 dB. It was found that such threshold slightly varies according to the display and lies around 31 dB for a parallax barrier display and about 33 dB for full-resolution polarized projection displays. Thus an average value of 32dB can be defined for the lower quality view when the other is very high quality.

When comparing SNR scaling with spatial scaling, the former should be used at high bitrates (above the previously stated threshold) as it results in better perceived quality. When low bitrates are used (e.g., below 28dB PSNR on the auxiliary view) spatial scaling tends to perform better than SNR scaling. When operating in some quality range between these two thresholds symmetric coding is preferable over other options. Note that display technology also influences the subjective quality obtained by symmetric/asymmetric coding [43]. The results displayed in Figure 2.25 show that in terms of R-D SNR scaling outperforms spatial scaling when above the defined thresholds. In terms of perceived quality subjective tests have provided the results to sustain this theory. Figures 2.26, 2.27 and 2.28 present the subjective scores (the method will be explained in section 2.5) obtained from both symmetric and asymmetric coding (SNR or spatial) above the 32 dB threshold and below (at 30 and 27 dB). Also in the three figures the results are discriminated according to the type of display used.

An indication of objective thresholds that can be used in asymmetric encoding of stereoscopic video is shown in Figure 2.29, obtained from the results presented in [41]. In the first case two SNR scalable streams are produced and then the enhancement layer on



(a)



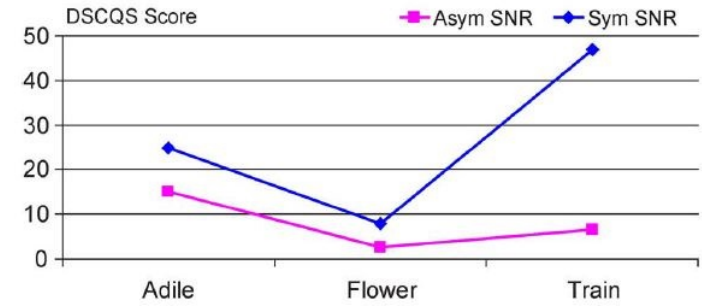
(b)

Figure 2.25: RD performance of SNR and spatial scaling (1/4) for (a) Adile and (b) Flower. [41]

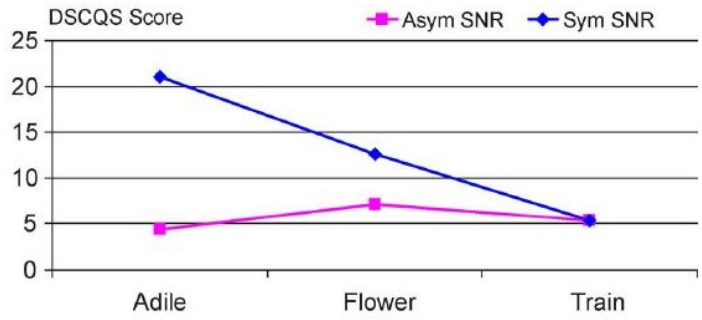
one of them might be cut down to the threshold of 32 dB. It is possible to achieve the lower bitrates with this scheme but maximum PSNR is compromised due to scalability overhead. Enhancements layers of both views can be used differently; hence one view can be extracted with maximum quality (using all enhancement layers) and the other with minimum quality (using only the base layer). In the other option, only one view is encoded with SNR scalability providing higher quality variation range. The other view is coded with H.264/AVC. The view coded with AVC provides a gain in compression efficiency that may be used to extend adaptation capability in the other view. It is possible to exploit asymmetry both if maximum or minimum quality is received for the SVC-coded view, with the difference that in the first case the SVC-coded view works as the higher quality view and in the second case the AVC-coded view is the higher quality one.

2.4.4 The effect of dominant eye

Although asymmetric coding algorithms can be used to reach greater coding efficiency at either small or no cost to the viewer's perceived quality, when a lower quality view corresponds to the viewer's dominant eye, some users can perceive the global quality



(a)



(b)

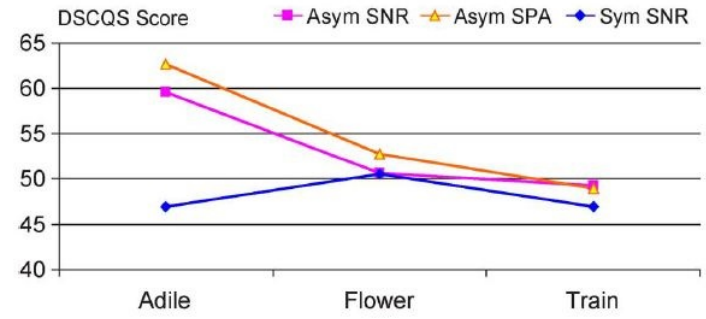
Figure 2.26: Subjective test scores for symmetric and asymmetric coding above the threshold PSNR for a (a) parallax barrier display and (b) a polarized projector. [41]

lower than others [44]. Therefore, such ocular dominance effect may lead to an overall perceived quality not equal to that of the higher quality view as commonly expected.

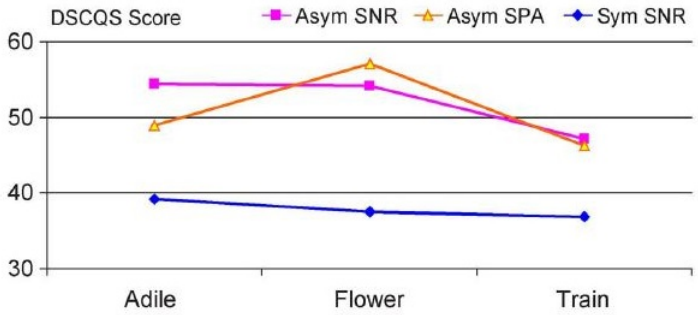
To reduce the impact of the ocular dominance effect it is possible to cross-switch the low and high quality views along the time [45]. However the interleaving of the views' quality along the time can also result in a noticeable effect, similar to flickering. To mitigate this effect an adaptive algorithm should be implemented in asymmetric encoders such that views' quality cross-switching occur at scene cuts, where a perceptual masking effect of the HVS occurs. The group of pictures (GOP) can also be used as a reference for cross-switching, as the frames inside a GOP have high correlation [46]. Even so the GOP size should not be small to avoid flickering effects. A possible alternative is to use unbalanced coding in horizontal slices with smoothing on the slice edges, in both views [47]. In this case, the high and low quality slices in each view should be located in complementary spatial positions in order to mitigate the effect of eye dominance.

2.4.5 Regions of Interest in 3D Video

Although it is a extensively exploited subject in traditional 2D video there is significantly less work done on region of interest (ROI) coding in 3D content. Using ROI techniques it is possible to have a finer approach and achieve spatial asymmetric coding based on



(a)

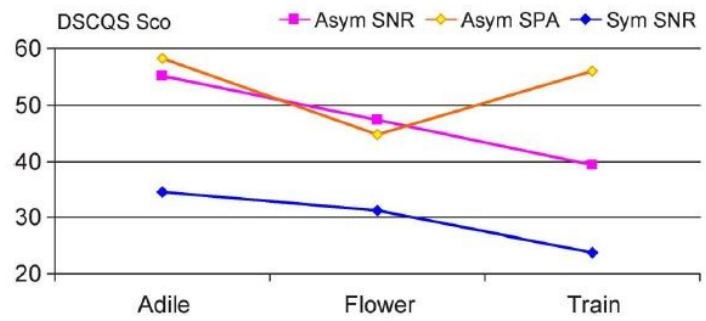


(b)

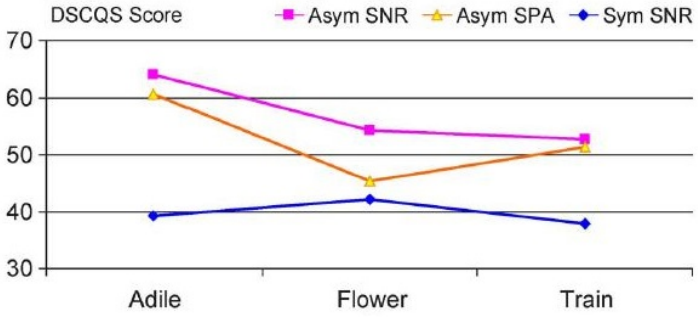
Figure 2.27: Subjective test scores for symmetric and asymmetric coding below threshold PSNR (at 30 db) for a (a) parallax barrier display and (b) a polarized projector. [41]

identification and encoding different regions of the stereo images according to their perceptual relevance. Relevant regions might be identified through a combined approach of depth variation thresholds and specific texture detection in order to differentiate the background and foreground of a 3D scene. Bit rate savings of 28% can be achieved by using a method where the scene background is given less relevance [48].

In [48] 3DTV while using V+D format is considered , such as in Figure 2.30. In order to increase perceived quality at limited bitrates the quality in regions interesting to the viewer is also increased. On the other hand regions outside of the interest area receive a reduction in quality. Two methods are used to retrieve the ROI. First the depth information is used and an adaptive threshold based on the histogram of depth values is calculated in order to identify the first third of depth values (pixels nearest to the screen plane). Then a binary map is defined for each texture frame that identifies if the pixel is in the ROI. Also, as the test sequences include talking heads, a skin detection algorithm has been incorporated in order to ensure faces are always in the ROI.



(a)



(b)

Figure 2.28: Subjective test scores for symmetric and asymmetric coding below threshold PSNR (at 27 db) for a (a) parallax barrier display and (b) a polarized projector. [41]

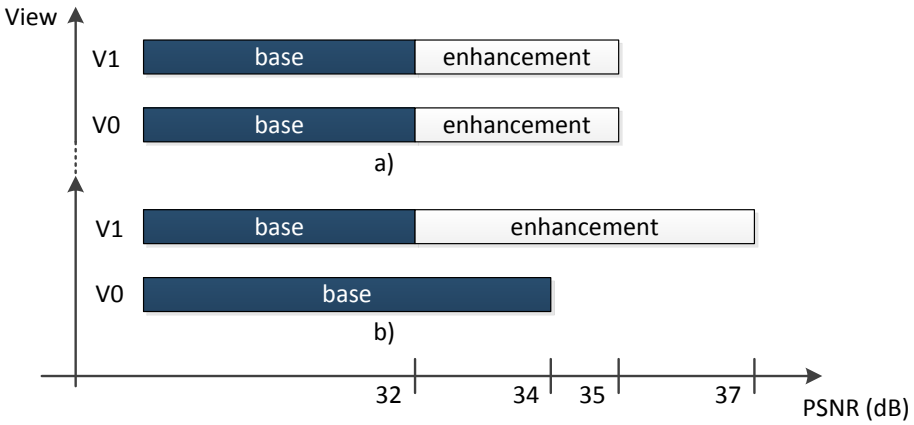


Figure 2.29: Thresholds for asymmetric coding (a) both views coded with SVC and (b) only one view coded with SVC [41].

Figure 2.30: Example of V+D video frame subject to ROI encoding [48].

2.5 Standards for Subjective Evaluation of 3D Video

ITU is the international organization responsible for video subjective evaluation recommendations ITU-R BT.500 [33] and ITU-T P.910 [49]. These subjective evaluation methods can be divided in two main categories: single stimulus and double stimulus. In single stimulus methods an observer evaluates the quality without a reference sequence for comparison with the same content. On the other hand, when dealing with double stimulus methods, a reference sequence precedes the sequence under evaluation and the observer scores the relative quality taking into account the given reference.

Since no specific video subjective evaluation methods exists for 3D, the recommendation ITU-R BT.1438 [50] is followed. According to this recommendation double stimulus methods presented in ITU BT.500 should be used as long as a reference sequence is available. When a reference sequence is not available Absolute Category Rating (ACR) methods should be used (ITU-T P.910).

Traditional monoscopic TV quality factors are based on resolution, colour rendition, motion portrayal, overall quality, sharpness, etc. Although these same factors can be applied to 3D subjective quality assessment there are many other factors that are specific to 3DTV. As so the following factors must be added to the list of relevant assessment factors in 3D content:

- Depth resolution;
- Depth motion;
- Puppet theater effect - objects unnaturally large or small;
- Cardboard effect - objects unnaturally thin.

In the following sections the main methods presented in these recommendations will be presented in more detail. Also, the evaluation environment specifications such as lightning will be described.

2.5.1 Recommendation ITU-R BT.500

According to recommendation ITU-R BT.500 the video subjective assessment should run under the following general viewing conditions when in a laboratory environment:

- Ratio of luminance of inactive screen to peak luminance: $\leqslant 0.02$;
- Ratio of the luminance of the screen, when displaying only black level in a completely dark room, to that corresponding to peak white: ≈ 0.01;
- Display brightness and contrast according to ITU-R BT.814 and ITU-R BT.815;

- Maximum observation angle relative to the normal (this number applies to CRT displays, whereas the appropriate numbers for other displays are under study): 30^o;
- Ratio of luminance of background behind picture monitor to peak luminance of picture: ≈ 0.15;
- Chromaticity of background: D_{65};
- Other room illumination: low;

Table 2.2: Preferred viewing distance table [33].

Screen diagonal 4/3 ratio	Screen diagonal 16/9 ratio	Screen height (m)	PVD (H)
12	15	0.18	9
15	18	0.23	8
20	24	0.30	7
29	36	0.45	6
60	73	0.91	5
100+	120+	1.53+	3-4

When conducting subjective assessment in a home environment there are some other conditions that should also apply. Mainly, for both 4/3 and 16/9 format display ratio, there is a preferred viewing distance (PVD) at which the observer should be placed. This distance is unchanged for standard definition (SD) or HD TV and can be seen in table 2.2. There is also a peak luminance of 200 cd/m^2 and an environmental illuminance in the screen (measured perpendicularly to the screen) of 200 lux.

Furthermore there are basic concepts that are explained in this recommendation as well as some additional features that the people conducting the assessment should be aware. Concepts such as psychoperceptual quality evaluation, user-centered quality evaluation, user-centered design or multimedia quality are defined. It is also recommended to have at least 15 participant observers that should be experts in subjective assessment. These observers must be subject to vision screening and receive instructions about the assessment method, possible artifacts, rating scores and timing. Also, to allow the observers to relax and adapt to the assessment method, test sequences with different content must be presented prior to the real assessment session. It is recommended to not exceed 30 minutes with the complete procedure.

Double Stimulus Impairment Scale (DSIS)

In Double Stimulus Impairment Scale (DSIS), the video sequences are presented in a double stimulus format. This is a typical method to evaluate a new system or to rate the effect of a certain transmission path. Once the assessment material is significant enough to conduct the assessment session the assessment parameters must be defined in order to

comprise the whole range of a certain impairment. The test conditions should also be defined as presented before.

As a double stimulus method, the observers are cyclicly presented with an unimpaired reference followed with a video sequence that may be impaired due to the phenomena under study. For each pair of sequences a rating is required from the observer. This rating should represent the subjective quality of the second sequence to the observer, taking into account the quality of the first sequence. The presentation sequence should be randomized to reduce bias and the unimpaired sequence must be included amongst those which are impaired.

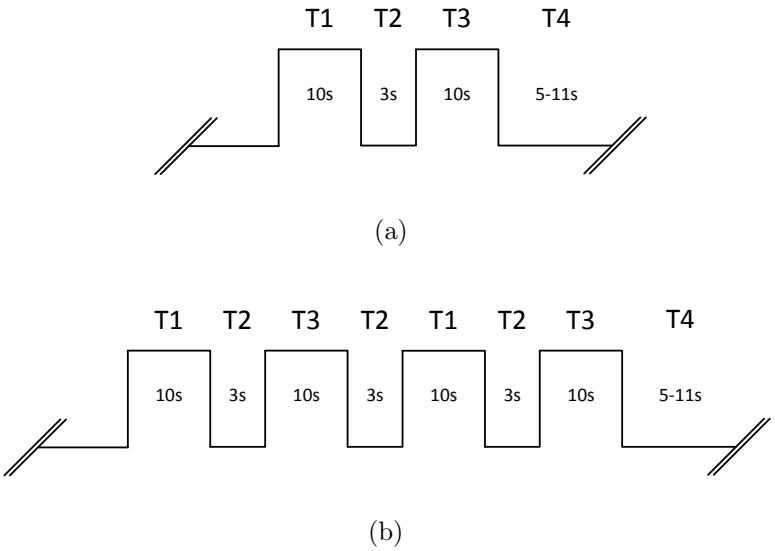


Figure 2.31: DSIS presentation variants I (a) and II (b).

In Figure 2.31 the two temporal structures of the method can be observed. T1 is the viewing time for the reference sequence, lasting about 10 seconds, followed by a 3 second mid-grey interval (T2). Then the impaired sequence is presented (T3) and the rating period (T4) starts lasting no less than 5 seconds or more than 11. According to the Recommendation observers should be asked to look at the picture for the whole of the duration of T1 and T3 and voting should be allowed only during T4. The rating is done according to the 5 point discrete scale:

- Imperceptible;
- Perceptible, but not annoying;
- Slightly annoying;
- Annoying;
- Very annoying.

A second variant of the method exists, also depicted in Figure 2.31 which differs in

the number of times that both the reference and the impaired sequence are displayed to the user. Although more time-consuming, due to the repetition of both the reference and the impaired sequences, this variant helps to better discriminate small impairments and is more adequate when dealing with moving sequences.

Once again it is recommended to randomize the sequence list before each assessment session and to avoid repetitions of the same video sequence (even with a different type or amount of impairments). There should be enough material to cover all the possible ratings on the scale, aiming for a final mean around 3. The whole procedure should last at most 30 minutes, including the preliminary sequences and explanations.

Double Stimulus Continuous Quality Scale (DSCQS)

Using a similar structure as DSIS, the Double Stimulus Continuous Quality Scale (DSCQS) differs in the rating scale. In this method, the observers' scores are continuous over a scale ranging from Bad to Excellent. Furthermore the reference sequence is hidden. This means that the observer does not know which one of the sequences (in slot T1 or T3 if we take into account Figure 2.31) is the impaired one.

DSCQS is recommended when 5 discrete levels are not enough to rate the content quality. This method can also be used if it is required for the observers to have a greater range of freedom to rate the different sequences.

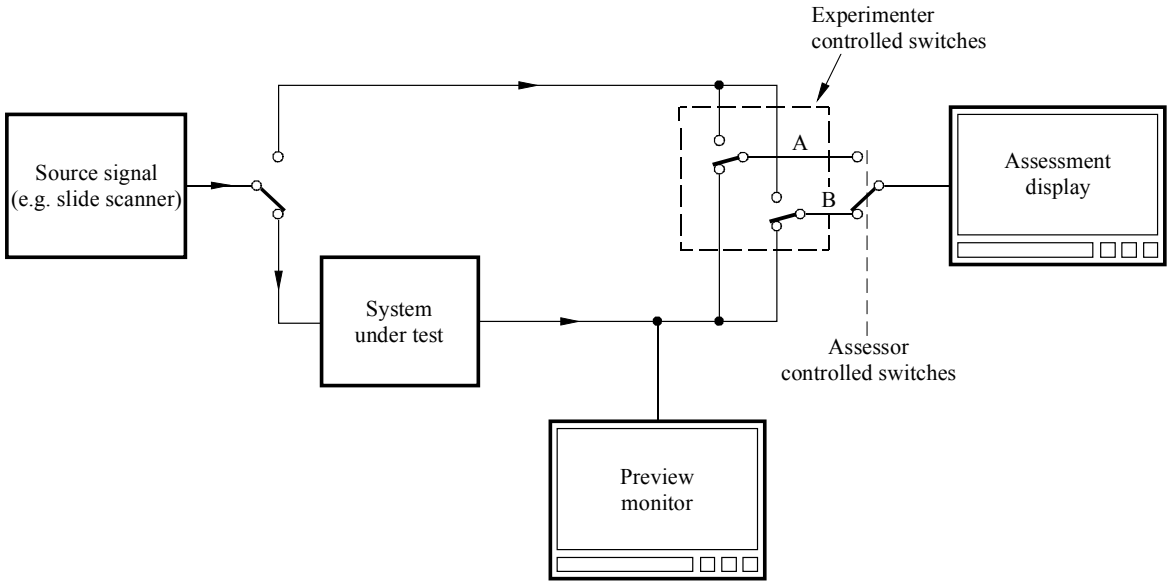


Figure 2.32: General arrangement for a DSCQS test system [33].

There are also 2 variants of this method, illustrated in Figure 2.32 using switches A and B. In the first variant the observer may choose which of the sequences he wishes to

see and he/she may watch them as many times as he/she needs to rate them. When using the second variant, the assessment is controlled by the person conducting the assessment making it more straightforward to the observer that just has to watch and rate as in the DSIS case.

To analyse the results coming from the continuous scale it is recommended to normalize it to a scale from 0 to 100. The score represents the difference between each pair of sequences.

Additional Assessment Tools

In this recommendation several other methods and tools are also presented. One of the most meaningfull is the Single Stimulus Continuous Quality Evaluation (SSCQE). In this method the sequences' quality is evaluated without a reference using a continuous scale similar to DSCQS. It is best used in long duration stimuli.

Acceptance threshold methods are also used for long sequences (around 60 seconds). Observers are given a MIDI controller and they aim to identify the unacceptable quality parts of the sequence. In the end, an 11-point scale is used for an overall quality rating and a binary scale of *yes* and *no* is used to rate the overall acceptance of the sequence.

Additional tools such as free-choice profiling can also be used along any other to allow the observers to describe the sequence characteristics using their own words. It can be specially usefull when future users can give suggestions about the system under test in order to make it more user-friendly, for instance.

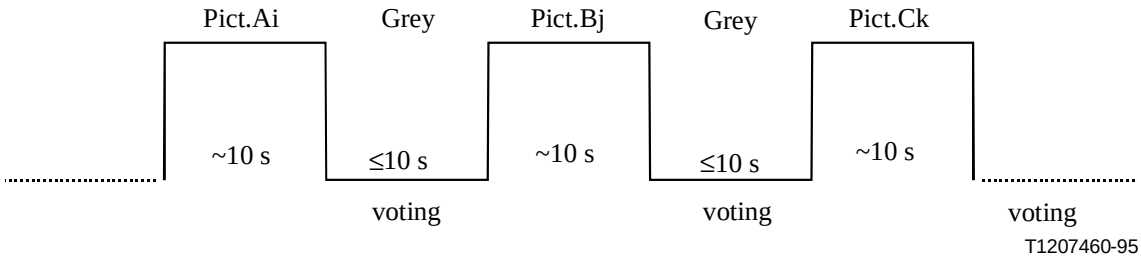
2.5.2 Absolute Category Rating (ACR)

ITU Recommendation ITU-T P.910 [49], Subjective Video Quality Assessment Methods for Multimedia Applications, describes subjective assessment methods to evaluate the overall video quality in multimedia applications, such as videoconference. Environment conditions and number of observers differ from recommendation BT.500, although the same conditions as the former should be used, as recommended in BT.1438.

Once again, prior to the assessment, session participants must receive written information about the test scenarios, the type of assessment and the stimuli presentation. Furthermore preliminary, non-recorded, assessments should be representative of the type of impairments that may appear during the assessment session but should also use different content.

ACR is one of the two main methods presented in this recommendation. It is a single stimulus method, where sequences are presented one at a time and then independently rated from one another. It can also appear simply under the name of Single Stimulus

method.



Ai Sequence A under test condition i
Bj Sequence B under test condition j
Ck Sequence C under test condition k

Figure 2.33: Temporal structure of the ACR method [49].

Test sequences last around 10 seconds and are observed one after the other, without a reference sequence. Rating is done immediately after the sequence is finished within a period of 10 seconds. This procedure can be described as shown in Figure 2.33. The evaluation is done according to a 5-point discrete scale or, in alternative, a 9 or 11 point scale if a more detailed rating is required.

This method allows to categorically (5 categories or 9 if higher discriminative power is needed) judge the quality of a video, from Bad to Excellent. The implementation of the method is straightforward and, since the presentation order is similar to most systems, it is pretty accurate to assess the final quality of a multimedia service.

2.5.3 Subjective Assessment of Multimedia Video Quality (SAMVIQ)

Subjective Assessment of Multimedia Video Quality (SAMVIQ) [51] is yet another method described in recommendation ITU-R BT.1788 [52] for subjective assessment of video quality. However this method differs from single and double stimulus methods, as it allows the viewer to access several versions of the sequence at a time.

In this method, the viewer is presented a graphical interface such as the one shown in Figure 2.34. The different versions of a sequence can be accessed through it using stop, review and selector controls. It is also possible to modify the given score to each version of the sequence at any given time. Once all the versions of the given sequence are rated the viewer is then allowed to move on to the next sequence content.

Similarly to double stimulus methods, an explicit reference is included (hidden references can also be present if desired) that can also be seen whenever desired. The rating is done in a continuous quality scale similar to that of the DSCQS method (100 points)

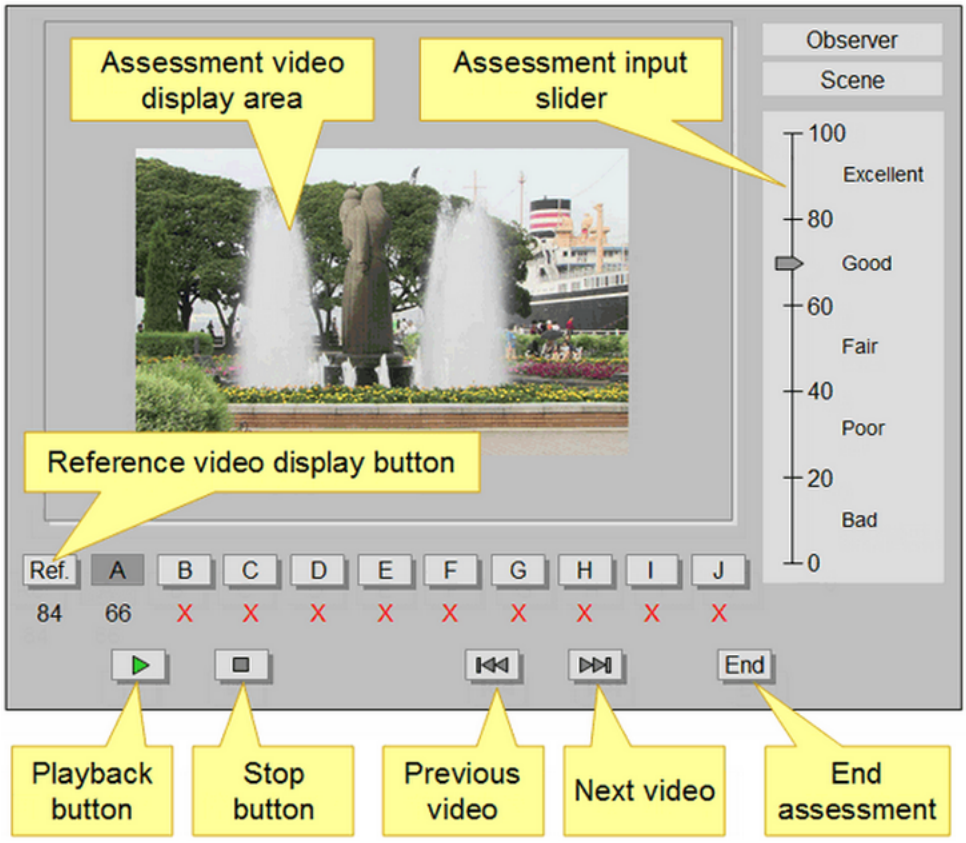


Figure 2.34: Example of a graphical interface for the SAMVIQ method [53].

that has 5 quality levels linearly arranged, once again from bad to excellent. Short stimuli with 15 seconds at most are adequate for this method.

2.5.4 Comparison of methods

Table 2.3: Comparison of the presented methods.

	DSIS	DSCQS	SSCQE	ACR	SAMVIQ
Reference:	Explicit	Hidden	—	—	Both
Comparison:	Double stimulus	Double stimulus	Single stimulus	Single stimulus	Multi stimulus
Rating:	Discrete (5 points)	Continuous	Continuous	Discrete (5 points or more)	Continuous
Stimulus:	Short (10s)	Short (10s)	Long (1min or more)	Short (10s)	Short (15s)

Table 2.3 presentes a short summary of the main characteristics of the previously presented subjective assessment methods. Double stimulus methods offer more transparency and fidelity than single stimulus methods, due to the presence of explicit references. Degradation Category Rating (DCR) (similar to DSIS but presented in P.910) and, consequently, DSIS have long been used to assess television and other high quality

systems [49]. The rating scale of these methods supports the discrimination of the impact of the perceptible (or not) impairments.

On the other hand, ACR is faster to implement and the presentation of stimuli is similar to common systems. Also, assessment time required by this method is half of that of methods such as DSIS or DCR, while maintaining good stability. Tests conducted on DSCQS, ACR, DCR and ACR with hidden reference proved that Pearson and Spearman rank correlation coefficients of the MOS across all methods is rather high, although with DCR's MOS having lower 95% confidence intervals [54].

2.5.5 Handling results

Once the results of the subjective assessment are retrieved, the mean scores, $\bar{u}_{jkr}$ (Equation 2.1), of all N observers (i) for each sequence (k), test condition (j) and repetition (r) are calculated, as well as the standard deviation, S_{jk} (Equation 2.2). Also 95% confidence intervals should be displayed when presenting the results.

$$\bar{u}_{jkr} = \frac{1}{N}\sum_{i=1}^{N} u_{ijkr} \tag{2.1}$$

$$S_{jk} = \sqrt{\sum_{i=1}^{N} \frac{(\bar{u}_{jk} - u_{ijkr})^2}{N-1}} \tag{2.2}$$

When less than 20 observers are used in the evaluation it is recommended to run an algorithm to screen them and check if any of the observers must be excluded. Kurtosis coefficients (β_{2jkr}) are calculated (Equations 2.3 and 2.4) to verify if the scores for test presentations follow a normal distribution. If β_{2jkr} is higher than 2 and lower than 4, the distribution may be taken as normal. Each observer score (u_{ijkr}) is then compared with the associated mean value for that presentation plus and minus the associated standard deviation times 2 (Equations 2.5 and 2.6) or times the square root of 20 if the distribution is not normal (Equations 2.7 and 2.8). Every observer's score above u_{jkr} increments a P_i counter. The same way a Q_i counter is incremented each time a score is below u_{jkr}. Finally the observer is eliminated if his scores comply with Equation 2.9 or Equation 2.10.

$$\beta_{2jkr} = \frac{m_4}{(m_2)^2} \tag{2.3}$$

$$m_x = \frac{\sum_{i=1}^{N}(u_{ijkr} - \bar{u}_{ijkr})^x}{N} \tag{2.4}$$

if $2 \leq \beta_{2jkr} \leq 4$, then:

$$u_{ijkr} \geq +2S_{jkr} \Rightarrow P_i = P_i + 1 \tag{2.5}$$

$$u_{ijkr} \leq -2S_{jkr} \Rightarrow Q_i = Q_i + 1 \tag{2.6}$$

else:

$$u_{ijkr} \geq +\sqrt{20}S_{jkr} \Rightarrow P_i = P_i + 1 \tag{2.7}$$

$$u_{ijkr} \leq -\sqrt{20}S_{jkr} \Rightarrow Q_i = Q_i + 1 \tag{2.8}$$

Reject observer i if:

$$\frac{P_i + Q_i}{J * K * R} > 0.05 \tag{2.9}$$

or

$$\left|\frac{P_i - Q_i}{P_i + Q_i}\right| < 0.3 \tag{2.10}$$

Chapter 3

Non-uniform Asymmetric Coding

This chapter presents the proposed asymmetric coding method based on regions of perceptual relevance proposed in this dissertation. In the following sections the method used to define the regions of a stereo pair with different perceptual relevance is described. Then the implementation of the necessary modifications to obtain asymmetric encoding, both uniform and non-uniform (using regions of perceptual relevance) in a reference H.264/MVC encoder are explained. Due to the high complexity of the codec source code, only the main points of action will be targeted. Finally, the software developed during this research work is presented as well as additional implemented functionalities that provide possible directions for future work.

3.1 Regions of Perceptual Relevance

In this dissertation we look to advance the state of the art in asymmetric coding by adding a new spatial quality dimension to the asymmetry of the auxiliary view. To distinguish between the two SNR asymmetric coding methods we define the traditional methods as uniform asymmetric coding and the presented method using regions of perceptual relevance as non-uniform asymmetric coding. We achieve non-uniform asymmetric coding by defining regions within the stereo pair which have different perceptual relevance to the observer. We classify them as being of either high or low perceptual relevance for the 3D user experience in order to later encode the most relevant regions with higher fidelity and the less relevant regions with lower quality. Once again it must be noted that the principle behind such approach relies on the fact that high disparity image regions are those that mostly contribute to the perceived depth in stereoscopic images. Thus, the non-matching regions in each stereoscopic pair of images (i.e., those with high differences) identify the spatial locations of higher disparity, which in general correspond to object borders and their neighbouring regions.

These regions are initially defined at pixel level, based on an estimate of disparity, and then converted to a MB level. In order to reach different quality levels, according to the perceptual relevance of each region, different quantisation factors are applied during the encoding of the auxiliary view. This is done by changing the QP of each auxiliary view MB during the encoding phase according to the perceptual relevance of the region the MB is in. Doing this we achieve a non-uniform asymmetric coding in the way that the auxiliary view of a stereo pair is not uniformly coded (with a single QP) as traditionality happens. Instead, there are two different possible sets of QP to encode each MB according to its relevance (quantized in either high or low). The main view of the stereo pair is encoded normally, without any significant extra processing, following the principles of asymmetric quality coding.

In a practical implementation the regions of perceptual relevance must be defined prior to the encoding of a stereo pair. Then, a file identifying the regions of the auxiliary view frame which have higher perceptual relevance needs to be fed to the modified encoder in a way that makes it possible to be associated with each MB in the MB coding loop. As so, the aforementioned regions with different perceptual relevance are defined, giving rise to a binary pixel mask, by calculating a disparity estimate between the pixels of the 2 images for each stereo pair. This means that for a video sequence there will be n binary masks corresponding to the n frames of each view. The regions of perceptual relevance are first determined from the absolute pixel differences $AD_{f_k}(x, y)$ between stereo image pairs ($fl_k(x, y)$ and $fr_k(x, y)$), using equation 3.1. Then, after removing noise using median filtering, the absolute differences greater than a fixed threshold (Th1), from which lower values are equal to 0 and equal or higher values are equal to 1 (normalised), define the pixel positions in the region of higher perceptual relevance

$$AD_{f_k}(x, y) = |fl_k(x, y) - fr_k(x, y)| \quad 0 \leq x \leq M-1; 0 \leq y \leq N-1 \tag{3.1}$$

Before feeding the regions of perceptual relevance information to the encoder in order to encode them with different quality the pixel-based mask is processed, giving rise to a MB-based mask, where each region is comprised of an integer number of MB. The actual MB-based region with high perceptual importance is defined by those MB that verify the condition imposed by equation 3.2, where X_{ij} are the binary values of the pixel mask in each spatial position ij. At this point, from a black and white YUV file, we get a smaller (16 times, corresponding to the width and height size of the MB), more blocky and also black and white mask where each pixel represents a MB. This later mask is visually binary as it contains only zeros (black) for pixels outside the region with higher perceptual relevance and 255 (white) for pixels within. In order to further improve the regions of perceptual relevance at a MB level, the MB-based mask is subject to a morphological

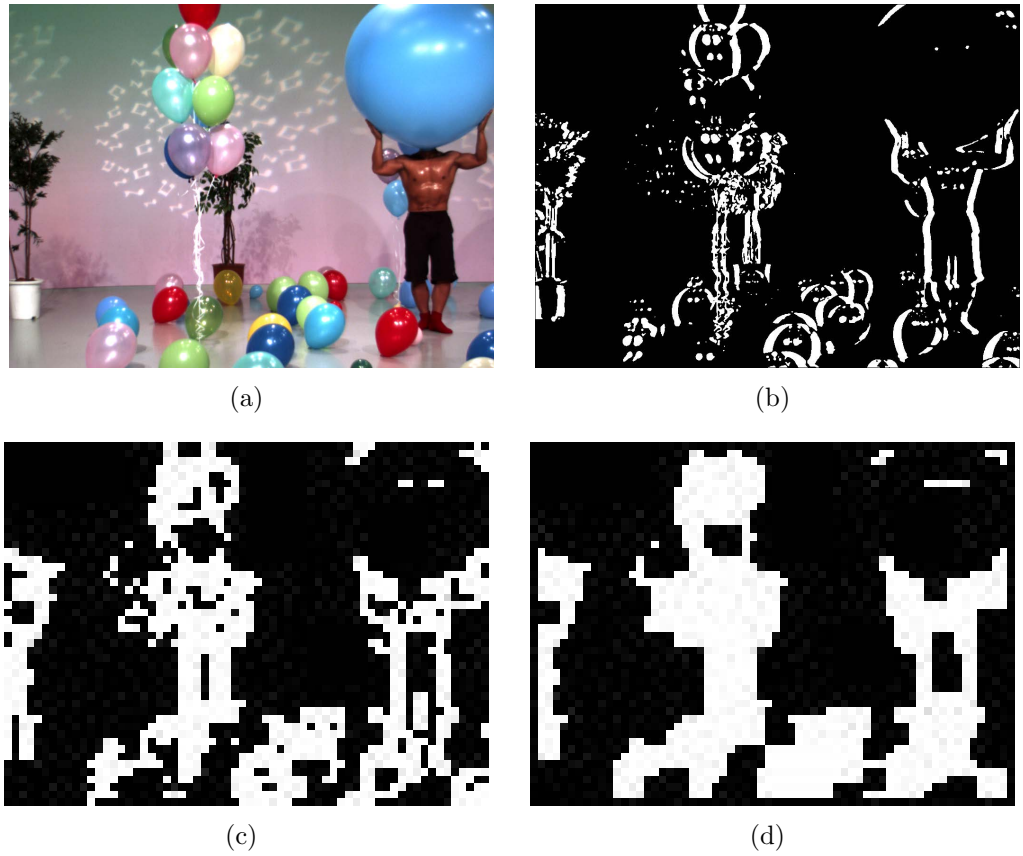


Figure 3.1: Sequence Balloons: auxiliary view (a), regions of perceptual relevance at a pixel level (b), MB level (c) and after applying morphological closing (d).

binary operation of closing, which contributes to smooth the mask and make it fit better with the objects. An example of the process can be seen in figure 3.1.

In this work both thresholds were empirically defined such that MB masks define the regions of higher relevance for 3D perception, i.e., those containing most edges and surrounding areas as pointed out above. Also the closing morphological operation was chosen due to being the one among the experimented ones, using the *bwmorph* function in Matlab, considered to better fit the objects borders and neighbouring regions in the scenes. The closing operation consists of applying a morphological dilation on the binary image followed by a morphological erosion operation.

$$\sum_{(i,j)\epsilon W} |X_{ij}| \geq Th_2 \tag{3.2}$$

After computing the regions of perceptual relevance for all pairs of frames in the stereoscopic sequence the file containing them is just a normal monoscopic video in YUV format, with the peculiarities that it has a very small spatial resolution and binary (black and white). This file contains a single byte for each MB of each auxiliary view frame that

identifies if it belongs to the region with higher perceptual relevance and can be easily read as a binary file at the start of the encoding process.

3.2 Asymmetric Encoding Implementation on the JM Encoder

To evaluate the coding performance obtained from implementation of the regions of perceptual relevance, asymmetric encoding was implemented in a reference H.264/MVC software. Due to familiarity with the previous version of the codec, the JM 18.0 was selected. The JM reference software is currently in its 18.5 version and can be freely downloaded from [55] along with it's manual and further documentation.

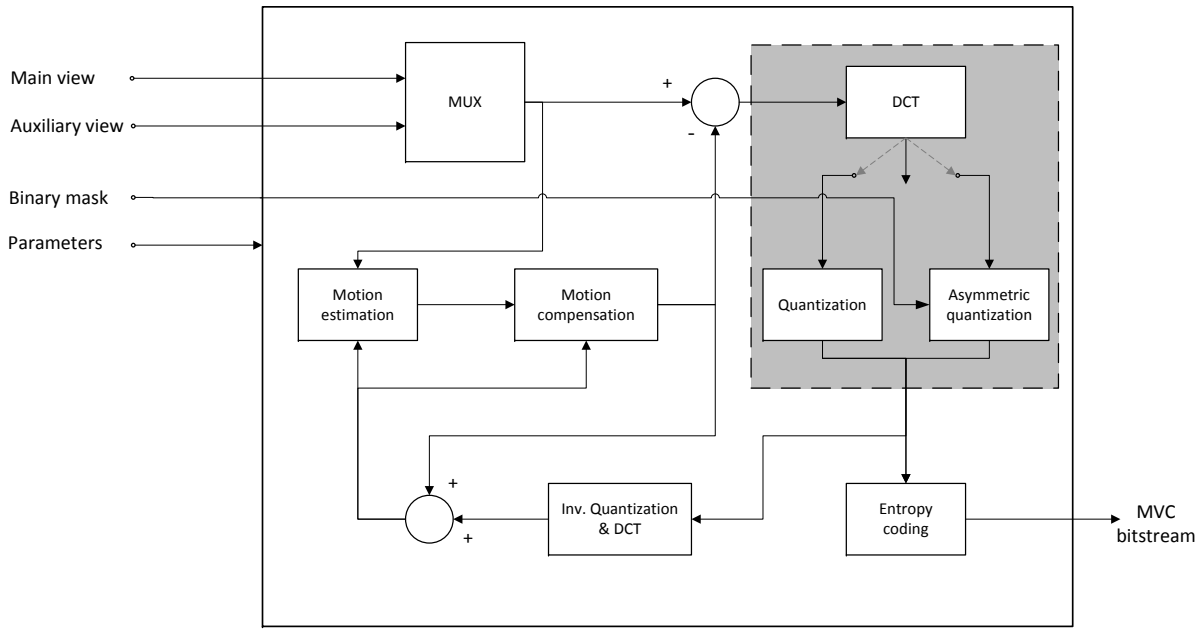


Figure 3.2: Schematic representation of the modified encoder.

Figure 3.2 depicts the main modifications that were implemented in the JM encoder. In this Figure the encoder side is represented by the large box enclosing the H.264/MVC operational blocks. The operational blocks of the encoder which were modified in order to implement asymmetric encoding are enclosed in the dark gray area in the right side. To achieve asymmetric coding through manipulation of the QP one must change the quantization routines that happen after the DCT transform block. Uniform SNR asymmetry can be implemented as well as non-uniform SNR asymmetry. In the former, the frames from the same view have a similar QP configuration (i.e. similar quality) across all MBs, while a different quality is assigned to the complimentary view. In the later, the quality

is not uniform inside each frame. To control these asymmetric coding methods, namely (uniform or non-uniform) SNR asymmetric, new inputs have to be added, along with the routines to read and handle them once inside the encoder. As so new configuration parameters are created in order to define the type of asymmetry and the QPs to use in each case. Also an auxiliary file (binary mask) identifying the MBs that are within the region perceived to have greater perceptual relevance is fed to the encoder.

The JM software consists of a Visual Studio solution of 4 projects that can be independently built:

- ldecod;
- lencod;
- rtploss;
- rtpdump.

The two last projects, *rtploss* and *rtpdump*, are useful to simulate Real-time Transport (RTP) protocol packet losses but were not used as no packet loss tests were made and RTP packetisation was not used. The main implementation work in this dissertation was developed around the *ldecod* and *lencod* projects which are, respectively, the H.264 reference decoder and encoder. Asymmetric encoding has been implemented purely in the encoder side and requires no modifications to the decoder in order to be decodable, which makes it standard compliant. On the decoder side of the transmission chain several modifications have been implemented, not regarding asymmetric encoding but instead for providing further information or enabling future functionalities.

The JM encoder is a rather large project, consisting of 87 header files and 104 source files in C language, as would be expected from such a complex process as compressing a stereoscopic 3D video to usable bitrates. However, modifications in order to implement asymmetric encoding require changes in only a few processes, thus only a few source files need to be modified. These changes will be presented in more detail in this section. The whole encoding process can be controlled through a configuration file with a *.cfg* extension, editable by any simple text editor. In this file all the encoding parameters can be changed, from the input files to QP or prediction modes. Even though further information about the configuration parameters can be found in the user's manual, the main parameters that need to be changed in order to encode a stereoscopic video (with a brief description) are listed below:

- *InputFile* and *YUVFormat* - the input sequence for the main view to be encoded and it's YUV format);
- *FramesToBeEncoded* and *FrameRate*;
- *SourceWidth*, *SourceHeight*, *OutputWidth* and *OutputHeight*;

- *OutputFile* - the resulting encoded sequence in a *.264* extension;
- *NumberOfViews* and View1ConfigFile - in case of a stereoscopic a second configuration file should be provided;
- *ProfileIDC* - *i.e.* 128 for Stereo High Profile;
- *IntraPeriod* and *IDRPeriod* along with *NumberBFrames* to define the IPB format;
- *QPISlice*, *QPPSlice* and *QPBSlice* - QP for I, P and B slices, respectively;
- *NumberReferenceFrames*;
- *RateControlEnable* and *Bitrate* if rate control is enabled.

As seen in the previous settings if one is to encode a stereoscopic video, then an auxiliary configuration should be provided to deal with the auxiliary view. This additional configuration file is simpler, having significantly less parameters, and it is mainly used to feed the encoder with the input file corresponding to the auxiliary view. The majority of the parameters already configured for the main view are also kept for the auxiliary view. However, some of them such as the ones corresponding to search and mode types used in each kind of slice are possible to change. Appendix B provides an example of a configuration file of a stereoscopic video sequence. Note that the given example has the parameters added during this dissertation located in the auxiliary view configuration file.

In order to perform asymmetric SNR encoding using the configuration of different sets of QP for each view one must go as deep in the encoding process as to the point where the DCT transform is applied and the quantization is done for each macroblock. As so one needs to look at the coding functions. Until reaching the specific macroblock encoding functions, where we are deep in the code to the point where motion estimation and the prediction and sub-prediction modes are chosen, the encoding procedure goes through these main functions:

- *encode_sequence*;
- *encode_one_frame*;
- *perform_encode_frame*;
- *frame_picture*;
- *code_a_picture*;
- *code_a_plane*;
- *encode_one_slice*;
- *encode_one_macroblock*.

Following this and looking at the main structures of the program the main QP controller variable, *qp*, can be found in the *VideoParameters* structure. This variable which can be accessed by using the pointer *p_Vid*. This same structure and pointer can be used in order to get the input parameters (read from the configuration files) as a pointer to the *InputParameters* structure is included in the *VideoParameters* structure.

3.2.1 Adding Input Parameters to the Encoder

This leads us to the addition of further input parameters in the configuration file of the encoder. This can be done by modifying the header file responsible for defining the format of the configuration files (*configfile.h*) and the input data structure defined in *params.h*. As so the additional parameters inserted in the configuration file are presented below for the auxiliary view along with a brief explanation. As there is no asymmetric coding with only one view, if the encoder is used to encode a monoscopic sequence these additional parameters are not needed and all the flags signaling asymmetric coding or the use of information from a previous decoder stage are defaulted to their appropriate values.

- *Transcoding* - Flag (0=off, 1=on) to enable the use of transcoding info from previous decoder outputted files;
- *UseDecoderMVs* - Flag (0=off, 1=on) enabling the use of motion vectors info from previous decoder outputted files;
- *TranscodingInfoFile* - name of the binary file with transcoding info, i.e. "*infofile.bin*";
- *TranscodingMVFile* - name of the binary file with motion vectors info, i.e. "*mvfile.bin*";
- *AssymCoding* - Flag for view 1 Asymmetric Coding (0=off, 1=Quality/QP Based, 2=Spatial Resolution, 3=Temporal);
- *nonbaseQPISlice* - QP for I Slices (0-51);
- *nonbaseQPPSlice* - QP for P Slices (0-51);
- *nonbaseQPBSlice* - QP for B slices (0-51);
- *nonbaseQPSPSlice* - QP for SP-Slices for Prediction Error (0-51);
- *nonbaseQPSISlice* - QP for SI-Slices for Prediction Error (0-51);
- *MBAssymCodingFlag* - Flag (0=off, 1=on) to enable the use of different QP for MBs in the region with greater perceptual relevance;
- *MBAssymCodingQP* - QP for MBs in the region with greater perceptual relevance for all slice types (0-51);
- *RelevantMBsFile* - name of the file with the binary mask identifying the MBs in the region with greater perceptual relevance, i.e. "*mask_Balloons.yuv*".

As seen above the parameters added to the auxiliary view configuration file allow the configuration of different QP for I, P, B, SP and SI slices of the auxiliary view (in the conducted tests only I, P an B slices were used) and a single QP for MBs that are within the region of higher perceptual relevance, whatever the type of slice. It must be noted that only QP parameters for each type of slice were added. There are no additional parameters to configure, i.e., possible small changes in QP according to the type of frame

(eg. reference frame). Nevertheless, it is possible to manipulate QP parameters in the main view configuration file in order to achieve that for the auxiliary view as well, even if asymmetric coding is enabled.

3.2.2 Uniform Asymmetric Coding - Frame Level

In the first stage of uniform asymmetric encoding implementation, a single QP for a whole frame of the auxiliary view, is used. This is done at the frame encoding level as the input parameters are available during the coding loops for each frame through the use of the previously mentioned pointers. Note that in our tests each frame corresponds to a unique slice. In order to correctly select the QP to use at start of each frame, modifications were made in function *encode_one_frame* (file *image.c*), which runs at the start of the encoding of each frame. Thus, when asymmetric coding is enabled, through the *AssymCodingFlag* flag in the configuration file, the QP for the current frame ($p\_Vid \rightarrow qp$) is set according to the type of slice and the QP selected for that same type in the configuration file. As the encoder was not prepared to handle different sets of QP for different view frames not only the QP for frames of the auxiliary view have to be changed but also the QP for frames of the main view. This must happen in order to reset the default QP for the type slice (defined in the main view configuration file) when encoding a main view frame after a auxiliary view one. The processing flowchart of the implementation is represented in Figure 3.3.

As seen in the Figure, once the encoding is initialized and the parameters are read (including the perceptual relevance information, which is not needed at this stage), at the start of each frame it's view ID is used as a condition to select the appropriate QP. If the frame belongs to the main view, the set of QPs (for each type of slice) that is configured in the main view configuration file is used. On the other hand, if the view ID is 1, then it belongs to the auxiliary view and QPs from the additional input parameters in the auxiliary view configuration file are used. This results in a uniform asymmetric coding as the QP is uniform throughout each view but different for each one. The previously mentioned *AssymCodingFlag* flag disables the first decision point and defaults all frames to use QPs as if they were from the auxiliary view.

3.2.3 Non-uniform Asymmetric Coding - Macroblock Level

At this point, on top of the already implemented uniform asymmetric coding, for frames of the auxiliary view, a new QP decision level needs to be implemented in order to achieve non-uniform asymmetric coding. To do that we go deeper into the coding loop, down to the MB encoding cycle, in order to identify MBs that are within the region with greater

perceptual relevance defined in the binary mask used as input to the encoder. Then (as depicted in Figure 3.3) if the MB relevance is set to high (1) the QP configured for MBs with higher perceptual relevance is used. On the other hand if the MB relevance is lower (0) the QP coming from the frame loop is kept.

As mentioned before, the binary mask contains a single byte of data for each MB, identifying if it is within the region with higher perceptual relevance. These values come from an YUV file and, as so, were kept as 0 (black) for MBs outside the said region and 255 (white) for MBs inside that same region. To check if a MB is within the region with greater perceptual relevance changes were made to the function *start_macroblock* in *macroblock.c*. This function runs at the start of the encoding of each macroblock initializing the necessary variables for it. As seen by the excerpt of code below, resorting to the information read from the binary mask, if the current MB with address *mbAddrX* results in a corresponding value in the binary mask of 255 it's QP will be changed to that defined in the auxiliary view configuration file for MBs inside the region with higher perceptual relevance. On the other hand, and especially in the case where a MB inside the region was coded previously to one outside, an else condition is added in order to reset the QP of the MB to the default value, similarly to what happens when defining the QP for the frame.

```
if (p_Vid->p_Inp->MBAssymCodingFlag){
  if (*(p_Vid->p_Transcod->relevantMBs[(int)(p_Vid->frame_no)] +
            ((*currMB)->mbAddrX)) == 255 && p_Vid->view_id == 1) {
          mb_qp = p_Vid->p_Inp->MBAssymCodingQP;
   }else{
            ...
       }
```

This process maintains compliance with the default decoder as the change in the MB's QP is included in the bitstream, through the variable *mb_qp_delta*. Empirical tests have shown that only differences between adjacent MB's QP value above an approximate value of 30 will cause the bitstream to be invalid, reporting the said variable to be out of range. As these are values too high to be practically meaningful that is not a problem to the current implementation.

3.2.4 Input of Region Definition in the Encoder

In order to accommodate the information retrieved from the file (binary mask) identifying the regions of perceptual relevance and also from the other possible options explained before (for transcoding and/or adaptation) new data structures were created. This structures, defined in the also newly created *transcoding.h* header file, are able to accommodate information outputted by a modified decoder (namely motion vectors) but are also responsible for the accommodation

of the binary masks information. The main structure, defined as *transcoding_info*, will point the space where the MBs information is stored and also keep track, by calculating it at the beginning, of the number of frames and MBs per frame. A pointer to this structure was added to the previously introduced *VideoParameters* structure for easiness of access.

At the very start of the coding process, only after the initiation of the encoder, in case the asymmetric coding flag is enabled the program will try to read the file indicated as containing the MBs perceptual relevance information. This function, called *get_relevant_mbs* starts by calculating the number of MBs in a frame using the following expression, where *source.width* and *source.height* are, respectively, the width and height of a frame in pixels and *MB_PIXELS* is the number of pixels in a MB (defined as 256):

```
p_Vid->p_Transcod->MBcount = (p_Inp->source.width[0]*p_Inp->source.height[0])/
MB_PIXELS;
```

Then, the number of frames with a binary mask defined is calculated by dividing the total file size (obtained through a fseek function) by the size of a single frame, considering a frame, in bytes, has a size equal to it's MB count. This is done in function *get_nframes*. This is done as the amount of frames having a binary mask may differ from the amount of frames to code in the stereoscopic video and to enable retrieving the information of regions of perceptual relevance from all the frames at once.

At this point the allocation of memory to store a byte per MB in the sequence is done. This has been done such as there is a pointer per each frame (through a *malloc* memory allocation) and each one of these pointers points to another pointer that points the start of a vector of bytes (unsigned chars), where each byte corresponds to a MB. Writing to the allocated addresses is then done recurring to binary file reading functions, a frame at a time. This step finishes the reading of information of perceptual relevance regions and lets the program proceed with the encoding process, where this information will be used as explained in the previous section.

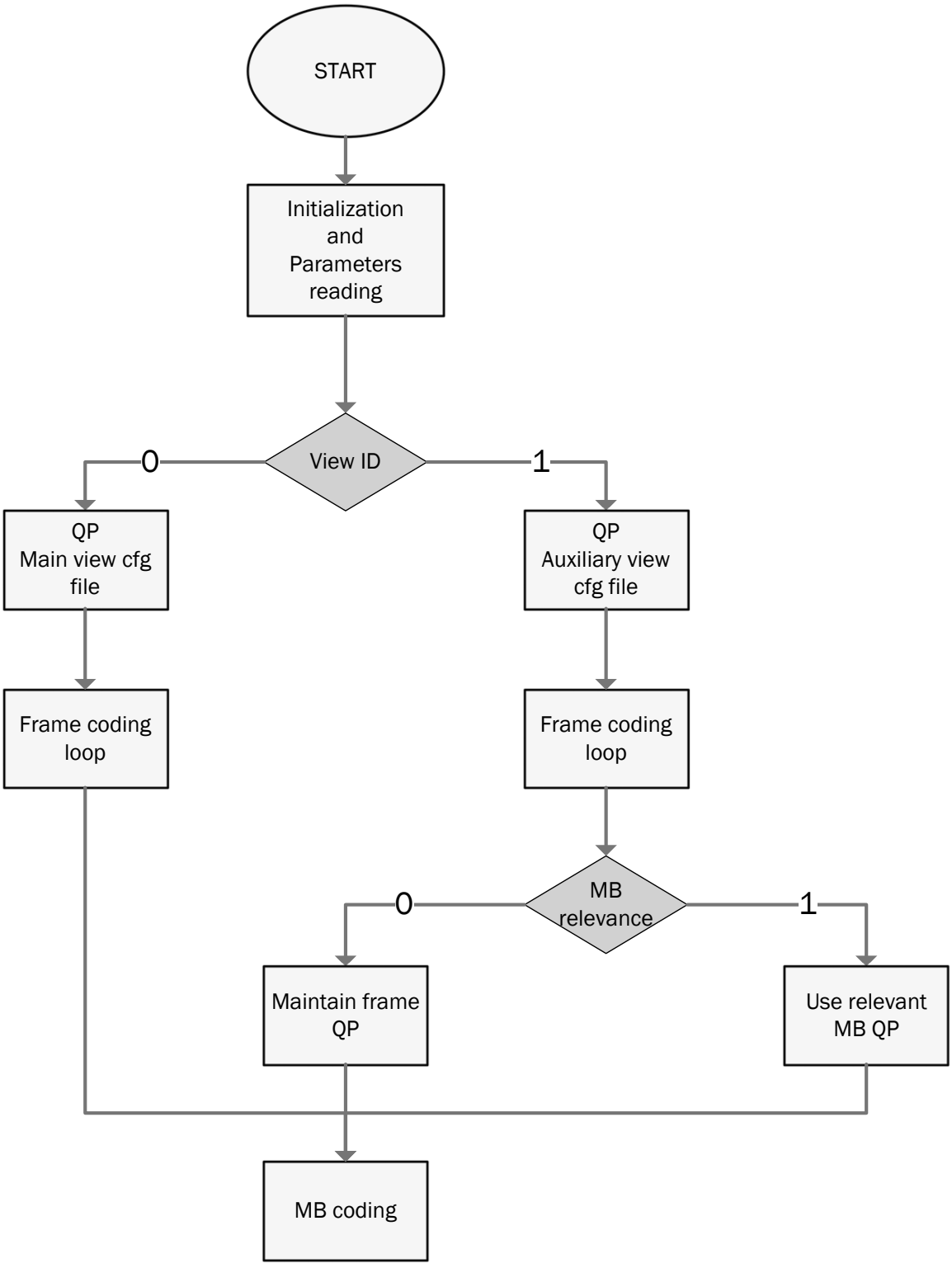


Figure 3.3: Simplified flowchart for the selection of QP to use in each MB.

3.3 Other Features and Support Software

Previous sections described the implementation of non-uniform SNR asymmetric coding through the manipulation of the QPs from MBs considered to have more or less perceptual relevance for the observer. This section will cover the implementation of other functionalities, namely in the decoder, for analysis and debugging purposes. Also some additional software developed in order to automate repetitive manual tasks, i.e., to present the video sequences and still images for subjective testing.

The decoder implementation, in opposition to the encoder, is the one predicted in the recomendation [2]. Thus it must follow the recommendation, leaving less room to maneuver within the code. As so its routines are simpler and more straightforward than the encoder ones. However, for a total of 38 header files and 51 source files, it is still quite a piece of code. The reference decoder receives an encoded video, which usually uses the previously mentioned *.264* extension, and a configuration file such as the one presented in Appendix C. Note that this example of a configuration file for decoding 20 frames of a VGA size encoded stereoscopic sequence does not need to be given that information. This happens because that same information is implicit in the encoded video file structure. Apart from features that have been implemented during this dissertation that are present in the example there's only a few parameters such as the name of the resulting YUV output file and it's format, the packetisation mode and the number of frames to decode, the use of simple concealment methods and the use of deblocking filters.

For debugging purposes and also to be able to easily retrieve additional information a version of the encoder was modified in order to output information such as bitrates per type of slice, QP and SNR statistics in organized tables at the end of each encoding. This can be easily done by accessing the routines and variables already existent, mainly in the report.c. On the other hand a significantly higher amount of information is outputted by the modified JM decoder. Namely the decoder has a data structure that contains the encoding parameters contained in the decoded bitstream. This includes a panoply of information from frame width and height to the encoding profile and thus can be used to automatically configure a follow up encoder if needed. The same happens to a structure that contains more specific information for each frame. Finally the motion vectors from each MB in a frame are also stored in a binary file that contains information that allows a encoder the identification and matching of that frame and MB to the one being encoded at the moment. Routines to read the information outputted by the decoder have been implemented in the encoder, however, the use of it has not. This can be done, i.e., by placing the motion vectors as initial guesses for the EPSZ block motion search routines (files me_epzs_int.c and me_epzs_sub.c and changing the encoding initialisation functions both for the sequence and for each frame to include some of the information coming from

the decoder.

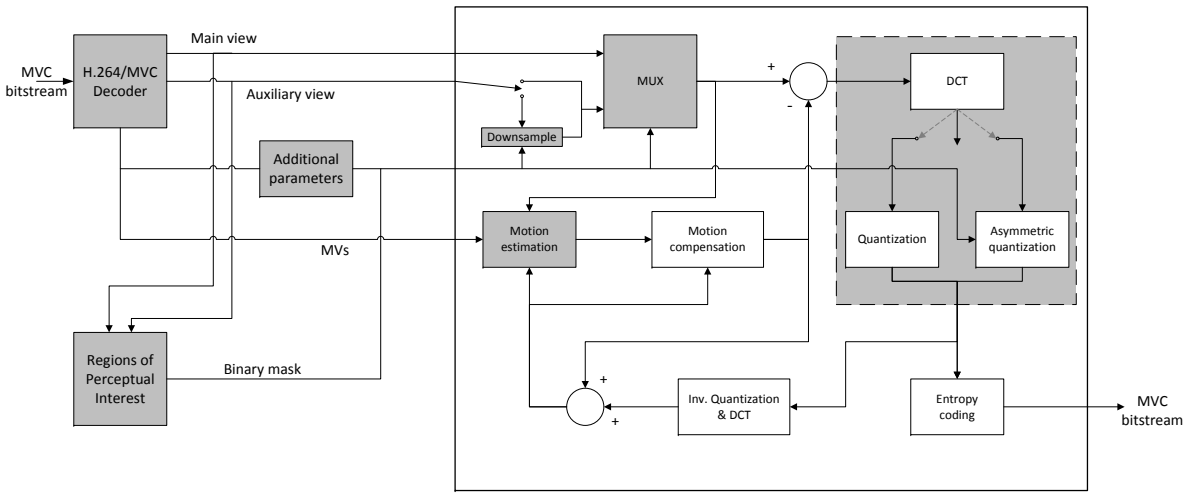


Figure 3.4: Schematic representation of the modified encoder and decoder.

Figure 3.4 presents a possible application for both the method of non-uniform asymmetric coding presented in this dissertation but also for some of the other features described in the paragraphs above. In this application a smart transcoding or adaptation framework is implemented in order to, according to the network feedback, control the type of asymmetry used and the quality output of the following encoder (through the additional inputs created) in order to adjust the bitrate to the target network or end device.

In the Figure a H.264/MVC decoder is presented in the top-left corner, which receives a standard MVC bitstream containing a stereo video and outputs the raw views and additional information provided in the MVC bitstream received by the decoder. Information such as QP and coding modes are added to the additional parameters block, which provides the configurations to run the encoder, including the control of asymmetry. The right-half large square represents the modified JM encoder, using both modifications described before (implementation of SNR-asymmetric encoding and routines to read the information outputted by the decoder). SNR-asymmetric encoding is enabled through the use of the binary mask calculated in the Regions of Perceptual Interest block, through the use of both raw views. The implementation of additional asymmetric coding methods may also be performed in this scheme. One of the simplest ways of achieving spatial asymmetry is to implement a downsampling cycle at the auxiliary view input but, as seen before, normally results in a non-normative encoder. The implementation of temporal asymmetry can be easily done by dropping auxiliary view frames at the input view multiplexer and replacing them by the previous one. Doing this will cause the second

(or following) of the two (or more) equal frames in a row to have a coded rate almost negligenciable due to interframe predictions motion estimation values being null. Finally, the use of motion vectors (MVs) outputted by the decoder can be used in the motion estimation process to reduce the computing time as, once identified, can be fed, i.e., as accurate initial estimations to the motion prediction schemes of the encoder, namely the Enhanced Predictive Zonal Search Algorithm (EPZS) routines, drastically reducing the amount of iterations needed to retrieve the optimal motion predictor.

In order to setup subjective assessment tests new software was developed and also both software previously developed and freely available software were used as well. Additional software created by members of the workgroup was also used in 2 operations. First, to calculate the estimate of disparity and output the first stage pixel-based mask. The follow up mask operations are done through Matlab scripts. Second, to concatenate the encoded views (once decoded) side-by-side. In order to present these sequences with Nvidia 3D Vision software they were encoded using the ffmpeg software, freely available at [56]. In case of still images the frame is replicated in order to provide a 10 second stereoscopic video that can be shown in the existing subjective assessment framework. An example of this operation may be seen below.

```
copy Balloons-SbS-HQ.yuv Balloons-SbS-HQ-250f.yuv
FOR /L %%G IN (1,1,249) DO
copy /b Balloons-SbS-HQ-250f.yuv+Balloons-SbS-HQ.yuv Balloons-SbS-HQ-250f.yuv
ffmpeg.exe -s 2048x768 -i Balloons-SbS-HQ-250f.yuv -vcodec copy -r 25 -y
    Balloons-SbS-HQ.avi
del Balloons-SbS-HQ-250f.yuv
```

To setup subjective assessment sessions, in particular to create playlists with the test content, a previously created software toolbox was used. One of the tools it provides, given a text file with the name of the file for each sequence, can create several scrambled lists that are be used to define the playing order of the content in the test session. This is done by feeding the playlist to a new software developed for the purpose that reads the playlist in the defined format and is able to manage the reproduction of each sequence and the respective reference at the given time according to the DSIS method.

A software that computes the PSNR of each relevance regions has also been developed. It has been used in order to try and find a definable threshold for the performance of non-uniform and uniform asymmetric coding methods based on the PSNR difference of the relevance regions. Although a threshold was not found to this point the results the software provided can be seen in Appendix E. The program uses as input the original and encoded auxiliary views of the video sequence as well as the binary mask that defines the regions of perceptual relevance and the additional parameters spatial resolution, YUV

format and the component of the video in which to compute the PSNR (usually Y).

Chapter 4

Asymmetric Coding of Stereoscopic Still Images

As seen in previous chapters 3D stereoscopic signals require a huge amount of bandwidth and storage capacity but this burden can be reduced if the latest video codec are used. Furthermore the properties of the HVS can be exploited in order to further reduce the encoded video bitrate while maintaining the video's perceived quality. This can be achieved through asymmetrical coding of the different views of the video.

In this chapter, we further exploit these asymmetric properties of the HVS by defining stereoscopic image regions with different perceptual relevance. These regions are used to identify the areas (at a MB level) of the stereo pair with higher impact in the observer's perceived quality. This is based on the underlying idea that higher disparity regions consist mainly of objects' borders and their neighboring regions and thus these are the most important regions for 3D perception. In the encoding process, for auxiliary view images, the regions with higher perceptual relevance are encoded with higher quality and lower relevance regions are, on the other hand, coarsely encoded. This means that a new level of granularity is being added to traditional asymmetric coding, based on frames with different qualities.

Such regions are identified by a MB-based binary mask based on an estimate of disparity determined by absolute difference between stereo images followed by pixel-to-macroblock expansion and a closing morphological operation, as seen in Section 3.1. Then, the mask is fed to a modified and standard compliant MVC encoder is used to selectively encode each region such that those with lower perceptual relevance are encoded with lower quality, using a higher QP, and those with higher perceptual relevance are encoded with higher quality, using a lower QP.

We aim to prove demonstrate with this technique, for certain coding regions and for a given subjective quality, it is possible to reach higher compression gains than if traditional

uniform asymmetric coding methods (i.e., the whole image has roughly the same quality) were used. To support this goal an experimental study based on subjective assessment of stereoscopic content with different characteristics was performed. The study described in this chapter deals with stereoscopic still images, corresponding to single frames taken from known 3D video sequences. This framework can be summarized in the following 3 man steps.

- Two regions with different perceptual relevance (high or low) are computed in each pair of stereoscopic frames. Such regions are identified by using a binary mask.
- This binary mask, once processed and given to the encoder, defines the regions where the target quality is higher or lower, i.e., the binary mask is fed to a modified H.264/MVC encoder (JM 18.0) along with the 2 views of the stereo content.
- Finally, subjective testing is performed to assess the perceived quality of the 3D images encoded with regions of perceptual relevance (non-uniform asymmetric) when compared with symmetric and uniform asymmetric coded sequences.

To enable the asymmetric coding of stereo images with this additional level of asymmetry, the previously introduced modified H.264/MVC reference encoder (JM version 18.0) was used. The changes implemented enable SNR-asymmetric coding by allowing the configuration of different QP for the main view and for each one of the regions of relevance in the auxiliary view. Furthermore this modified encoder receives final MB-based binary mask as input auxiliary data for each stereo pair to differentiate the compression ratio of the two different image regions of the auxiliary view, according to their perceptual relevance. This maintains normative decoding compliance.

In the following sections the subjective assessment framework, including material and methodology, will be presented. Then, the results it provided will be analyzed and discussed. Finally the dominant eye of each observer will be taken into account in order to assess the impact of this factor in the subjective quality ratings.

4.1 Experimental Framework

4.1.1 Test Material

As stated before a subjective experimental study was carried out in order to compare the performance of uniform and non-uniform asymmetric coding methods in stereoscopic images. Eight different stereoscopic sequences (as shown in table 4.1) were used to provide single stereo images, allowing a relevant amount of different scenes. The coding scenes used in the assessment sessions were divided in the following coding methods:

- Symmetric coding;
- Uniform asymmetric coding;
- Non-uniform asymmetric coding.

In each coding method several operating points are considered, each having a set of common characteristics and sharing asymmetry modes but slightly differing in the QP configuration. There are 2 symmetric coding operating points, the HQ reference and a second one with both view at approximately 37dB PSNR. The HQ symmetric reference is used as the input for encoding. Non-uniform operating points, in a total of 5, were obtained from the first uniform asymmetric operating point configurations (higher quality of this method) by increasing the QP in the lower relevance regions of the auxiliary view in steps of 3 (total of 3 points). The additional lower quality uniform asymmetric operating points (which total 3) have approximately the same auxiliary view objective quality of the third and fifth non-uniform asymmetric coding points, approximately 35 and 33dB PSNR respectively, for better comparison. Apart from the QP configuration every encoded used the exactly same parameters. An example of a configuration file for a video sequence using the same coding parameters (apart from QP) can be seen in Appendix B. The subjective testing was divided in 2 assessment sessions and to maintain high correlation between the subjective scores of the 2 sessions the 40dB symmetric reference and soft uniform asymmetric coding points were used as anchor points. This was done as a first session showed that additional data was needed (using different RD points with more severe asymmetry) to be conclusive.

Table 4.1: Characteristics of the evaluated images.

Sequence	Resolution
Balloons	1024x768
Bike	1024x576
BMX	1024x576
Cafe	1024x576
Car	1024x576
Champagne Tower	1024x768
Kendo	1024x768
Notebook	1024x576

4.1.2 Subjective Assessment Methodology

The subjective assessment sessions took place in a room with conditions conforming to [33]. The stereo images were displayed at 120Hz using a 3D capable projector synchronized with shutter glasses. The sequence order was randomized before the assessment to minimize bias. The assessors sit at a distance of approximately 3 meters from the center

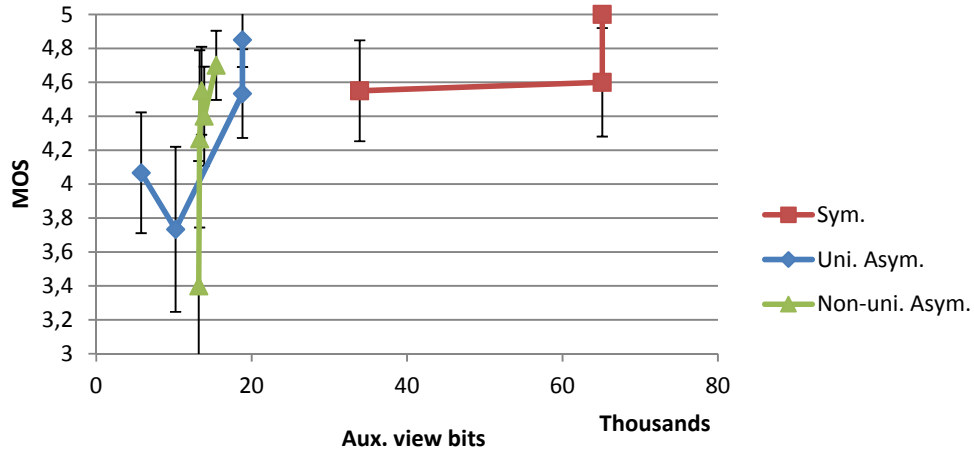
of the screen and were allowed to adjust their position to the most comfortable viewing position.

According to the DSIS method variant I, each stereo image was displayed for 10 seconds with a 3 second mid-gray interval between the reference and the impaired one. As so, the reference image was presented followed by the mid-gray interval, the impaired image and the evaluation period, in this respective order. The participants were given up to 10 seconds to evaluate each sequence according to the 5 point discrete scale: (5) imperceptible, (4) perceptible but not annoying, (3) slightly annoying, (2) annoying or (1) very annoying. Before the subjective assessment session the participants were given written instructions explaining the procedure and the rating scale and also spoken instructions making them aware of possible impairments. A training period with different sequences preceded the evaluation session to allow the participants to train and make them familiar with the procedure. The results from these training sessions were not included in the evaluation set of results. After this initial period, the participants were free to ask any question before the start of the actual test. For each participant the total time spent in an evaluation test was about 30 minutes.

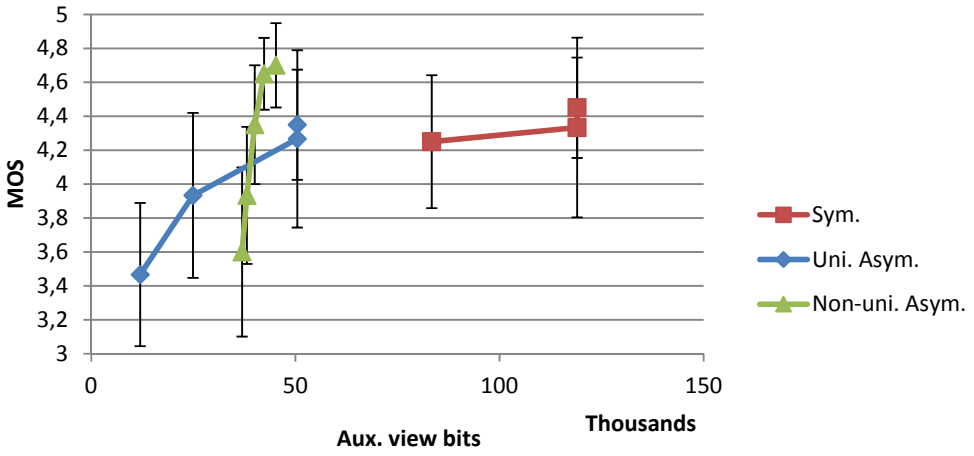
All participants in the subjective tests were screened for normal stereopsis using a random dot stereogram test. All the 35 participants, 32 males and 3 females, divided between the two subjective tests sessions, passed the random dot stereogram test and were then presented with both synthetic and natural 3D sequences, in order to allow them to adjust to stereoscopic viewing conditions. The age of the participants ranges from 21 to 47 years old, with a mean of 26.3 years old. The participants are, mainly, students, teachers and staff from the university campus, with no expertise in the area. None of the participants was rejected based on the screening for the DSIS method proposed in [33].

4.2 Analysis of the Results

In figure 4.1 the mean opinion scores (MOS) are plotted against the size (in bits) of the coded auxiliary view of the stereoscopic picture individually for sequences Balloons and BMX. Subjective scores for symmetrically coded images are shown in red, while the subjective scores for uniform asymmetric and non-uniform asymmetric with regions of perceptual relevance coding points are in blue and green, respectively. Confidence intervals of 95% are also displayed for each point. In this figure it is possible to see that, both objectively and perceptually, traditional uniform asymmetric coding is better for lower bitrates. However, when working at higher rates the use of preceptual relevance regions to encode with better quality the regions of the picture (auxiliary view) more relevant to the observer's 3D perception of quality results in better subjective results



(a)



(b)

Figure 4.1: MOS versus auxiliary view size for images Balloons (a) and BMX (b).

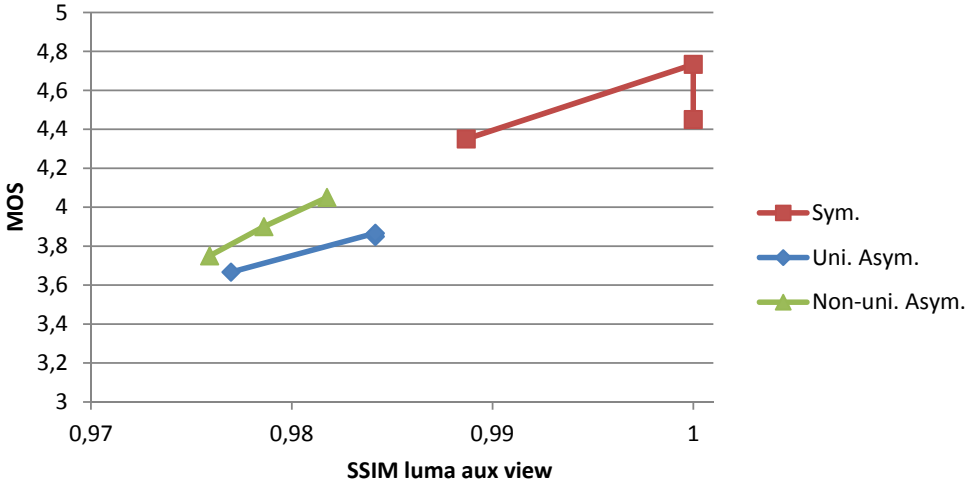


Figure 4.2: MOS versus SSIM for image Kendo.

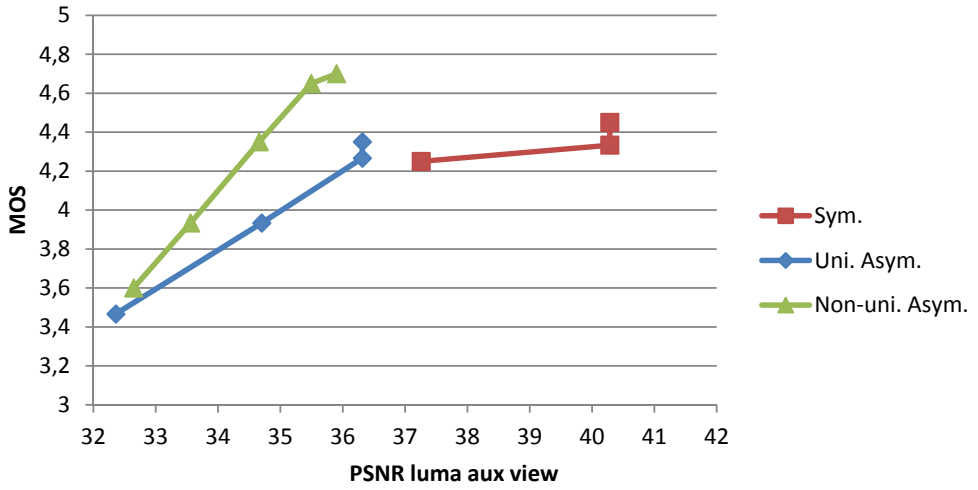
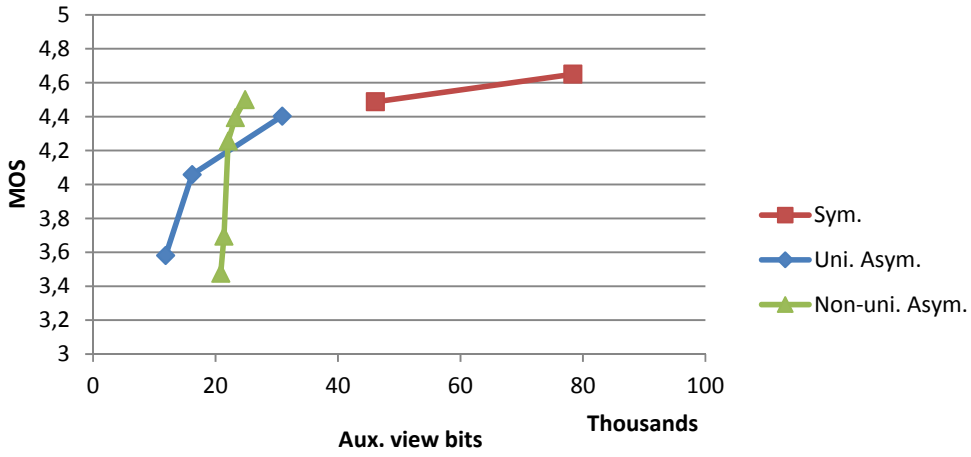
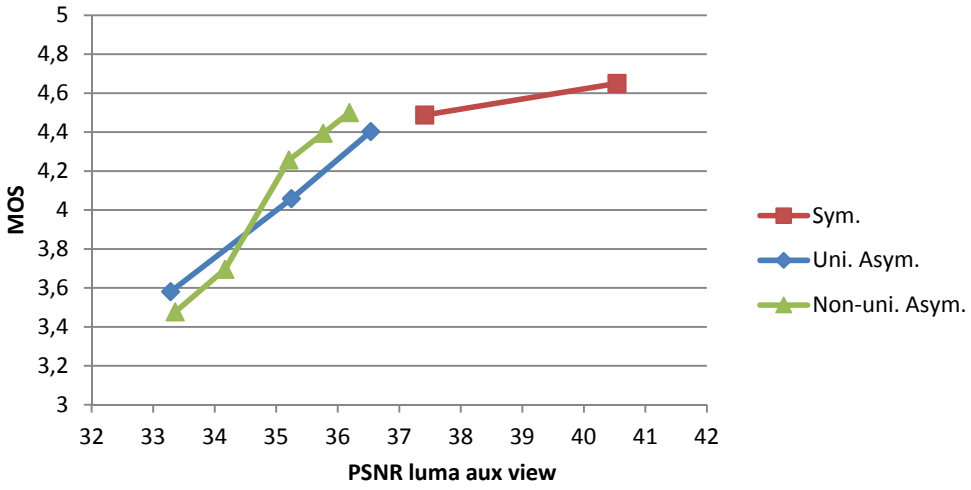


Figure 4.3: MOS versus PSNR for image BMX.



(a)



(b)

Figure 4.4: Mean subjective scores across scenes according to auxiliary view size (a) and PSNR of the auxiliary view (b).

than uniform asymmetric coding for the same rate. This can be observed in 6 out of the 8 cases tested, Cafe and Car being the exceptions. However, it is not possible to define a threshold in coded image size where the method with the better performance changes. In most cases the first 2 or 3 test points using non-uniform asymmetric coding (i.e., higher quality) provide better subjective results than the first points using uniform asymmetric coding and the opposite happens to the last 2 to 3 test points.

The structural similarity (SSIM) results were computed in the luminance channel of the auxiliary view for every test point and are displayed in figure 4.2 for the Kendo image. MOS scores for the non-uniform asymmetric operating points across all test scenes are generally better than those of the similar SSIM test points using uniform asymmetric coding. The only exception happens for the Car image. All test points for each scene with the higher SSIM scores were also rated higher in MOS when coded using perceptual relevance masks. In 5 out of 8 scenes the mask coding also provided better subjective scores at lower SSIM points. Overall, the SSIM scores for all scenes and test points were rather high, corroborating the theory that both techniques provide an efficient way to reduce image/video size/rate while maintaining a good subjective quality. This can be seen by the MOS scores that are, in the vast majority, above 3 (corresponding to the slightly annoying rating).

PSNR was also measured in the luminance channel of the auxiliary view and MOS scores in function of the PSNR is plotted figure 4.3 for the BMX image. The results shown are similar to those obtained with SSIM. However, when looking at the PSNR axis, the objective quality range is much bigger than using SSIM. This can be explained because SSIM is more related to subjective scores and for that reason results in overall high scores. On the other hand, PSNR is more sensitive to distortions in pixel values that may not have strong impact to influence the subjective opinion of the observers, resulting in a broader range of values.

It could also be seen in auxiliary view image size graphics (Figure 4.1) and both SSIM and PSNR (Figures 4.2 and 4.3) ones that in half of the test images (Balloons, Bike, BMX and Champagne Tower) the higher quality (rate, SSIM and PSNR measured) test point with non-uniform asymmetric coding achieved better subjective quality than symmetric coding with higher objective quality (approximately double the rate or 2dB PSNR more). In the remaining images the subjective quality was considered equal or only slightly worse.

The mean subjective scores for all images are presented in figure 4.4. One can see that, generally, the proposed non-uniform asymmetric coding achieves better subjective scores for higher bitrates than uniform asymmetric coding. This can be seen, i.e., in the first (right to left) operating point in non-uniform asymmetric coding when compared with the same operating point in uniform asymmetric coding. Despite the two operating points

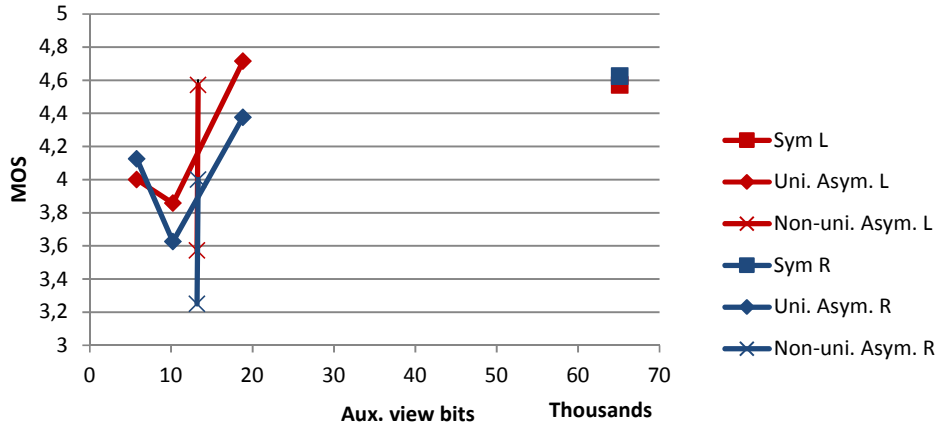
having approximately the same PSNR the non-uniform asymmetric coding one presents a considerable decrease in the size of the image and, even so, a slight higher MOS. This gain is high enough to achieve the same subjective score than symmetric coding where the auxiliary view has double the rate. In measured PSNR that means the same subjective score for an objective quality of the auxiliary view almost 2dB below. This suggests that non-uniform asymmetric coding can be efficiently applied resulting in better subjective performance than a similar rate or objective quality uniform asymmetric coding. Also it it shows it is possible to maintain the subjective quality of symmetric coding at a significantly higher rate. On the other hand when dealing with lower image sizes uniform asymetric coding provides better results. Non-uniform asymmetric coding at these smaller image sizes was accomplished using higher discrepancies between the QPs of the high and low perceptual relevance regions. In these operating points it was noticed that reducing the quality (by increasing the QP) only in the regions of low perceptual relevance does not result in significant reduction in the image size, mainly due to the characteristics of these regions, however it results in a severe MOS decay.

4.3 The Influence of the Dominant Eye

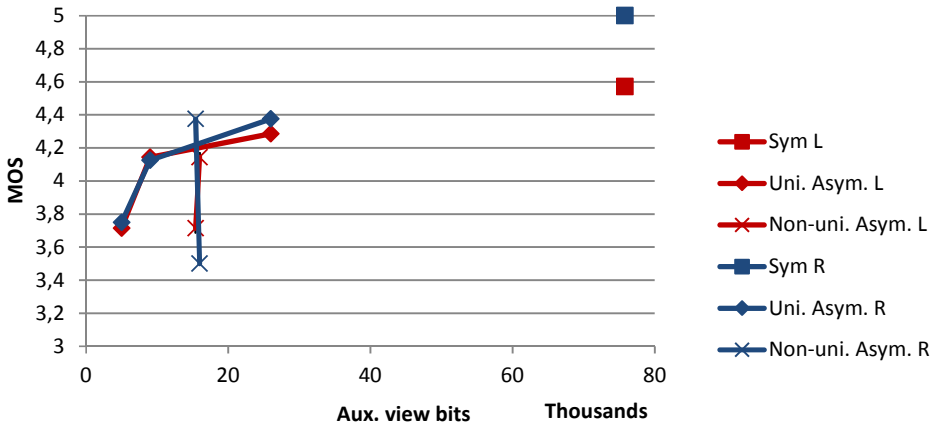
In this section we analyse the impact of the dominant eye factor in the subjective perception of quality of the observers in asymmetrically coded content. Both uniform asymmetric coding methods and the method proposed in this work result in a view of the stereo pair being worse in quality than the other. It has been proved before (as seen in section 2 that in the presence of two different quality inputs the HVS tends to mask the artifacts in the lower quality one. However it may be expected that people that have eye dominance in the eye corresponding to the lower quality view (the auxiliary one in this case) are more sensible to the artifacts in that same view.

In order to assess the impact of this factor in the method presented the observers taking part in one of the subjective assessment sessions were asked to do a simple Miles test [57] and to register their dominant eye in the given forms. To perform the Miles test participants are asked to extend their arms in front of them and make a small triangle between their hands. With both eyes open they aim at an object in the screen plane (a red circle in this case). By blinking an eye the eye dominance is found to be on that same eye if the object moves off the triangle or on the open eye if it remains within the triangle.

From the 15 test participants 8 of them were found to be right-eye dominant and 7 to be left-eye dominant. This test was performed during one of the assessment sessions described in section 4.1 and, as so, the observers watched 6 operating points per scene: symmetric coding HQ reference, 3 uniform asymmetric coding points and the lower quality



(a)



(b)

Figure 4.5: Individual results for left and right eye dominant observers for images Balloons (a) and Bike (b).

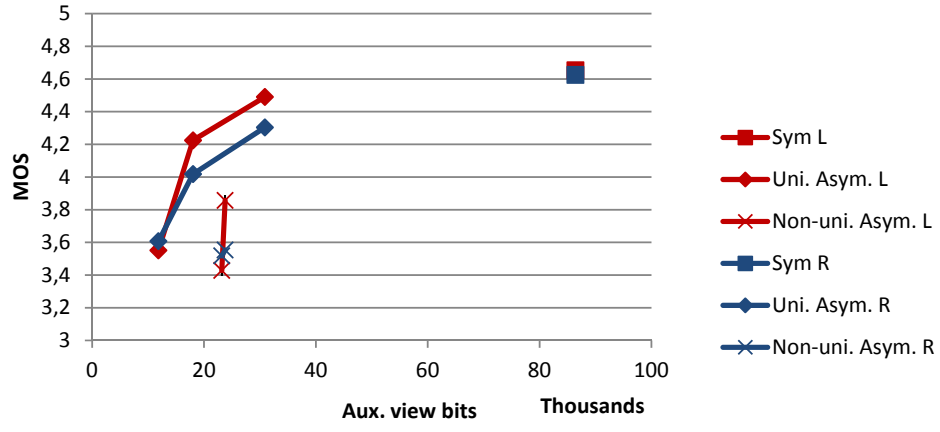


Figure 4.6: Mean results for left and right eye dominant observers across all scenes.

2 non-uniform asymmetric coding points using regions of perceptual relevance (this means we are testing the method at the maximum difference between views). Image Kendo was not considered in the results as there was no equivalent non-uniform asymmetric coding points due to QP restrictions.

In Figure 4.5 the individual results for images Balloons and Bike are presented, while in Figure 4.6 are presented the averaged results for all images. In both figures results from left-eye dominant participants can be seen in red and right-eye dominant participants have their scores shown in blue. For each type of eye dominance the 3 coding types are shown in a different line (symmetric, uniform asymmetric and non-uniform asymmetric coding). The greater difference in the subjective scores can be found in the case of Balloons. In 4 of the 5 asymmetrically coded points left-eye dominant observers gave higher scores to the stereoscopic images than right-eye dominant observers. However this difference is, on average, 0.25 on a 0-5 scale, which is not significant. On the other hand, in images like Bike no noticeable differences can be found. From the mean results one can see that differences in results are too small to clearly state that the observers with eye dominance corresponding to the view that has lower quality in the asymmetrically coded stereo pair are more affected than those who have eye dominance corresponding to the higher quality view. However one must note that the exhibited content has a short duration. Additional testing may be required to assess if the eye dominance factor becomes more relevant in the presence of longer stimulus. There is the hypothesis that, besides quality perception, eye dominance corresponding to the view with the lower quality may result in faster eye strain.

Chapter 5

Asymmetric 3D Video Coding using Regions of Perceptual Relevance

This chapter deals with asymmetric coding of 3D video using the same principles as described on the previous one. The aim of this work is to demonstrate that non-uniform asymmetric video coding based on regions with different perceptual relevance extends the current state-of-the-art concept of SNR-asymmetric coding, which uses uniform coding in the auxiliary view. In order to achieve this goal a new assessment framework was developed. The new framework is, similarly to the previous chapter, based in subjective assessment of the 3D video quality and uses the same non-uniform asymmetric coding method presented in this dissertation.

In the following sections the experimental framework, including material and methodology, will be presented. Due to the methodology being similar to the previous chapter only the differences will be explained below. Then results are presented and discussed, showing that the non-uniform asymmetric coding method presented in this dissertation can be effectively used to extend the state of the art SNR-asymmetric coding.

5.1 Experimental Framework

5.1.1 Test Material

The subjective testing was carried out using the stereoscopic 3D video sequences Balloons, Champagne Tower and Kendo (see Table 4.1 for spatial resolution information). These sequences differ greatly in content, motion and depth (disparity between views), which allows quite different variability in the type of content. The JM 18.0 modified encoder was used to encode each of the three sequences at different operating points: symmetrically (1 point), asymmetrically with uniform QP in the auxiliary view (1 point) and asym-

metrically using the proposed non-uniform asymmetric coding method with regions of perceptual relevance (3 points). The encoding configuration used during this framework is similar to that presented in the example of Appendix B for all the encoding processes, with a GOP of 10, 3 reference pictures and 7 B-type frames (per view), using a stereo-high profile.

The regions of different perceptual relevance are defined in each stereo pair by a binary mask in a similar way to the framework previously described. However, in this case the masks were not subject to morphological closing, as these tests were performed before that implementation. This results in areas of relevance that are more irregular in both it's contour and it's filling. The implementation of morphological closing happens after these tests have been made in an attempt to solve the lacy aspect of the regions of perceptual relevance when at a MB level and to to make them fit better with the objects they are representing. The images of the steps leading to the creation of the binary masks defining the regions of perceptual relevance can be observed in Figure 5.1. This process, in exception to the morphological operations, is the same as the one described in section 3.1 and it resulted in the following percentages of MB inside the region with higher perceptual relevance (i.e., encoded with lower QP):

- Balloons − 25.18%;
- Champagne Tower − 39.96%;
- Kendo − 18.67%.

The 5 operating points per sequence were defined as follow: symmetric coding at 37dB in both views (used as reference for the subjective testing), uniform asymmetric coding at 37dB PSNR (main view) and 33dB PSNR (auxiliary view) as a starting configuration for comparison with non-uniform asymmetric coding defined by regions of perceptual relevance and, finally, non-uniform asymmetric coding points achieved by increasing the quantisation parameter of the MB in the region with lower relevance. These values follow from a previous study on uniform asymmetric coding [41], corresponding to the maximum quality asymmetry that can be used without being noticed by users. In this work, such limit is further extended by encoding different regions with different quality, according to their perceptual relevance within a view frame. To achieve the different operating points the quantisation parameter of the MBs in the region with lower relevance were increased in steps of either 2 (Kendo) or 3 (Balloons and Champagne Tower).

5.1.2 Subjective Assessment Methodology

Following the creation of the test material subjective assessment took place using a similar methodology and the same equipment as explained in subsection 4.1.2. The assessment

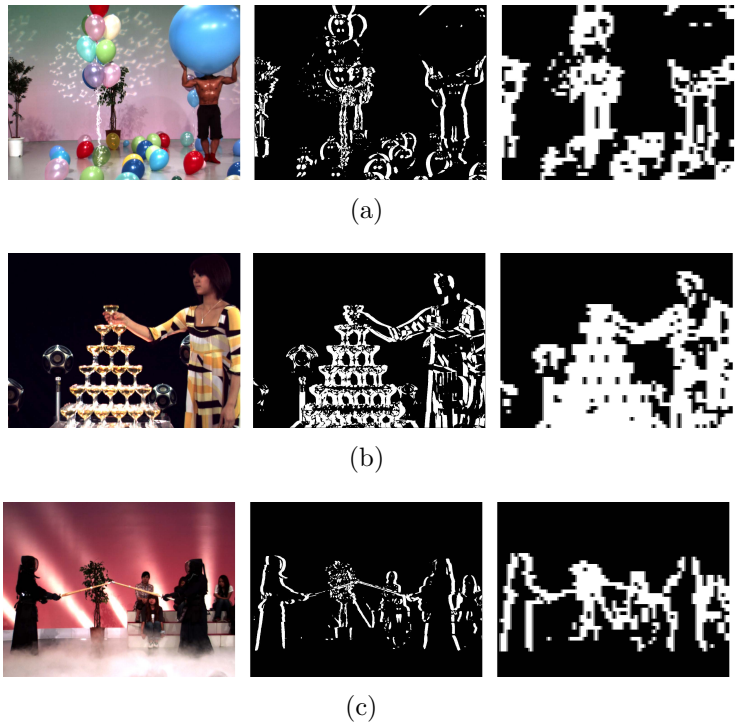


Figure 5.1: From left to right: view frame, pixel mask and MB mask for sequences: (a) Balloons, (b) Champagne Tower and (c) Kendo.

room conformed to [33] and 3D videos were displayed at 120Hz using a 3D capable projector synchronized with shutter glasses. The assessors sit at a distance of approximately 3 meters from the center of the screen and were allowed to adjust their position to the most comfortable viewing position. The sequence order was randomized before the assessment to minimize bias.

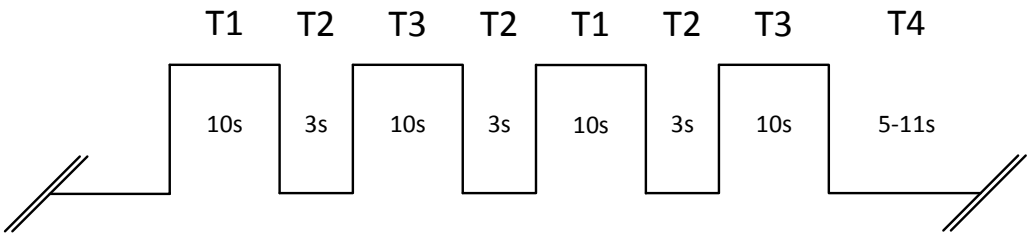


Figure 5.2: DSIS presentation variant II.

However, in this assessment session DSIS presentation variant II was used. This way, and as we are dealing with short stereoscopic videos instead of still images, observers have a second chance and thus more time to identify little difference which leads to more accurate results. This means that each stereo video and it's respective reference were displayed twice, each for 10 seconds with a 3 second mid-gray interval between each other, and followed by an evaluation period up to 10 seconds. The temporal scheme for each

video sequence using this method can be seen in Figure 5.2, in which T1 is the reference sequence, T3 is the impaired sequence, T2 are intervals and T4 is the evaluation period. The 3D video sequences were evaluated according to the same 5 point discrete scale: (5) imperceptible, (4) perceptible but not annoying, (3) slightly annoying, (2) annoying or (1) very annoying. Written and spoken instructions about the assessment procedure, rating scale and possible impairments was given prior to a training period in which results were not included in the evaluation set of results. Preceding the start of the actual test participants were also given a time to ask any question. The total evaluation time for each participant was no longer than 30 minutes.

19 observers participated in the subjective testing, 16 of them male and 3 female, with age ranging from 21 to 46 years old, with a mean of 25.5 years old. The participants passed a random dot stereogram test proving to have normal stereopsis and are naive to video subjective assessment. Both synthetic and natural 3D sequences were provided in order to adjust participants' viewing to the room and 3D viewing conditions. None of the participants was rejected based on the screening for the DSIS method proposed in [33].

5.2 Analysis of the Results

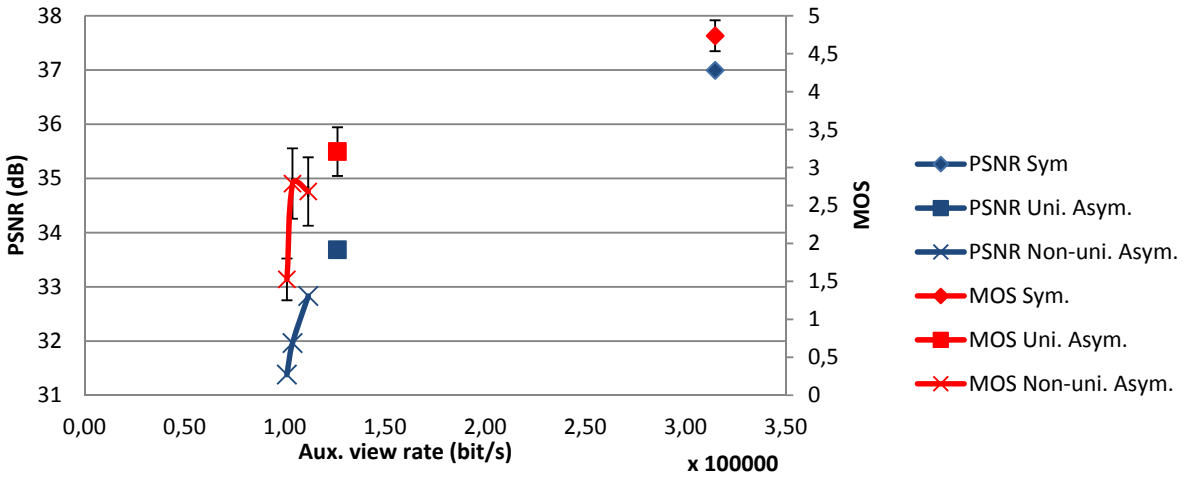
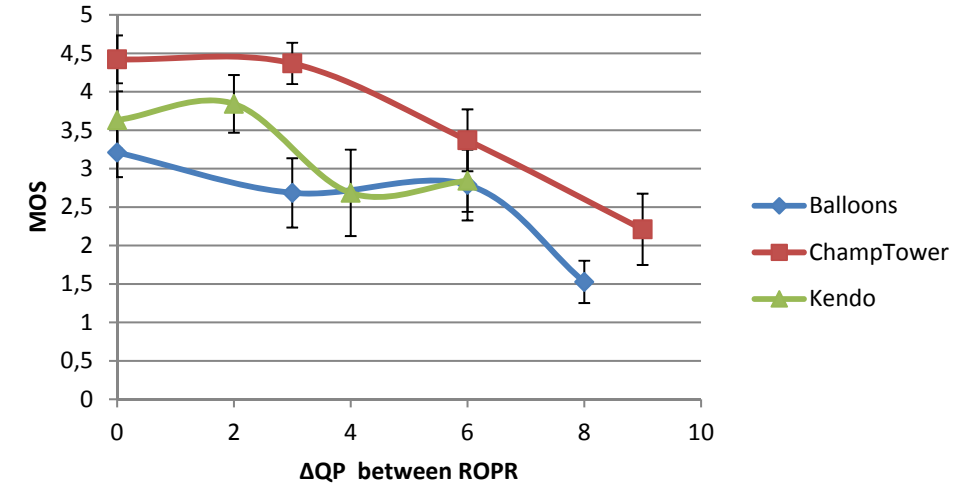
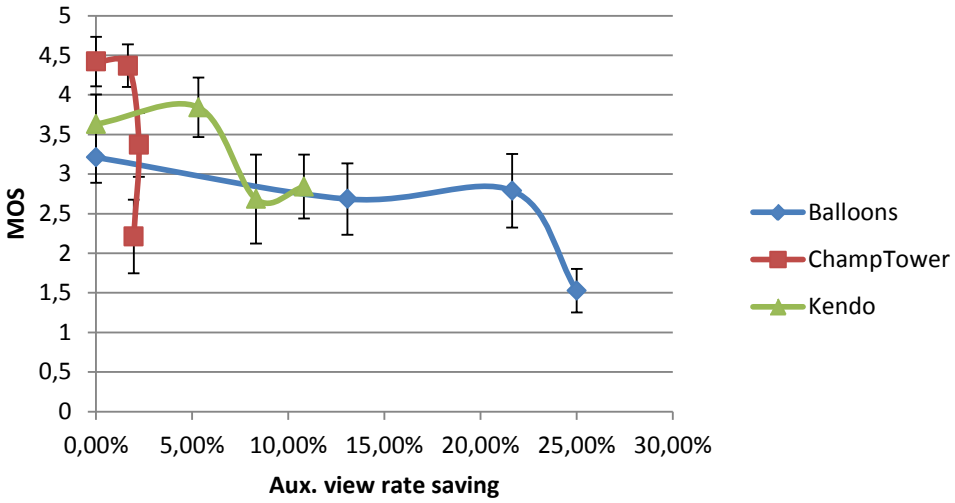


Figure 5.3: R-D and MOS results for the sequence Balloons.

Figure 5.3 presents the subjective scores, MOS (red lines and right Y axis), and the RD results (blue lines and left Y axis), for the sequence Balloons. The MOS scores for each test point along with the 95% confidence interval of the statistical distribution are shown in the figure for each type of coding (symmetric, uniform asymmetric and non-uniform asymmetric) according to the legend. From this figure one can confirm that these operating points correspond to the boundary of the perceptual threshold in these video sequences as the MOS rapidly decreases after the first few test points (i.e.,



(a)



(b)

Figure 5.4: MOS as function of: ΔQP between regions of perceptual relevance (a) and rate saving in the auxiliary view (b).

moving from right to the left in the graphs). It can be seen, i.e., in sequence Balloons that for PSNR values at below 32dB the MOS decreases rapidly reaching as low as 1.5 (below the annoying level). Individually, sequence Balloons produced the most meaningful results, exhibiting the greater bitrate saving for the same approximate subjective quality. However, individual results for other videos (Kendo and Champagne Tower) may be found in Appendix F as well as complimentary data resulting from the subjective tests framework, such as the exact bitrates, rating scores, means and confidence intervals that resulted in the graphics plotted.

In figure 5.4 the subjective scores in MOS are plotted against the the ΔQP used to differentiate the regions of perceptual relevance (a) and the rate saving obtained by using the proposed method in comparison with traditional uniform asymmetric coding (b). In

the first case it is shown that a roughly constant MOS can be obtained when coarser quantisation outside the regions of perceptual relevance is used. In the cases where ΔQP is equal to 3 there is almost no change in the MOS scores. This can be seen in the second operating point in the graphic (left to right) in all the three videos. In the Balloons video a ΔQP up to 6 can be used without significantly affecting the subjective quality, as can be seen in the third point (left to right) in the blue line that displays a MOS almost equal to that corresponding to a ΔQP value of 0. Despite the fact that these results show that performance is content dependent, this figure provides evidence that non-uniform asymmetric coding using regions of perceptual interest can be used within certain limits, without affecting the perceptual quality of 3D video.

In the second case it is seen that bitrate savings up to 20% can be achieved with almost no perceptible losses in some videos. This can be seen in the video Balloons, third point left to right, where the auxiliary view bitrate is reduced in 21.6% at the cost of less than half a point in MOS. However the results are worse in other type of videos, namely Champagne Tower, as the rate saving can be as low as 2%. This happens because the regions coded with higher QP (i.e., lower relevance) mainly consist in the background (uniform black in this video) and, as so, raising the QP in this area does not result in rate saving. This is also accompanied by a sharp decrease in MOS, once again due to the large flat background area of this sequence, which basically is little affected by coding and concentrates all distortion effects in the main region where all observers tend to focus their attention. The remaining video under test, Kendo, sits between the previous 2 extreme cases, reaching about 5% rate saving for a ΔQP of 2 and for a MOS a few tithes above the uniform asymmetry case (ΔQP equal to 0).

The wide range characteristics of this behavior suggests that a finer classification of regions in each stereo image pair of 3D video might be necessary to not exceed the range of acceptable results. However in the majority of natural scenes flat (and mainly dark) backgrounds are not so common making it possible to efficiently use the proposed method to extend the maximum range of asymmetry achievable with uniform asymmetric coding.

Chapter 6

Conclusions

This chapter concludes this dissertation, presenting general conclusions about this research work. Also, some future research options are discussed regarding non-uniform asymmetric video encoding and it's uses.

6.1 Overview

In this work an experimental study was carried out on a new non-uniform asymmetric coding method based on regions of perceptual relevance in the auxiliary view of stereoscopic video. The proposed method was thoroughly described and, using both still stereo images and video sequences, the main propositions and objectives for this work were demonstrated.

As the results show, it is possible, to a certain extent, to achieve compression gains using the proposed non-uniform asymmetric coding according to our method for the same or higher subjective quality level of uniform asymmetric coding. As non-uniform asymmetric coding preserves the quality (lower quantisation) at the perceptually relevant regions of the auxiliary view the presented results for the method are better at higher rate coding points. For the same reason it is not possible to efficiently achieve lower rates with our method, specially when the encoded content has smooth backgrounds. However it was not possible to find a general threshold at rate, PSNR or SSIM level that states where the non-uniform asymmetric coding starts to perform worse than uniform asymmetric coding, in part due to variable content characteristics.

Is must also be noted that, for the same objective quality, measured in both SSIM or PSNR in the luminance of the auxiliary view, non-uniform asymmetric coding provides better subjective quality, for the majority of content and coding points, than uniform asymmetric coding. Therefore, non-uniform asymmetric coding based in regions of per-

ceptual relevance for the 3D perception of quality can be efficiently used to reduce the rate of the auxiliary view while maintaining (and even improving it in some cases) the subjective quality of the overall 3D scene.

Using the proposed method it is also possible to further explore the suppression theory of the HVS by extending the traditional asymmetric coding concept into the spatial domain. Using this new level of asymmetry based on regions of perceptual relevance in asymmetric coding it is possible to achieve up to 20% rate savings in the auxiliary view without harming the viewer's perceived quality on top of the maximum asymmetry range enabled by uniform asymmetric coding. In this case a a ΔQP from 3 to 6 between these two regions with different perceptual relevance should be used in order to avoid negative impact in the viewers' perception of quality.

Furthermore eye dominance corresponding to the view with lower quality has proven to have no significant impact on the the perception of quality of the observers who have eye dominance corresponding to the view encoded with lower quality. While in these short sequences/images it was not possible to show a definite impact it may be possible that for prolonged stimulus eye strain or discomfort start to appear.

6.2 Contributions

In this dissertation a review of the state of the art is given in chapter 2, providing the necessary background of the (3D) video coding, 3D technologies and 3D asymmetric coding areas. Methods to assess the subjective quality of 3D content are also presented and compared, including hints about the future of 3D technology.

A detailed description of the experimental framework used to implement asymmetric coding and conduct subjective tests using both still images and video sequences is given. The main results of this work were presented in two international conferences and one national conference. It was also awarded a *Best Poster Award* in the Plenoptics training school (July 2013, Sweden). The complete list of publications is included in Appendix D.

Overall it must be considered that this dissertation has contributed to the advance of the state of the art revolving asymmetric coding of 3D video. It also provides a valid starting point to future research as pointed out in the next section.

6.3 Future work

Future work should try to better define these newly found subjective thresholds in order to devise an algorithm that can preemptively know if it is more efficient to use the newly proposed method. Such algorithm can then be integrated in a non-uniform asymmetric

coding methodology in a standard coding framework. This framework can also be used to apply non-uniform asymmetric coding if the range of asymmetry is slightly greater than the maximum range of traditional uniform asymmetric coding, as was shown in video subjective assessment results.

A non-uniform asymmetry based framework may as well be included in a network node in order to perform adaptation or transcoding operations in an efficient way, by exploring some of the additional functionalities implemented in the codec and introduced in section 3.3. According to the network state or the end device the configuration settings can be automatically adjusted to better fit those characteristics.

Scene-dependent masks can also be a topic for future research. There are research works that aim to model the visual attention of a subject resulting in the so called saliency maps. An algorithm that calculates this kind of maps may as well be integrated and subjective assessments tests can be run in order to check if binary masks defined using saliency maps provide better results than those based on an estimate of disparity.

Finally, further research on eye dominance can be conducted. This can be done mainly using longer stimulus to assess if the perceived quality starts to degrade. Additional fields can be included in an observer inquiry to assess eye strain, visual discomfort or other type of body discomfort such as headaches.

Bibliography

[1] A. Vetro, A. Tourapis, K. Muller, and T. Chen, "3D-TV content storage and transmission," *IEEE Transactions on Broadcasting*, vol. 57, pp. 384 – 394, 2011.

[2] *Advanced video coding for generic audiovisual services, ITU-T Recommendation H.264 and ISO/IEC 14496-10 (MPEG-4 AVC)*, ITU-T and ISO/IEC JTC 1 Std., 2010.

[3] T. Wiegand, G. Sullivan, G. Bjontegaard, and A. Luthra, "Overview of the H.264/AVC video coding standard," *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 13, no. 7, pp. 560 –576, July 2003.

[4] Y. Chen, Y.-K. Wang, K. Ugur, M. M. Hannuksela, J. Lainema, and M. Gabbouj, "The emerging MVC standard for 3D video services," *EURASIP J. Appl. Signal Process.*, vol. 2009, pp. 8:1–8:13, January 2008.

[5] A. Vetro, T. Wiegand, and G. 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, April 2011.

[6] *High Efficiency Video Coding*, ITU Video Coding Experts Group (VCEG) and ISO/IEC Moving Picture Experts Group (MPEG) ISO/IEC JTC1/SC29/WG11 Std. ITU-T Recommendation H.265 and ISO/IEC 23 008-2 (HEVC), 2013.

[7] G. J. Sullivan, J.-R. Ohm, W.-J. Han, and T. Wiegand, "Overview of the high efficiency video coding (HEVC) standard," *IEEE Trans. Circuits and Systems for Video Technology*, December 2012.

[8] B. Julesz, *Foundations of Cyclopean Perception*. Univ. Chicago Press, 1971.

[9] L. Stelmach, W. J. 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, March 2000.

[10] W. J. Tam, "Image and depth quality of asymmetrically coded stereoscopic video for 3D-TV," in *JVT-W094, San Jose, CA*, April 2007.

[11] N. Holliman, N. Dodgson, G. Favalora, and L. Pockett, "Three-dimensional displays: A review and applications analysis," *Broadcasting, IEEE Transactions on*, vol. 57, no. 2, pp. 362 –371, June 2011.

[12] F. Speranza, R. Renaud, A. Vincent, and W. Tam, "Perceived picture quality of frame-compatible 3DTV video formats," in *IEEE International Conference on Multimedia and Expo (ICME)*, pp.640-645, Melbourne, Australia, July 2012.

[13] Text of ISO/IEC MPEG2011/N12543, "Additional profiles and SEI messages," San Jose, USA, February 2012.

[14] G. Ballocca, P. D"Amato, M. Grangetto, and M. Lucenteforte, "Tile format: A novel frame compatible approach for 3d video broadcasting," in *IEEE International Conference on Multimedia and Expo (ICME), pp.1-4, Barcelone, Spain*, July 2011.

[15] K. Muller, P. Merkle, and T. Wiegand, "3-D video representation using depth maps," *Proceedings of the IEEE, vol.99, no.4, pp.643-656*, April 2011.

[16] E.-K. Lee and Y.-S. Ho, "Generation of high-quality depth maps using hybrid camera system for 3-d video," *Journal of Visual Com-munication and Image Representation, vol.22, Issue 1, pp. 73-84, ISSN 1047-3203*, January 2011.

[17] C. Fehn, "Depth-image-based rendering (DIBR), compression and transmission for a new approach on 3DTV," in *Proc. SPIE Stereoscopic Displays and Virtual Reality Systems XI, pp.93-104, San Jose, USA*, January 2004.

[18] A. Smolic, K. Mueller, P. Merkle, P. Kauff, and T. Wiegand, "An overview of available and emerging 3D video formats and depth enhanced stereo as efficient generic solution," in *Picture Coding Symposium, pp.1-4, Chicago, USA*, May 2009.

[19] W. J. Tam, L. B. Stelmach, F. Speranza, and R. Renaud, "Cross-switching in asymmetrical coding for stereoscopic video," in *Stereoscopic Displays and Virtual Reality Systems IX, vol. 4660, pp. 95-104, Burlingame, CAL, USA*, January 2002.

[20] *Final Draft Amendment 3*, Amendment 3 to ITU-T Recommendation H.262 and ISO/IEC 13818-2 (MPEG-2 Video), MPEG document N1366 Std., September 1996.

[21] T. Chen and Y. Kashiwagi, "Subjective picture quality evaluation of MVC stereo high profile for full-resolution stereoscopic high-definition 3D video applications," in *Proc. IASTED Conference on Signal and Image Processing, Maui, HI, USA*, August 2010.

[22] Y. Su, A. Vetro, and A. Smolic, "Common test conditions for multiview video coding," in *Joint Video Team, Doc. JVT-U211, Hangzhou, China*, October 2006.

[23] G. Bjontegaard, "Calculation of average psnr differences between rd- curves," in *ITU-T SG16/Q.6, Doc. VCEG-M033, Austin, TX*, April 2001.

[24] D. Tian, P. Pandit, P. Yin, and C. Gomila, "Study of MVC coding tools," in *Joint Video Team, Doc. JVT-Y044, Shenzhen, China*, October 2007.

[25] Y.-J. Jeon, J. Lim, and B.-M. Jeon, "Report of MVC performance under stereo condition," in *ITU-T & ISO/IEC JTC1/SC29/WG11 Docs. JVT-AE016, London, UK*, 2009.

[26] T. Chen, Y. Kashiwagi, C. Lim, and T. Nishi, "Coding performance of stereo high profile for movie sequences," in *ITU-T & ISO/IEC JTC1/SC29/WG11 Doc. JVT-AE022, London, UK*, 2009.

[27] F. Bossen, B. Bross, K. Suhring, and D. Flynn, "HEVC complexity and implementation analysis," *IEEE Trans. Circuits and Systems for Video Technology, vol. 22, no. 12*, December 2012.

[28] M. Domanski, T. Grajek, D. Karwowski, K. Klimaszewski, J. Konieczny, M. Kurc, A. Luczak, R. Ratajczak, J. Siast, O. Stankiewicz, J. Stankowski, and K. Wegner, "Coding of multiple video+depth using HEVC technology and reduced representations of side views and depth maps," in *Picture Coding Symposium, Krakow, Poland*, 2012.

[29] ISO/IEC 23002-3:2007 - Information technology, "Mpeg video technologies - part 3: Representation of auxiliary video and supplemental information," 2007.

[30] A. Vetro and D. Tian, "Analysis of 3D and multiview extensions of the emerging HEVC standard," in *SPIE Conference on Applications of Digital Image Processing XXXV, Paper 8499-33, San Diego, USA*, August 2012.

[31] S. Yea and A. Vetro, "View synthesis prediction for multiview video coding," *Signal Processing: Image Communication, vol. 24, no. 1+2, pp. 89-100*, April 2009.

[32] F. Shao, G. Jiang, X. Wang, M. Yu, and K. Chen, "Stereoscopic video coding with asymmetric luminance and chrominance qualities," *IEEE Transactions on Consumer Electronics, 56, pp. 2460-2468*, November 2010.

[33] *Recommendation BT.500-12, Methodology for the subjective assessment of the quality of television pictures*, ITU-R Std., 2009.

[34] H. Brust, A. Smolic, K. Mueller, G. Tech, and T. Wiegand, "Mixed resolution coding of stereoscopic video for mobile devices," in *3DTV Conference: The True Vision - Capture, Transmission and Display of 3D Video, 2009*, May 2009, pp. 1 –4.

[35] J. Quan, M. Hannuksela, and H. Li, "Asymmetric spatial scalability in stereoscopic video coding," in *3DTV Conference: The True Vision - Capture, Transmission and Display of 3D Video (3DTV-CON), 2011*, May 2011, pp. 1 –4.

[36] Y. Chen, S. Liu, Y.-K. Wang, M. Hannuksela, H. Li, and M. Gabbouj, "Low-complexity asymmetric multiview video coding," in *IEEE International Conference on Multimedia and Expo*, April 2008.

[37] C. Fehn, P. Kauff, S. Cho, H. Kwon, N. Hur, and J. Kim, "Asymmetric coding of stereoscopic video for transmission over T-DMB," in *3DTV Conference*, May 2007.

[38] A. Aksay, C. Bilen, E. Kurutepe, T. Ozcelebi, G. B. Akar, M. R. Civanlar, and A. M. Tekalp, "Temporal and spatial scaling for stereoscopic video compression," *EURASIP European Signal Processing Conference*, 2006.

[39] L. Stelmach, W. Tam, D. Meegan, A. Vincent, and P. Corriveau, "Human perception of mismatched stereoscopic 3D inputs," in *International Conference on Image Processing (ICIP)*, vol. 1, 2000.

[40] F. Shao, G. Jiang, M. Yu, K. Chen, and Y.-S. Ho, "Asymmetric coding of multi-view video plus depth based 3-D video for view rendering," *IEEE Transactions on Multimedia*, vol. 14, no. 1, pp. 157 –167, February 2012.

[41] G. Saygili, C. Gurler, and A. 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, June 2011.

[42] ——, "Quality assessment of asymmetric stereo video coding," in *17th IEEE International Conference on Image Processing (ICIP)*, September 2010.

[43] ——, "3D display dependent quality evaluation and rate allocation using scalable video coding," in *16th IEEE International Conference on Image Processing (ICIP)*, November 2009.

[44] M. Reiss and G. Reiss, "Dominance: Some family data," in *Laterality: Asymmetries of Body, Brain and Cognition, vol. 2, pp. 7-16*, 1997.

[45] W. J. Tam, L. B. Stelmach, and S. Subramaniam, "Stereoscopic video: asymmetrical coding with temporal interleaving," in *Proc. SPIE 4297, Stereoscopic Displays and Virtual Reality Systems VIII, pp. 299-306, San Jose, USA*, January 2001.

[46] S. Liu, F. Liu, J. Fan, and H. Xia, "Asymmetric stereoscopic video encoding algorithm based on subjective visual characteristic," in *International Conference on Wireless Communications Signal Processing, pp. 1-5, Nanjing, China*, November 2009.

[47] M. Azimi, S. Valizadeh, X. Li, L. Coria, and P. Nasiopoulos, "Subjective study on asymmetric stereoscopic video with low-pass filtered slices," in *International Conference on Computing, Networking and Communica-tions, pp. 719 -723, Maui, Hawaii, USA*, January 2012.

[48] L. Karlsson and M. Sjostrom, "Region-of-interest 3D video coding based on depth images," in *3DTV Conference: The True Vision - Capture, Transmission and Display of 3D Video, pp. 141-144, Istanbul, Turkey*, May 2008.

[49] *Recommendation P.910, Subjective video quality assessment methods for multimedia applications*, ITU-T Std., 2008.

[50] *Recommendation BT.1438, Subjective assessment of stereoscopic television pictures*, ITU-R Std., 2000.

[51] Q. Huynh-Thua, M. Brothertonc, D. Hands, K. Brunnstrom, and M. Ghanbari, "Examination of the SAMVIQ methodology for the subjective assessment of multimedia quality." [Online]. Available: https://www.acreo.se/sites/default/files/public/acreo.se/upload/publications/10.pdf

[52] *Recommendation BT.1788, Methodology for the subjective assessment of video quality in multimedia applications*, ITU-R Std.

[53] [Online]. Available: http://www.ntt.co.jp/qos/qoe/eng/technology/visual/01_5_7.html

[54] T. Kawano, K. Yamagishi, and T. Hayashi, "Performance comparison of subjective assessment methods for 3D video quality," in *4th International Workshop on Quality of Multimedia Experience (QoMEX), pp.218-223, Yarra Valley, Australia*, July 2012.

[55] Jm - h.264/avc reference software. [Online]. Available: http://iphome.hhi.de/suehring/tml/

[56] http://www.ffmpeg.org/.

[57] W. R. Miles, "Ocular dominance-method and results," *Psychological Bulletin*, vol. 25, pp. 155–156, 1928.

Appendix A

Test sequences

This appendix shows the original test stereoscopic sequences. The sequences Balloons, Champagne Tower, Kendo and Pantonime were obtained under the permission of Tanimoto Lab at Nagoya University (available at `http://www.tanimoto.nuee.nagoya-u.ac.jp/`).

(a) View ID 3 (b) View ID 2

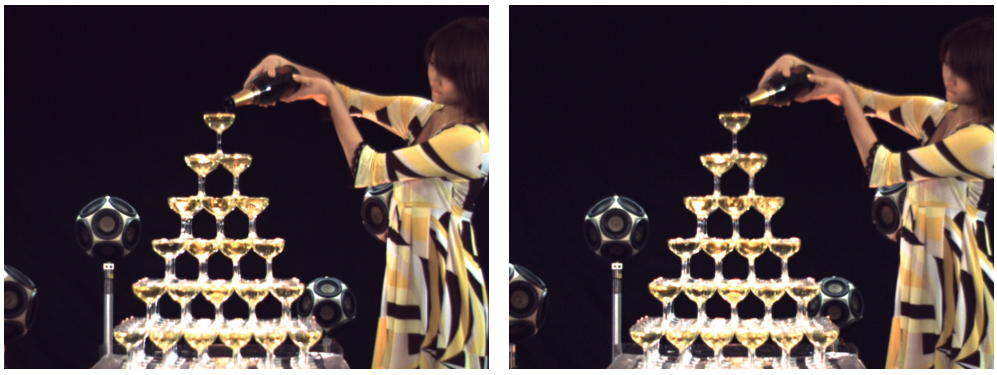
Figure A.1: Stereo sequence Balloons, 1024×768.

(a) View ID 3 (b) View ID 2

Figure A.2: Stereo sequence Kendo, 1024 × 768.

(a) View ID 40 (b) View ID 39

Figure A.3: Stereo sequence Champagne Tower, 1280 × 960.



(a) View ID 40 (b) View ID 39

Figure A.4: Cropped version of the stereo sequence Champagne Tower, 1024 × 768.

(a) View ID 2 (b) View ID 3

Figure A.5: Cropped version of the stereo sequence Pantonime, 1024 × 768.

Figure A.6: Stereo sequence Bike, 1024 × 576.

Figure A.7: Stereo sequence BMX, 1024 × 576.

Figure A.8: Stereo sequence cafe, 1024 × 576.

Figure A.9: Stereo sequence car, 1024×576.

Figure A.10: Stereo sequence notebook, 1024×576.

Appendix B

Example of a configuration file for the modified encoder

The following configuration example was used in order to asymmetrically encode a stereoscopic video (Balloons) using regions of perceptual relevance and using the modified H.264 reference software JM 18.0 .

```
# New Input File Format is as follows
# <ParameterName> = <ParameterValue> # Comment
#
# See configfile.h for a list of supported ParameterNames
#
# For bug reporting and known issues see:
# https://ipbt.hhi.fraunhofer.de


###############################################################################
# Files
###############################################################################
InputFile             = "Balloons-40dB.yuv"       # Input sequence
InputHeaderLength     = 0      # If the inputfile has a header, state it's length in byte here
StartFrame            = 0      # Start frame for encoding. (0-N)
FramesToBeEncoded     = 1      # Number of frames to be coded
FrameRate             = 25.0   # Frame Rate per second (0.1-100.0)
Enable32Pulldown      = 0      # Enforce 3:2 pulldown methods
                               # 0 = disabled
                               # 1 = A, B, Bt|Cb, Ct|Db, D
                               # 2 = A, B, C, Ct|Db, D
SourceWidth           = 1024    # Source frame width
SourceHeight          = 768    # Source frame height
SourceResize          = 0      # Resize source size for output
OutputWidth           = 176    # Output frame width
OutputHeight          = 144    # Output frame height
ProcessInput          = 0      # Filter Input Sequence
Interleaved           = 0      # 0: Planar input, 1: Packed input
StandardRange         = 0      # 0: Standard range 1: Full range (RGB input)
VideoCode             = 1      # Video codes for RGB ==> YUV conversions
                               # 0 = NULL,
                               # 1 = ITU_REC709,
                               # 2 = CCIR_601,
                               # 3 = FCC,
                               # 4 = ITU_REC624BG,
                               # 5 = SMPTE_170M,
                               # 6 = SMPTE_240M,
                               # 7 = SMPTE_260M,
                               # 8 = ITU_REC709_EXACT

TraceFile             = "encoding_trace.txt"      # Trace file
ReconFile             = "Balloons-mask5.yuv"       # Reconstruction YUV file
OutputFile            = "Balloons-mask5.264"           # Bitstream
```

```
StatsFile            = "encoding_stats.dat"         # Coding statistics file

NumberOfViews        = 2                    # Number of views to encode (1=1 view, 2=2 views)
View1ConfigFile      = "encoder_view1_Balloons_40dB.cfg"   # Config file name for second view
###############################################################################
# Encoder Control
###############################################################################
Grayscale            = 0   # Encode in grayscale (Currently only works for 8 bit YUV 420 input)
ProfileIDC           = 128 # Profile IDC (66=baseline, 77=main, 88=extended; FREXT Profiles: 100=High, 110=High 10, 122=High 4:2:2, 244=High 4:4:4, 44=CAVLC 4:4:4 Intra,
IntraProfile         = 0   # Activate Intra Profile for FREXt (0: false, 1: true)
                           # (e.g. ProfileIDC=110, IntraProfile=1  =>  High 10 Intra Profile)
LevelIDC             = 40  # Level IDC   (e.g. 20 = level 2.0)

IntraPeriod          = 10   # Period of I-pictures   (0=only first)
IDRPeriod            = 10   # Period of IDR pictures (0=only first)
AdaptiveIntraPeriod  = 0   # Adaptive intra period
AdaptiveIDRPeriod    = 0   # Adaptive IDR period
IntraDelay           = 0   # Intra (IDR) picture delay (i.e. coding structure of PPIPPP... )
EnableIDRGOP         = 0   # Support for IDR closed GOPs (0: disabled, 1: enabled)
EnableOpenGOP        = 0   # Support for open GOPs (0: disabled, 1: enabled)
QPISlice             = 34  # Quant. param for I Slices (0-51)
QPPSlice             = 34  # Quant. param for P Slices (0-51)
FrameSkip            = 0   # Number of frames to be skipped in input (e.g 2 will code every third frame).
                           # Note that this now excludes intermediate (i.e. B) coded pictures
ChromaQPOffset       = 0   # Chroma QP offset (-51..51)

DisableSubpelME      = 0   # Disable Subpixel Motion Estimation (0=off/default, 1=on)
SearchRange          = 32  # Max search range

MEDistortionFPel     = 0   # Select error metric for Full-Pel ME    (0: SAD, 1: SSE, 2: Hadamard SAD)
MEDistortionHPel     = 2   # Select error metric for Half-Pel ME    (0: SAD, 1: SSE, 2: Hadamard SAD)
MEDistortionQPel     = 2   # Select error metric for Quarter-Pel ME (0: SAD, 1: SSE, 2: Hadamard SAD)
MDDistortion         = 2   # Select error metric for Mode Decision  (0: SAD, 1: SSE, 2: Hadamard SAD)
SkipDeBlockNonRef    = 0   # Skip Deblocking (regardless of DFParametersFlag) for non-reference frames (0: off, 1: on)
ChromaMCBuffer       = 1   # Calculate Color component interpolated values in advance and store them.
                           # Provides a trade-off between memory and computational complexity
                           # (0: disabled/default, 1: enabled)
ChromaMEEnable       = 0   # Take into account Color component information during ME
                           # (0: only first component/default,
                           #  1: All Color components - Integer refinement only
                           #  2: All Color components - All refinements)
ChromaMEWeight       = 1   # Weighting for chroma components. This parameter should have a relationship with color format.

NumberReferenceFrames = 3   # Number of previous frames used for inter motion search (0-16)

PList0References     = 0   # P slice List 0 reference override (0 disable, N <= NumberReferenceFrames)
Log2MaxFNumMinus4    = 0   # Sets log2_max_frame_num_minus4 (-1 : based on FramesToBeEncoded/Auto, >=0 : Log2MaxFNumMinus4)
Log2MaxPOCLsbMinus4  = -1  # Sets log2_max_pic_order_cnt_lsb_minus4 (-1 : Auto, >=0 : Log2MaxPOCLsbMinus4)

GenerateMultiplePPS  = 1   # Transmit multiple parameter sets. Currently parameters basically enable all WP modes (0: diabled, 1: enabled)
SendAUD              = 0   # Send Access Delimiter Unit NALU (for every access unit)
ResendSPS            = 2   # Resend SPS  (0: disabled, 1: all Intra pictures, 2: only for IDR, 3: for IDR and OpenGOP I)
ResendPPS            = 0   # Resend PPS (with pic_parameter_set_id 0) for every coded Frame/Field pair (0: disabled, 1: enabled)

MbLineIntraUpdate    = 0   # Error robustness(extra intra macro block updates)(0=off, N: One GOB every N frames are intra coded)
RandomIntraMBRefresh = 0   # Forced intra MBs per picture

###############################################################################
# PSlice Mode types
###############################################################################
PSliceSkip           = 1   # P-Slice Skip mode consideration  (0=disable, 1=enable)
PSliceSearch16x16    = 1   # P-Slice Inter block search 16x16 (0=disable, 1=enable)
PSliceSearch16x8     = 1   # P-Slice Inter block search 16x8  (0=disable, 1=enable)
PSliceSearch8x16     = 1   # P-Slice Inter block search  8x16 (0=disable, 1=enable)
PSliceSearch8x8      = 1   # P-Slice Inter block search  8x8  (0=disable, 1=enable)
PSliceSearch8x4      = 1   # P-Slice Inter block search  8x4  (0=disable, 1=enable)
PSliceSearch4x8      = 1   # P-Slice Inter block search  4x8  (0=disable, 1=enable)
PSliceSearch4x4      = 1   # P-Slice Inter block search  4x4  (0=disable, 1=enable)

###############################################################################
# BSlice Mode types
###############################################################################

BSliceDirect         = 1   # B-Slice Skip mode consideration  (0=disable, 1=enable)
BSliceSearch16x16    = 1   # B-Slice Inter block search 16x16 (0=disable, 1=enable)
BSliceSearch16x8     = 1   # B-Slice Inter block search 16x8  (0=disable, 1=enable)
BSliceSearch8x16     = 1   # B-Slice Inter block search  8x16 (0=disable, 1=enable)
BSliceSearch8x8      = 1   # B-Slice Inter block search  8x8  (0=disable, 1=enable)
```

```
BSliceSearch8x4       = 1   # B-Slice Inter block search  8x4  (0=disable, 1=enable)
BSliceSearch4x8       = 1   # B-Slice Inter block search  4x8  (0=disable, 1=enable)
BSliceSearch4x4       = 1   # B-Slice Inter block search  4x4  (0=disable, 1=enable)

BiPredSearch16x16     = 1   # B-Slice Bi-prediction block search 16x16 (0=disable, 1=enable)
BiPredSearch16x8      = 1   # B-Slice Bi-prediction block search 16x8  (0=disable, 1=enable)
BiPredSearch8x16      = 1   # B-Slice Bi-prediction block search 8x16  (0=disable, 1=enable)
BiPredSearch8x8       = 0   # B-Slice Bi-prediction block search 8x8   (0=disable, 1=enable)

DisableIntra4x4        = 0  # Disable Intra 4x4 modes
DisableIntra16x16      = 0  # Disable Intra 16x16 modes
DisableIntraInInter    = 0  # Disable Intra modes for inter slices
IntraDisableInterOnly  = 0  # Apply Disabling Intra conditions only to Inter Slices (0:disable/default,1: enable)
Intra4x4ParDisable     = 0  # Disable Vertical & Horizontal 4x4
Intra4x4DiagDisable    = 0  # Disable Diagonal 45degree 4x4
Intra4x4DirDisable     = 0  # Disable Other Diagonal 4x4
Intra16x16ParDisable   = 0  # Disable Vertical & Horizontal 16x16
Intra16x16PlaneDisable = 0  # Disable Planar 16x16
ChromaIntraDisable     = 0  # Disable Intra Chroma modes other than DC
EnableIPCM             = 1  # Enable IPCM macroblock mode

DisposableP            = 0  # Enable Disposable P slices in the primary layer (0: disable/default, 1: enable)
DispPQPOffset          = 0  # Quantizer offset for disposable P slices (0: default)

PreferDispOrder        = 1  # Prefer display order when building the prediction structure as opposed to coding order (affects intra and IDR periodic insertion, a
PreferPowerOfTwo       = 0  # Prefer prediction structures that have lengths expressed as powers of two
FrmStructBufferLength  = 16 # Length of the frame structure unit buffer; it can be overriden for certain cases

ChangeQPFrame          = 0  # Frame in display order from which to apply the Change QP offsets
ChangeQPI              = 0  # Change QP offset value for I_SLICE
ChangeQPP              = 0  # Change QP offset value for P_SLICE
ChangeQPB              = 0  # Change QP offset value for B_SLICE
ChangeQPSI             = 0  # Change QP offset value for SI_SLICE
ChangeQPSP             = 0  # Change QP offset value for SP_SLICE

###############################################################################################
# B Slices
###############################################################################################

NumberBFrames          = 7  # Number of B coded frames inserted (0=not used)
PReplaceBSlice         = 0  # Replace B-coded slices with P-coded slices when NumberBFrames>0
QPBSlice               = 34 # Quant. param for B slices (0-51)
BRefPicQPOffset        = -1 # Quantization offset for reference B coded pictures (-51..51)
DirectModeType         = 1  # Direct Mode Type (0:Temporal 1:Spatial)
DirectInferenceFlag    = 1  # Direct Inference Flag (0: Disable 1: Enable)
BList0References       = 0  # B slice List 0 reference override (0 disable, N <= NumberReferenceFrames)
BList1References       = 1  # B slice List 1 reference override (0 disable, N <= NumberReferenceFrames)
                            # 1 List1 reference is usually recommended for normal GOP Structures.
                            # A larger value is usually more appropriate if a more flexible
                            # structure is used (i.e. using HierarchicalCoding)

BReferencePictures    =  0  # Referenced B coded pictures (0=off, 1=B references for secondary layer, 2=B references for primary layer)

HierarchicalCoding      =  3  # B hierarchical coding (0= off, 1= 2 layers, 2= 2 full hierarchy, 3 = explicit)
HierarchyLevelQPEnable  =  1  # Adjust QP based on hierarchy level (in increments of 1). Overrides BRefPicQPOffset behavior.(0=off, 1=on)
ExplicitHierarchyFormat = "b3r0b1r1b0e2b2e2b5r1b4e2b6e2" # Explicit Enhancement GOP. Format is {FrameDisplay_orderReferenceQP}.
                                                        # Valid values for reference type is r:reference, e:non reference.
ExplicitSeqCoding     =  0    # Enable support for explicit sequence coding
ExplicitSeqFile       =  "explicit_seq.cfg"
LowDelay              =  0    # Apply HierarchicalCoding without delay (i.e., encode in the captured/display order)
ReferenceReorder      =  0    # Reorder References according to Poc distance for HierarchicalCoding (0=off, 1=enable, 2=use when LowDelay is set)
PocMemoryManagement   =  0    # Memory management based on Poc Distances for HierarchicalCoding (0=off, 1=on, 2=use when LowDelay is set)
SetFirstAsLongTerm    =  0    # Set first frame as long term

BiPredMotionEstimation = 1   # Enable Bipredictive based Motion Estimation (0:disabled, 1:enabled)
BiPredMERefinements    = 3   # Bipredictive ME extra refinements (0: single, N: N extra refinements (1 default)
BiPredMESearchRange    = 16  # Bipredictive ME Search range (8 default). Note that range is halved for every extra refinement.
BiPredMESubPel         = 2   # Bipredictive ME Subpixel Consideration (0: disabled, 1: single level, 2: dual level)


###############################################################################################
# SP Frames
###############################################################################################

SPPicturePeriodicity  = 0                # SP-Picture Periodicity (0=not used)
SPSwitchPeriod        = 0                # Switch period (in terms of switching SP/SI frames) between bitstream 1 and bitstream 2
QPSPSlice             = 36               # Quant. param of SP-Slices for Prediction Error (0-51)
QPSISlice             = 36               # Quant. param of SI-Slices for Prediction Error (0-51)
```

```
QPSP2Slice           = 35                # Quant. param of SP/SI-Slices for Predicted Blocks (0-51)
SI_FRAMES            = 0                 # SI frame encoding flag (0=not used, 1=used)
SP_output            = 0                 # Controls whether coefficients will be output to encode switching SP frames (0=no, 1=yes)
SP_output_name       = "low_quality.dat"  # Filename for SP output coefficients
SP2_FRAMES           = 0                 # switching SP frame encoding flag (0=not used, 1=used)
SP2_input_name1      = "high_quality.dat" # Filename for the first swithed bitstream coefficients
SP2_input_name2      = "low_quality.dat"  # Filename for the second switched bitstream coefficients

###############################################################################################
# Output Control, NALs
###############################################################################################

SymbolMode            =  1  # Symbol mode (Entropy coding method: 0=UVLC, 1=CABAC)
OutFileMode           =  0  # Output file mode, 0:Annex B, 1:RTP
PartitionMode         =  0  # Partition Mode, 0: no DP, 1: 3 Partitions per Slice

###############################################################################################
# CABAC context initialization
###############################################################################################

ContextInitMethod       =  1     # Context init (0: fixed, 1: adaptive)
FixedModelNumber        =  0     # model number for fixed decision for inter slices ( 0, 1, or 2 )

###############################################################################################
# Interlace Handling
###############################################################################################

PicInterlace            =  0     # Picture AFF    (0: frame coding, 1: field coding, 2:adaptive frame/field coding)
MbInterlace             =  0     # Macroblock AFF (0: frame coding, 1: field coding, 2:adaptive frame/field coding, 3: frame MB-only AFF)
IntraBottom             =  0     # Force Intra Bottom at GOP Period

###############################################################################################
# Weighted Prediction
###############################################################################################

WeightedPrediction      =  0     # P picture Weighted Prediction (0=off, 1=explicit mode)
WeightedBiprediction    =  0     # B picture Weighted Prediciton (0=off, 1=explicit mode,  2=implicit mode)
ChromaWeightSupport     =  1     # Enable consideration of weights for Chroma components
UseWeightedReferenceME  =  1     # Use weighted reference for ME (0=off, 1=on)
WPMethod                =  1     # WP method (0: DC based, 1: LMS based)
WPIterMC                =  0     # Iterative Motion compensated based weighted prediction method
EnhancedBWeightSupport  =  0     # Enhanced B Weight support (needs revisit if we wish to merge with WPMethod)
WPMCPrecision           =  0     # Improved Motion Compensation Precision using WP based methods.
                                 # Clones WP references with slightly modified rounding offsets (Requires RDPictureDecision and GenerateMultiplePPS) :
                                 # 0: disabled (default)
                                 # 1: Up to one additional coding pass. Ref0 is 0, ref1 is 0 with a -1 offset
                                 # 2: Up to two additional coding passes. (1) Ref0 is 0, ref1 is 0 with a -1 offset, (1) Ref0 is 0 with a -1 offset, ref1 is 0
WPMCPrecFullRef         =  0     # Increases the number of references in the reference picture lists to account
                                 # for the lost reference slot when reordering is used during a coding pass in WPMCPrecision for reference replication.
                                 # The number of references in non-reordered passes stays unchanged
WPMCPrecBSlice          =  1     # 2: Apply rounding on every B slice. This efectively disables the evaluation of alternative QPs during RDPictureDecision.
                                 # 1: Disable rounding for non-reference B slices. Non-reference B slices are evaluated for alternative QPs during RDPictureDecision.
                                 # 0: Disable rounding for B slices.

###############################################################################################
# Picture based Multi-pass encoding
###############################################################################################

RDPictureDecision       =  0     # Perform multiple pass coding and make RD optimal decision among them
RDPSliceBTest           =  0     # Perform Slice level RD decision between P and B slices.
RDPictureMaxPassISlice  =  1     # Max number of coding passes for I slices, valid values [1,3], default is 1
RDPictureMaxPassPSlice  =  2     # Max number of coding passes for P slices, valid values [1,6], default is 2
RDPictureMaxPassBSlice  =  3     # Max number of coding passes for B slices, valid values [1,6], default is 3
RDPictureFrameQPPSlice  =  0     # Perform additional frame level QP check (QP+/-1) for P slices, 0: disabled (default), 1: enabled
RDPictureFrameQPBSlice  =  0     # Perform additional frame level QP check (QP+/-1) for B slices, 0: disabled, 1: enabled (default)
RDPictureDeblocking     =  0     # Perform another coding pass to check non-deblocked picture, 0: disabled (default), 1: enabled
RDPictureDirectMode     =  0     # Perform another coding pass to check the alternative direct mode for B slices, , 0: disabled (default), 1: enabled

###############################################################################################
# Deblocking filter parameters
###############################################################################################

DFParametersFlag        = 0      # Configure deblocking filter (0=parameters below ignored, 1=parameters sent)
                                 # Note that for pictures with multiple slice types,
                                 # only the type of the first slice will be considered.
DFDisableRefISlice      = 0      # Disable deblocking filter in reference I coded pictures (0=Filter, 1=No Filter).
DFAlphaRefISlice        = 0      # Reference I coded pictures Alpha offset div. 2, {-6, -5, ... 0, +1, .. +6}
DFBetaRefISlice         = 0      # Reference I coded pictures Beta offset div. 2, {-6, -5, ... 0, +1, .. +6}
```

```
DFDisableNRefISlice      = 0      # Disable deblocking filter in non reference I coded pictures (0=Filter, 1=No Filter).
DFAlphaNRefISlice        = 0      # Non Reference I coded pictures Alpha offset div. 2, {-6, -5, ... 0, +1, .. +6}
DFBetaNRefISlice         = 0      # Non Reference I coded pictures Beta offset div. 2, {-6, -5, ... 0, +1, .. +6}
DFDisableRefPSlice       = 0      # Disable deblocking filter in reference P coded pictures (0=Filter, 1=No Filter).
DFAlphaRefPSlice         = 0      # Reference P coded pictures Alpha offset div. 2, {-6, -5, ... 0, +1, .. +6}
DFBetaRefPSlice          = 0      # Reference P coded pictures Beta offset div. 2, {-6, -5, ... 0, +1, .. +6}
DFDisableNRefPSlice      = 0      # Disable deblocking filter in non reference P coded pictures (0=Filter, 1=No Filter).
DFAlphaNRefPSlice        = 0      # Non Reference P coded pictures Alpha offset div. 2, {-6, -5, ... 0, +1, .. +6}
DFBetaNRefPSlice         = 0      # Non Reference P coded pictures Beta offset div. 2, {-6, -5, ... 0, +1, .. +6}
DFDisableRefBSlice       = 0      # Disable deblocking filter in reference B coded pictures (0=Filter, 1=No Filter).
DFAlphaRefBSlice         = 0      # Reference B coded pictures Alpha offset div. 2, {-6, -5, ... 0, +1, .. +6}
DFBetaRefBSlice          = 0      # Reference B coded pictures Beta offset div. 2, {-6, -5, ... 0, +1, .. +6}
DFDisableNRefBSlice      = 0      # Disable deblocking filter in non reference B coded pictures (0=Filter, 1=No Filter).
DFAlphaNRefBSlice        = 0      # Non Reference B coded pictures Alpha offset div. 2, {-6, -5, ... 0, +1, .. +6}
DFBetaNRefBSlice         = 0      # Non Reference B coded pictures Beta offset div. 2, {-6, -5, ... 0, +1, .. +6}

###############################################################################################
# Error Resilience / Slices
###############################################################################################

SliceMode             =  0   # Slice mode (0=off 1=fixed #mb in slice 2=fixed #bytes in slice 3=use callback)
SliceArgument         = 50   # Slice argument (Arguments to modes 1 and 2 above)

num_slice_groups_minus1 = 0  # Number of Slice Groups Minus 1, 0 == no FMO, 1 == two slice groups, etc.
slice_group_map_type    = 0  # 0:  Interleave, 1: Dispersed,    2: Foreground with left-over,
                             # 3:  Box-out,    4: Raster Scan   5: Wipe
                             # 6:  Explicit, slice_group_id read from SliceGroupConfigFileName
slice_group_change_direction_flag = 0    # 0: box-out clockwise, raster scan or wipe right,
                                         # 1: box-out counter clockwise, reverse raster scan or wipe left
slice_group_change_rate_minus1    = 85   #
SliceGroupConfigFileName          = "sg0conf.cfg"   # Used for slice_group_map_type 0, 2, 6

UseRedundantPicture   = 0    # 0: not used, 1: enabled
NumRedundantHierarchy = 1    # 0-4
PrimaryGOPLength      = 10   # GOP length for redundant allocation (1-16)
                             # NumberReferenceFrames must be no less than PrimaryGOPLength when redundant slice enabled
NumRefPrimary         = 1    # Actually used number of references for primary slices (1-16)

###############################################################################################
# Search Range Restriction / RD Optimization
###############################################################################################

RestrictSearchRange    =  2  # restriction for (0: blocks and ref, 1: ref, 2: no restrictions)
RDOptimization         =  1  # rd-optimized mode decision
                             # 0: RD-off (Low complexity mode)
                             # 1: RD-on (High complexity mode)
                             # 2: RD-on (Fast high complexity mode - not work in FREX Profiles)
                             # 3: with losses
I16RDOpt               =  0  # perform rd-optimized mode decision for Intra 16x16 MB
                             # 0: SAD-based mode decision for Intra 16x16 MB
                             # 1: RD-based mode decision for Intra 16x16 MB
SubMBCodingState       =  1  # submacroblock coding state
                             # 0: lowest complexity, do not store or reset coding state during sub-MB mode decision
                             # 1: medium complexity, reset to master coding state (for current mode) during sub-MB mode decision
                             # 2: highest complexity, store and reset coding state during sub-MB mode decision
DistortionSSIM         =  0  # Compute SSIM distortion. (0: disabled/default, 1: enabled)
DistortionMS_SSIM      =  0  # Compute Multiscale SSIM distortion. (0: disabled/default, 1: enabled)
SSIMOverlapSize        =  8  # Overlap size to calculate SSIM distortion (1: pixel by pixel, 8: no overlap)
DistortionYUVtoRGB     =  0  # Calculate distortion in RGB domain after conversion from YCbCr (0:off, 1:on)
CtxAdptLagrangeMult    =  0  # Context Adaptive Lagrange Multiplier
                             # 0: disabled (default)
                             # 1: enabled (works best when RDOptimization=0)
FastCrIntraDecision    =  0  # Fast Chroma intra mode decision (0:off, 1:on)
DisableThresholding    =  0  # Disable Thresholding of Transform Coefficients (0:off, 1:on)
DisableBSkipRDO        =  0  # Disable B Skip Mode consideration from RDO Mode decision (0:off, 1:on)
BiasSkipRDO            =  0  # Negative Bias for Skip/DirectSkip modes (0: off, 1: on)
ForceTrueRateRDO       =  0  # Force true rate (even zero values) during RDO process
SkipIntraInInterSlices =  0  # Skips Intra mode checking in inter slices if certain mode decisions are satisfied (0: off, 1: on)
WeightY                =  1  # Luma weight for RDO
WeightCb               =  1  # Cb weight for RDO
WeightCr               =  1  # Cr weight for RDO

###############################################################################################
# Explicit Lambda Usage
###############################################################################################
UseExplicitLambdaParams  =  0    # Use explicit lambda scaling parameters (0:disabled, 1:enable lambda weight, 2: use explicit lambda value)
UpdateLambdaChromaME     =  0    # Update lambda given Chroma ME consideration
FixedLambdaISlice        =  0.1  # Fixed Lambda value for I slices
```

```
FixedLambdaPSlice        =  0.1  # Fixed Lambda value for P slices
FixedLambdaBSlice        =  0.1  # Fixed Lambda value for B slices
FixedLambdaRefBSlice     =  0.1  # Fixed Lambda value for Referenced B slices
FixedLambdaSPSlice       =  0.1  # Fixed Lambda value for SP slices
FixedLambdaSISlice       =  0.1  # Fixed Lambda value for SI slices

LambdaWeightISlice       =  0.65 # scaling param for I slices. This will be used as a multiplier i.e. lambda=LambdaWeightISlice * 2^((QP-12)/3)
LambdaWeightPSlice       =  0.68 # scaling param for P slices. This will be used as a multiplier i.e. lambda=LambdaWeightPSlice * 2^((QP-12)/3)
LambdaWeightBSlice       =  0.68 # scaling param for B slices. This will be used as a multiplier i.e. lambda=LambdaWeightBSlice * 2^((QP-12)/3)
LambdaWeightRefBSlice    =  0.68 # scaling param for Referenced B slices. This will be used as a multiplier i.e. lambda=LambdaWeightRefBSlice * 2^((QP-12)/3)
LambdaWeightSPSlice      =  0.68 # scaling param for SP slices. This will be used as a multiplier i.e. lambda=LambdaWeightSPSlice * 2^((QP-12)/3)
LambdaWeightSISlice      =  0.65 # scaling param for SI slices. This will be used as a multiplier i.e. lambda=LambdaWeightSISlice * 2^((QP-12)/3)

LossRateA                =  5  # expected packet loss rate of the channel for the first partition, only valid if RDOptimization = 3
LossRateB                =  0  # expected packet loss rate of the channel for the second partition, only valid if RDOptimization = 3
LossRateC                =  0  # expected packet loss rate of the channel for the third partition, only valid if RDOptimization = 3
FirstFrameCorrect        =  0  # If 1, the first frame is encoded under the assumption that it is always correctly received.
NumberOfDecoders         = 30  # Numbers of decoders used to simulate the channel, only valid if RDOptimization = 3
RestrictRefFrames        =  0  # Doesnt allow reference to areas that have been intra updated in a later frame.

###############################################################################
# Additional Stuff
###############################################################################

UseConstrainedIntraPred  =  0  # If 1, Inter pixels are not used for Intra macroblock prediction.

NumberofLeakyBuckets     =  8                     # Number of Leaky Bucket values
LeakyBucketRateFile      =  "leakybucketrate.cfg"  # File from which encoder derives rate values
LeakyBucketParamFile     =  "leakybucketparam.cfg" # File where encoder stores leakybucketparams

NumFramesInELayerSubSeq  = 0  # number of frames in the Enhanced Scalability Layer(0: no Enhanced Layer)

SparePictureOption        =  0   # (0: no spare picture info, 1: spare picture available)
SparePictureDetectionThr  =  6   # Threshold for spare reference pictures detection
SparePicturePercentageThr = 92   # Threshold for the spare macroblock percentage

PicOrderCntType           = 0    # (0: POC mode 0, 1: POC mode 1, 2: POC mode 2)

###############################################################################
#Rate control
###############################################################################

RateControlEnable       = 0     # 0 Disable, 1 Enable
Bitrate                 = 45020 # Bitrate(bps)
InitialQP               = 0     # Initial Quantization Parameter for the first I frame
                                # InitialQp depends on two values: Bits Per Picture,
                                # and the GOP length
BasicUnit               = 0     # Number of MBs in the basic unit
                                # should be a fraction of the total number
                                # of MBs in a frame ("0" sets a BU equal to a frame)
ChannelType             = 0     # type of channel( 1=time varying channel; 0=Constant channel)
RCUpdateMode            = 0     # Rate Control type. Modes supported :
                                # 0 = original JM rate control,
                                # 1 = rate control that is applied to all frames regardless of the slice type,
                                # 2 = original plus intelligent QP selection for I and B slices (including Hierarchical),
                                # 3 = original + hybrid quadratic rate control for I and B slice using bit rate statistics
                                #
RCISliceBitRatio        = 1.0   # target ratio of bits for I-coded pictures compared to P-coded Pictures (for RCUpdateMode=3)
RCBSliceBitRatio0       = 0.5   # target ratio of bits for B-coded pictures compared to P-coded Pictures - temporal level 0 (for RCUpdateMode=3)
RCBSliceBitRatio1       = 0.25  # target ratio of bits for B-coded pictures compared to P-coded Pictures - temporal level 1 (for RCUpdateMode=3)
RCBSliceBitRatio2       = 0.25  # target ratio of bits for B-coded pictures compared to P-coded Pictures - temporal level 2 (for RCUpdateMode=3)
RCBSliceBitRatio3       = 0.25  # target ratio of bits for B-coded pictures compared to P-coded Pictures - temporal level 3 (for RCUpdateMode=3)
RCBSliceBitRatio4       = 0.25  # target ratio of bits for B-coded pictures compared to P-coded Pictures - temporal level 4 (for RCUpdateMode=3)
RCBoverPRatio           = 0.45  # ratio of bit rate usage of a B-coded picture over a P-coded picture for the SAME QP (for RCUpdateMode=3)
RCIoverPRatio           = 3.80  # ratio of bit rate usage of an I-coded picture over a P-coded picture for the SAME QP (for RCUpdateMode=3)
RCMinQPPSlice           =  8    # minimum P Slice QP value for rate control
RCMaxQPPSlice           = 42    # maximum P Slice QP value for rate control
RCMinQPBSlice           =  8    # minimum B Slice QP value for rate control
RCMaxQPBSlice           = 42    # maximum B Slice QP value for rate control
RCMinQPISlice           =  8    # minimum I Slice QP value for rate control
RCMaxQPISlice           = 42    # maximum I Slice QP value for rate control
RCMinQPSPSlice          =  8    # minimum SP Slice QP value for rate control
RCMaxQPSPSlice          = 40    # maximum SP Slice QP value for rate control
RCMinQPSISlice          =  8    # minimum SI Slice QP value for rate control
RCMaxQPSISlice          = 42    # maximum SI Slice QP value for rate control
RCMaxQPChange           =  4    # maximum QP change for frames of the base layer

###############################################################################
```

```
#Fast Mode Decision
###############################################################################################
EarlySkipEnable         = 0     # Early skip detection (0: Disable 1: Enable)
SelectiveIntraEnable    = 0     # Selective Intra mode decision (0: Disable 1: Enable)

###############################################################################################
#FREXT stuff
###############################################################################################

YUVFormat               = 1     # YUV format (0=4:0:0, 1=4:2:0, 2=4:2:2, 3=4:4:4)
RGBInput                = 0     # 1=RGB input, 0=GBR or YUV input
SeparateColourPlane     = 0     # 4:4:4 coding: 0=Common mode, 1=Independent mode
SourceBitDepthLuma      = 8     # Source Bit Depth for Luma color component (8...14 bits)
SourceBitDepthChroma    = 8     # Source Bit Depth for Chroma color components (8...14 bits)
SourceBitDepthRescale   = 0     # Rescale bit depth of source for output (0: Disable 1: Enable)
OutputBitDepthLuma      = 8     # Output Bit Depth for Luma color component (8...14 bits)
OutputBitDepthChroma    = 8     # Output Bit Depth for Chroma color components (8...14 bits)

CbQPOffset              = 0     # Chroma QP offset for Cb-part (-51..51)
CrQPOffset              = 0     # Chroma QP offset for Cr-part (-51..51)
Transform8x8Mode        = 1     # (0: only 4x4 transform, 1: allow using 8x8 transform additionally, 2: only 8x8 transform)
ReportFrameStats        = 0     # (0:Disable Frame Statistics 1: Enable)
DisplayEncParams        = 0     # (0:Disable Display of Encoder Params 1: Enable)
Verbose                 = 1     # level of display verboseness
                                # 0: short, 1: normal (default), 2: detailed, 3: detailed/nvb
SkipGlobalStats         = 0     # Disable global stat accumulation  (Set to 1 to avoid bipred core dump)

###############################################################################################
#Q-Matrix (FREXT)
###############################################################################################
QmatrixFile             = "q_matrix.cfg"

ScalingMatrixPresentFlag = 0    # Enable Q_Matrix  (0 Not present, 1 Present in SPS, 2 Present in PPS, 3 Present in both SPS & PPS)
ScalingListPresentFlag0  = 3    # Intra4x4_Luma    (0 Not present, 1 Present in SPS, 2 Present in PPS, 3 Present in both SPS & PPS)
ScalingListPresentFlag1  = 3    # Intra4x4_ChromaU (0 Not present, 1 Present in SPS, 2 Present in PPS, 3 Present in both SPS & PPS)
ScalingListPresentFlag2  = 3    # Intra4x4_chromaV (0 Not present, 1 Present in SPS, 2 Present in PPS, 3 Present in both SPS & PPS)
ScalingListPresentFlag3  = 3    # Inter4x4_Luma    (0 Not present, 1 Present in SPS, 2 Present in PPS, 3 Present in both SPS & PPS)
ScalingListPresentFlag4  = 3    # Inter4x4_ChromaU (0 Not present, 1 Present in SPS, 2 Present in PPS, 3 Present in both SPS & PPS)
ScalingListPresentFlag5  = 3    # Inter4x4_ChromaV (0 Not present, 1 Present in SPS, 2 Present in PPS, 3 Present in both SPS & PPS)
ScalingListPresentFlag6  = 3    # Intra8x8_Luma    (0 Not present, 1 Present in SPS, 2 Present in PPS, 3 Present in both SPS & PPS)
ScalingListPresentFlag7  = 3    # Inter8x8_Luma    (0 Not present, 1 Present in SPS, 2 Present in PPS, 3 Present in both SPS & PPS)
ScalingListPresentFlag8  = 3    # Intra8x8_ChromaU for 4:4:4 (0 Not present, 1 Present in SPS, 2 Present in PPS, 3 Present in both SPS & PPS)
ScalingListPresentFlag9  = 3    # Inter8x8_ChromaU for 4:4:4 (0 Not present, 1 Present in SPS, 2 Present in PPS, 3 Present in both SPS & PPS)
ScalingListPresentFlag10 = 3    # Intra8x8_ChromaV for 4:4:4 (0 Not present, 1 Present in SPS, 2 Present in PPS, 3 Present in both SPS & PPS)
ScalingListPresentFlag11 = 3    # Inter8x8_ChromaV for 4:4:4 (0 Not present, 1 Present in SPS, 2 Present in PPS, 3 Present in both SPS & PPS)

###############################################################################################
#Rounding Offset control
###############################################################################################

OffsetMatrixPresentFlag  = 0    # Enable Explicit Offset Quantization Matrices  (0: disable 1: enable)
QOffsetMatrixFile        = "q_offset.cfg" # Explicit Quantization Matrices file

AdaptiveRounding         = 1    # Enable Adaptive Rounding based on JVT-N011 (0: disable, 1: enable)
AdaptRoundingFixed       = 0    # Enable Global Adaptive rounding for all qps (0: disable, 1: enable - default/old)
AdaptRndPeriod           = 1   # Period in terms of MBs for updating rounding offsets.
                                # 0 performs update at the picture level. Default is 16. 1 is as in JVT-N011.
AdaptRndChroma           = 1    # Enables coefficient rounding adaptation for chroma

AdaptRndWFactorIRef      = 8    # Adaptive Rounding Weight for I/SI slices in reference pictures /4096
AdaptRndWFactorPRef      = 8    # Adaptive Rounding Weight for P/SP slices in reference pictures /4096
AdaptRndWFactorBRef      = 8    # Adaptive Rounding Weight for B slices in reference pictures /4096
AdaptRndWFactorINRef     = 8    # Adaptive Rounding Weight for I/SI slices in non reference pictures /4096
AdaptRndWFactorPNRef     = 8    # Adaptive Rounding Weight for P/SP slices in non reference pictures /4096
AdaptRndWFactorBNRef     = 8    # Adaptive Rounding Weight for B slices in non reference pictures /4096

AdaptRndCrWFactorIRef    = 8    # Chroma Adaptive Rounding Weight for I/SI slices in reference pictures /4096
AdaptRndCrWFactorPRef    = 8    # Chroma Adaptive Rounding Weight for P/SP slices in reference pictures /4096
AdaptRndCrWFactorBRef    = 8    # Chroma Adaptive Rounding Weight for B slices in reference pictures /4096
AdaptRndCrWFactorINRef   = 8    # Chroma Adaptive Rounding Weight for I/SI slices in non reference pictures /4096
AdaptRndCrWFactorPNRef   = 8    # Chroma Adaptive Rounding Weight for P/SP slices in non reference pictures /4096
AdaptRndCrWFactorBNRef   = 8    # Chroma Adaptive Rounding Weight for B slices in non reference pictures /4096

###############################################################
# Rate Distortion Optimized Quantization
###############################################################
UseRDOQuant              =  0 # Use Rate Distortion Optimized Quantization (0=disable, 1=enable)
RDOQ_DC                  =  1 # Enable Rate Distortion Optimized Quantization for DC components (0=disable, 1=enable)
```

```
RDOQ_CR                =  1 # Enable Rate Distortion Optimized Quantization for Chroma components (0=disable, 1=enable)
RDOQ_DC_CR             =  1 # Enable Rate Distortion Optimized Quantization for Chroma DC components (0=disable, 1=enable)
RDOQ_QP_Num            =  5 # 1-9: Number of QP tested in RDO_Q (I/P/B slice)
RDOQ_CP_Mode           =  0 # copy Mode from first QP tested
RDOQ_CP_MV             =  0 # copy MV from first QP tested
RDOQ_Fast              =  0 # Fast RDOQ decision method for multiple QPs

###############################################################################################
#Lossless Coding (FREXT)
###############################################################################################

LosslessCoding          = 0    # Enable lossless coding when qpprime_y is zero (0 Disabled, 1 Enabled)

###############################################################################################
#Fast Motion Estimation Control Parameters
###############################################################################################

SearchMode              = 3    # Motion estimation mode
                               # -1 = Full Search
                               #  0 = Fast Full Search (default)
                               #  1 = UMHexagon Search
                               #  2 = Simplified UMHexagon Search
                               #  3 = Enhanced Predictive Zonal Search (EPZS)

UMHexDSR                = 1    # Use Search Range Prediction. Only for UMHexagonS method
                               # (0:disable, 1:enabled/default)
UMHexScale              = 3    # Use Scale_factor for different image sizes. Only for UMHexagonS method
                               # (0:disable, 3:/default)
                               # Increasing value can speed up Motion Search.

EPZSPattern             = 2    # Select EPZS primary refinement pattern.
                               # (0: small diamond, 1: square, 2: extended diamond/default,
                               # 3: large diamond, 4: SBP Large Diamond,
                               # 5: PMVFAST )
EPZSDualRefinement      = 3    # Enables secondary refinement pattern.
                               # (0:disabled, 1: small diamond, 2: square,
                               # 3: extended diamond/default, 4: large diamond,
                               # 5: SBP Large Diamond, 6: PMVFAST )
EPZSFixedPredictors     = 2    # Enables Window based predictors
                               # (0:disabled, 1: P only, 2: P and B/default)
EPZSTemporal            = 1    # Enables temporal predictors
                               # (0: disabled, 1: enabled/default)
EPZSSpatialMem          = 1    # Enables spatial memory predictors
                               # (0: disabled, 1: enabled/default)
EPZSBlockType           = 1    # Enables block type Predictors
                               # (0: disabled, 1: enabled/default)
EPZSMinThresScale       = 0    # Scaler for EPZS minimum threshold (0 default).
                               # Increasing value can speed up encoding.
EPZSMedThresScale       = 1    # Scaler for EPZS median threshold (1 default).
                               # Increasing value can speed up encoding.
EPZSMaxThresScale       = 2    # Scaler for EPZS maximum threshold (1 default).
                               # Increasing value can speed up encoding.
EPZSSubPelME            = 1    # EPZS Subpel ME consideration
EPZSSubPelMEBiPred      = 1    # EPZS Subpel ME consideration for BiPred partitions
EPZSSubPelThresScale    = 2    # EPZS Subpel ME Threshold scaler
EPZSSubPelGrid          = 1    # Perform EPZS using a subpixel grid

###############################################################################################
# SEI Parameters
###############################################################################################

ToneMappingSEIPresentFlag = 0    # Enable Tone mapping SEI  (0 Not present, 1 Present)
ToneMappingFile           = "ToneMapping.cfg"

GenerateSEIMessage        = 0                   # Generate an SEI Text Message
SEIMessageText            = "H.264/AVC Encoder"  # Text SEI Message

UseMVLimits               = 0                   # Use MV Limits
SetMVXLimit               = 512                 # Horizontal MV Limit (in integer units)
SetMVYLimit               = 512                 # Vertical MV Limit (in integer units)

###############################################################################################
# VUI Parameters
###############################################################################################
# the variables below do not affect encoding and decoding
# (many are dummy variables but others can be useful when supported by the decoder)

EnableVUISupport                              = 0      # Enable VUI Parameters
```

```
# display parameters
VUI_aspect_ratio_info_present_flag              = 0
VUI_aspect_ratio_idc                            = 1
VUI_sar_width                                   = 0
VUI_sar_height                                  = 0
VUI_overscan_info_present_flag                  = 0
VUI_overscan_appropriate_flag                   = 0
VUI_video_signal_type_present_flag              = 0
VUI_video_format                                = 5
VUI_video_full_range_flag                       = 0
VUI_colour_description_present_flag             = 0
VUI_colour_primaries                            = 2
VUI_transfer_characteristics                    = 2
VUI_matrix_coefficients                         = 2
VUI_chroma_location_info_present_flag           = 0
VUI_chroma_sample_loc_type_top_field            = 0
VUI_chroma_sample_loc_type_bottom_field         = 0
VUI_timing_info_present_flag                    = 0
VUI_num_units_in_tick                           = 1000
VUI_time_scale                                  = 60000
VUI_fixed_frame_rate_flag                       = 0

# nal hrd parameters
VUI_nal_hrd_parameters_present_flag             = 0
VUI_nal_cpb_cnt_minus1                          = 0
VUI_nal_bit_rate_scale                          = 0
VUI_nal_cpb_size_scale                          = 0
VUI_nal_bit_rate_value_minus1                   = 0
VUI_nal_cpb_size_value_minus1                   = 0
VUI_nal_vbr_cbr_flag                            = 0
VUI_nal_initial_cpb_removal_delay_length_minus1 = 23
VUI_nal_cpb_removal_delay_length_minus1         = 23
VUI_nal_dpb_output_delay_length_minus1          = 23
VUI_nal_time_offset_length                      = 24

# vlc hrd parameters
VUI_vcl_hrd_parameters_present_flag             = 0
VUI_vcl_cpb_cnt_minus1                          = 0
VUI_vcl_bit_rate_scale                          = 0
VUI_vcl_cpb_size_scale                          = 0
VUI_vcl_bit_rate_value_minus1                   = 0
VUI_vcl_cpb_size_value_minus1                   = 0
VUI_vcl_vbr_cbr_flag                            = 0
VUI_vcl_initial_cpb_removal_delay_length_minus1 = 23
VUI_vcl_cpb_removal_delay_length_minus1         = 23
VUI_vcl_dpb_output_delay_length_minus1          = 23
VUI_vcl_time_offset_length                      = 24
VUI_low_delay_hrd_flag                          = 0

# other parameters (i.e. bitsream restrictions)
VUI_pic_struct_present_flag                     = 0
VUI_bitstream_restriction_flag                  = 0
VUI_motion_vectors_over_pic_boundaries_flag     = 1
VUI_max_bytes_per_pic_denom                     = 0
VUI_max_bits_per_mb_denom                       = 0
VUI_log2_max_mv_length_vertical                 = 16
VUI_log2_max_mv_length_horizontal               = 16
VUI_num_reorder_frames                          = 16


# New Input File Format is as follows
# <ParameterName> = <ParameterValue> # Comment
#
# See configfile.h for a list of supported ParameterNames
#
# For bug reporting and known issues see:
# https://ipbt.hhi.de


###############################################################################################
# Second View configuration
###############################################################################################
InputFile            = "Balloons-40dB-view1.yuv"       # Input sequence 2 for the MVC profile
ReconFile             = "Balloons-mask5-view1.yuv"      # Reconstruction YUV file 2 for the MVC profile

###############################################################################################
# Encoder Control
###############################################################################################
```

```
SearchRange      = 32  # Max search range for view 1 (requires SepViewInterSearch)
DisableSubpelME  = 0   # Disable Subpixel Motion Estimation for view 1 (0=off/default, 1=on) (requires SepViewInterSearch)
DisableIntraInInter = 0 # Disable Intra modes for inter slices for view 1 (requires SepViewInterSearch)

PListORefs       = 0   # View 1 P slice List 0 reference override (0 disable, N <= (NumberReferenceFrames + 1))  (requires SepViewInterSearch)

##############################################################################################
# PSlice Mode types
##############################################################################################
PSliceSkip       = 1   # P-Slice Skip mode consideration  (0=disable, 1=enable) (requires SepViewInterSearch)
PSliceSearch16x16 = 1  # P-Slice Inter block search 16x16 (0=disable, 1=enable) (requires SepViewInterSearch)
PSliceSearch16x8 = 1   # P-Slice Inter block search 16x8  (0=disable, 1=enable) (requires SepViewInterSearch)
PSliceSearch8x16 = 1   # P-Slice Inter block search  8x16 (0=disable, 1=enable) (requires SepViewInterSearch)
PSliceSearch8x8  = 1   # P-Slice Inter block search  8x8  (0=disable, 1=enable) (requires SepViewInterSearch)
PSliceSearch8x4  = 1   # P-Slice Inter block search  8x4  (0=disable, 1=enable) (requires SepViewInterSearch)
PSliceSearch4x8  = 1   # P-Slice Inter block search  4x8  (0=disable, 1=enable) (requires SepViewInterSearch)
PSliceSearch4x4  = 1   # P-Slice Inter block search  4x4  (0=disable, 1=enable) (requires SepViewInterSearch)

##############################################################################################
# BSlice Mode types
##############################################################################################

BSliceDirect     = 1   # B-Slice Skip mode consideration  (0=disable, 1=enable) (requires SepViewInterSearch)
BSliceSearch16x16 = 1  # B-Slice Inter block search 16x16 (0=disable, 1=enable) (requires SepViewInterSearch)
BSliceSearch16x8 = 1   # B-Slice Inter block search 16x8  (0=disable, 1=enable) (requires SepViewInterSearch)
BSliceSearch8x16 = 1   # B-Slice Inter block search  8x16 (0=disable, 1=enable) (requires SepViewInterSearch)
BSliceSearch8x8  = 1   # B-Slice Inter block search  8x8  (0=disable, 1=enable) (requires SepViewInterSearch)
BSliceSearch8x4  = 1   # B-Slice Inter block search  8x4  (0=disable, 1=enable) (requires SepViewInterSearch)
BSliceSearch4x8  = 1   # B-Slice Inter block search  4x8  (0=disable, 1=enable) (requires SepViewInterSearch)
BSliceSearch4x4  = 1   # B-Slice Inter block search  4x4  (0=disable, 1=enable) (requires SepViewInterSearch)

MVCEnableInterViewFlag = 1   # enable inter view flag
MVCInterViewReorder    = 0   # Reorder References according to interview pictures (0=off, 1=enable)
QPOffset          = 0   # QP offset during rate control for View 1
SepViewInterSearch     = 0   # If set, allows different InterSearch modes to be set for each view (default is 0 - disabled)
NoResidueRDO      = 1   # Test no residue case for View 1 during RDO (0: disabled, 1: enabled - default)
MVCInterViewForceB     = 0   # Force B coded pictures for Enhancement layer (assuming list order is as desired, this may be able to provide some coding gains)

##############################################################################################
# B Slices
##############################################################################################

BListORefs        = 0  # View 1 B slice List 0 reference override (0 disable, N <= (NumberReferenceFrames + 1)) (requires SepViewInterSearch)
BList1Refs        = 1  # View 1 B slice List 1 reference override (0 disable, N <= NumberReferenceFrames) (requires SepViewInterSearch)

BiPredMESearchRange = 16  # Bipredictive ME Search range (8 default) for view 1. Note that range is halved for every extra refinement. Requires SepViewInterSearch.

##############################################################################################
# Deblocking filter parameters
##############################################################################################

DFDisableRefISlice   = 0      # Disable deblocking filter in reference I coded pictures (0=Filter, 1=No Filter).
DFAlphaRefISlice     = 0      # Reference I coded pictures Alpha offset div. 2, {-6, -5, ... 0, +1, .. +6}
DFBetaRefISlice      = 0      # Reference I coded pictures Beta offset div. 2, {-6, -5, ... 0, +1, .. +6}
DFDisableNRefISlice  = 0      # Disable deblocking filter in non reference I coded pictures (0=Filter, 1=No Filter).
DFAlphaNRefISlice    = 0      # Non Reference I coded pictures Alpha offset div. 2, {-6, -5, ... 0, +1, .. +6}
DFBetaNRefISlice     = 0      # Non Reference I coded pictures Beta offset div. 2, {-6, -5, ... 0, +1, .. +6}
DFDisableRefPSlice   = 0      # Disable deblocking filter in reference P coded pictures (0=Filter, 1=No Filter).
DFAlphaRefPSlice     = 0      # Reference P coded pictures Alpha offset div. 2, {-6, -5, ... 0, +1, .. +6}
DFBetaRefPSlice      = 0      # Reference P coded pictures Beta offset div. 2, {-6, -5, ... 0, +1, .. +6}
DFDisableNRefPSlice  = 0      # Disable deblocking filter in non reference P coded pictures (0=Filter, 1=No Filter).
DFAlphaNRefPSlice    = 0      # Non Reference P coded pictures Alpha offset div. 2, {-6, -5, ... 0, +1, .. +6}
DFBetaNRefPSlice     = 0      # Non Reference P coded pictures Beta offset div. 2, {-6, -5, ... 0, +1, .. +6}
DFDisableRefBSlice   = 0      # Disable deblocking filter in reference B coded pictures (0=Filter, 1=No Filter).
DFAlphaRefBSlice     = 0      # Reference B coded pictures Alpha offset div. 2, {-6, -5, ... 0, +1, .. +6}
DFBetaRefBSlice      = 0      # Reference B coded pictures Beta offset div. 2, {-6, -5, ... 0, +1, .. +6}
DFDisableNRefBSlice  = 0      # Disable deblocking filter in non reference B coded pictures (0=Filter, 1=No Filter).
DFAlphaNRefBSlice    = 0      # Non Reference B coded pictures Alpha offset div. 2, {-6, -5, ... 0, +1, .. +6}
DFBetaNRefBSlice     = 0      # Non Reference B coded pictures Beta offset div. 2, {-6, -5, ... 0, +1, .. +6}

##############################################################################################
#Fast Motion Estimation Control Parameters
##############################################################################################

SearchMode          = 3    # Defined same as above, but for view 1 (requires SepViewInterSearch to be set)

EPZSMinThresScale   = 0    # Enh layer scaler for EPZS minimum threshold (0 default).
                           # Increasing value can speed up encoding. EnableEnhLayerEPZSScalers must be 1.
```

```
EPZSMedThresScale   = 1    # Enh layer scaler for EPZS median threshold (1 default).
                           # Increasing value can speed up encoding. EnableEnhLayerEPZSScalers must be 1.
EPZSMaxThresScale   = 2    # Enh layer scaler for EPZS maximum threshold (1 default).
                           # Increasing value can speed up encoding. EnableEnhLayerEPZSScalers must be 1.
EPZSSubPelThresScale = 1   # Enh layer EPZS Subpel ME Threshold scaler. EnableEnhLayerEPZSScalers must be 1.


###############################################################################################
# L. PINTO ADDED PARAMETERS
###############################################################################################

Transcoding = 0   # Use transcoding info from previous decoder outputed files (0=off, 1=on) - MASTER ENABLE
UseDecoderMVs = 0   # Use motion vectors info from previous decoder outputed files (0=off, 1=on)
TranscodingInfoFile = "infofile_bin.bin" # binary file with transcoding info
TranscodingMVFile = "mvfile_bin.bin" # binary file with motion vectors info
AssymCoding = 1   # Flag for view 1 Assymetric Coding (0=off, 1=Quality/QP Based, 2=Spatial Resolution, 3=Temporal)
nonbaseQPISlice         = 51  # Quant. param for I Slices (0-51)
nonbaseQPPSlice         = 51  # Quant. param for P Slices (0-51)
nonbaseQPBSlice         = 51  # Quant. param for B slices (0-51)
nonbaseQPSPSlice        = 41  # Quant. param of SP-Slices for Prediction Error (0-51)
nonbaseQPSISlice        = 41  # Quant. param of SI-Slices for Prediction Error (0-51)
MBAssymCodingFlag       = 1   # Enable (1) or disable (0) the use of different QP for important (disparity) MBs
MBAssymCodingQP = 38  # Quant. Param for important (disparity) MBs (0-51) for both frames and all slice types
RelevantMBsFile = "mask_Balloons_morph.yuv" # File with the relevant MBs info
```

Appendix C

Example of a configuration file for the modified decoder

```
# This is a file containing input parameters to the JVT H.264/AVC decoder.
# The text line following each parameter is discarded by the decoder.
#
# For bug reporting and known issues see:
# https://ipbt.hhi.fraunhofer.de
#
# New Input File Format is as follows
# <ParameterName> = <ParameterValue> # Comment
#
###############################################################################################
# Files
###############################################################################################
InputFile             = "champtower_vga_20f.264"       # H.264/AVC coded bitstream
OutputFile            = "test_dec.yuv"   # Output file, YUV/RGB
RefFile               = "test_rec.yuv"   # Ref sequence (for SNR)
WriteUV               = 1                # Write 4:2:0 chroma components for monochrome streams
FileFormat            = 0                # NAL mode (0=Annex B, 1: RTP packets)
RefOffset             = 0                # SNR computation offset
POCScale              = 2                # Poc Scale (1 or 2)
###############################################################################################
# HRD parameters
###############################################################################################
#R_decoder             = 500000           # Rate_Decoder
#B_decoder             = 104000           # B_decoder
#F_decoder             = 73000            # F_decoder
#LeakyBucketParamFile  = "leakybucketparam.cfg" # LeakyBucket Params
###############################################################################################
# decoder control parameters
###############################################################################################
DisplayDecParams      = 0                # 1: Display parameters;
ConcealMode           = 0                # Err Concealment(0:Off,1:Frame Copy,2:Motion Copy)
RefPOCGap             = 2                # Reference POC gap (2: IPP (Default), 4: IbP / IpP)
POCGap                = 2                # POC gap (2: IPP /IbP/IpP (Default), 4: IPP with frame skip = 1 etc.)
Silent                = 0                # Silent decode
IntraProfileDeblocking = 1               # Enable Deblocking filter in intra only profiles (0=disable, 1=filter according to SPS parameters)
DecFrmNum             = 0                # Number of frames to be decoded (-n)
###############################################################################################
# MVC decoding parameters
###############################################################################################
DecodeAllLayers       = 1                 # Decode all views (-mpr)
###############################################################################################
# L. Pinto variables
###############################################################################################
OutputEncodingInfo    = 1                 # 0 -> disable, 1 -> enable
OutputInfoFile        = "infofile.txt"    # txt file to output the encoding info (if OutputEncodingInfo = 1)
OutputMVInfo          = 1                 # 0 -> disable, 1 -> enable
OutputMVFile          = "mvfile.txt"      # txt file to output the motion vectors info (if OutputMVInfo = 1)
```

Appendix D

Deliverables - Papers, posters, articles and book chapters

This appendix presents the published papers and presented posters, resulted from the research work done during this dissertation.

Published:

Pinto, L. and Assunção, P., *Asymmetric 3D Video Coding Using Regions of Perceptual Relevance*, International Conference on 3D Imaging (IC3D), Liège, Belgium, 3-5 December 2012.

Pinto, L. and Assunção, P., *Asymmetric 3D Video Coding Using Regions of Perceptual Relevance*, 9^{th} Conference on Telecommunications, Castelo Branco, Portugal, 8-10 May 2013.

Pinto, L. and Assunção, P., *Non-uniform Asymmetric Coding of Stereo Images*, 2^{nd} Romeo Workshop, Istanbul, Turkey, 9 July 2013.

Pedro Assunção, Luís Pinto, Sérgio Faria, chapter: *3D media representation and coding*, in book: *3D Future Internet Media*, T. Dagiuklas, A. Kondoz (eds), Springer (Publication in 2014).

Presented posters:

Pinto, L. and Assunção, P., *Asymmetric 3D Video Coding using Regions of Perceptual Interest*, COST Training School on 3D Media, UX and Computational Architectures, Tampere, Finland, 12-16 August 2012.

Pinto, L. and Assunção, P., *Non-uniform Asymmetric 3D Video Coding using Regions Of Perceptual Interest*, EU COST Training School on Plenoptic Capture, Processing and

Reconstruction , Sundsvall, Sweden, 16-20 June 2013. (This poster received one of the three *Best Poster Award* awarded during the training school.)

Unpublished:

Pinto, L. and Assunção, P., *Non-uniform Asymmetric 3D Video Coding Based on Regions of Perceptual Relevance* (10 page journal article awaiting submission).

Appendix E

Stereoscopic Still Images - Complimentary Data and Results

Avaliador -> Sequência		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	Média	Desvio Padrão	Intervalo Conf. 95% δ_{jkr}	S_{jkr}	**N** avals
200	Bike-SbS-HQ	5	5	5	5	5	5	5	5	5	5	5	5	5	4	5	5	4	4	5	5	4,850	0,366348	0,159848975	0,364728	20
402	cafe-SbS-asym40dB	5	4	3	4	5	3	5	5	5	5	5	4	5	4	4	5	5	3	4	4	4,350	0,74516	0,324679129	0,740821	20
501	ChampTower-SbS-sym40dB	4	5	5	5	5	5	1	4	5	4	5	5	5	4	4	5	4	4	5	5	4,450	0,944513	0,410240738	0,936047	20
801	notebook-SbS-sym40dB	3	4	5	4	5	5	5	5	4	5	5	5	5	3	5	5	5	3	5	5	4,550	0,759155	0,329623374	0,752102	20
600	Kendo-SbS-HQ	4	5	5	5	4	2	4	5	4	5	4	5	5	4	4	5	5	5	4	5	4,450	0,759155	0,328086301	0,748595	20
401	cafe-SbS-sym40dB	5	5	4	5	5	5	5	4	4	5	5	5	5	4	5	5	5	4	4	5	4,700	0,470162	0,203838016	0,465098	20
701	car-SbS-sym40dB	5	5	5	5	5	3	5	4	3	5	5	5	5	3	5	5	5	3	5	5	4,550	0,825578	0,358985398	0,819098	20
101	Balloons-SbS-sym40dB	5	4	5	3	5	4	5	5	5	5	5	5	5	4	4	5	5	3	4	5	4,550	0,686333	0,297376228	0,678524	20
203	Bike-SbS-mask1	5	5	5	5	5	5	5	5	5	5	5	5	5	4	4	5	4	4	4	5	4,750	0,444262	0,193076887	0,440544	20
504	ChampTower-SbS-mask2	5	5	5	5	5	4	5	5	5	5	5	5	5	3	4	5	5	4	5	5	4,750	0,55012	0,239786638	0,547122	20
105	Balloons-SbS-mask3	4	4	4	5	4	5	5	5	4	5	5	5	5	3	5	5	4	5	5	4	4,550	0,604805	0,259234886	0,591497	20
305	BMX-SbS-mask3	3	4	4	5	5	2	5	5	4	5	4	5	5	4	4	5	4	4	5	5	4,350	0,812728	0,350146601	0,79893	20
700	car-SbS-HQ	4	5	5	5	5	3	4	5	4	5	5	5	5	4	5	5	5	3	5	5	4,600	0,680557	0,295543318	0,674342	20
503	ChampTower-SbS-mask1	4	5	4	5	5	5	4	4	5	5	5	5	5	4	5	5	4	5	5	5	4,700	0,470162	0,203838016	0,465098	20
704	car-SbS-mask2	4	5	4	5	5	3	5	5	4	5	4	5	5	4	4	5	5	3	5	5	4,500	0,688247	0,29741872	0,678621	20
703	car-SbS-mask1	3	3	4	5	4	3	4	4	5	5	4	5	5	4	5	5	5	2	4	5	4,200	0,894427	0,38365862	0,875395	20
505	ChampTower-SbS-mask3	5	5	5	5	5	5	5	4	5	5	4	5	5	5	4	5	5	3	5	5	4,750	0,55012	0,239786638	0,547122	20
804	notebook-SbS-mask2	2	5	5	4	5	4	5	4	5	5	5	5	5	3	4	5	4	3	4	5	4,350	0,875094	0,377917437	0,862295	20
803	notebook-SbS-mask1	3	4	5	4	5	5	4	5	4	5	5	5	5	4	5	5	5	3	5	5	4,550	0,686333	0,297376228	0,678524	20
403	cafe-SbS-mask1	3	3	4	5	4	5	5	5	4	5	4	5	5	4	4	5	4	4	4	5	4,350	0,67082	0,286643938	0,654036	20
702	car-SbS-asym40dB	4	5	5	4	5	5	5	5	5	5	5	5	5	3	4	5	5	5	5	5	4,750	0,55012	0,239786638	0,547122	20
103	Balloons-SbS-mask1	4	5	5	4	5	5	5	4	5	5	5	5	4	5	4	5	5	4	5	5	4,700	0,470162	0,203838016	0,465098	20
604	Kendo-SbS-mask2	4	4	4	2	4	1	5	5	4	5	4	5	5	4	3	4	4	3	3	5	3,900	1,071153	0,456239042	1,041002	20
303	BMX-SbS-mask1	5	5	5	5	4	5	5	5	5	5	4	5	4	3	4	5	5	5	5	5	4,700	0,571241	0,248533467	0,567079	20
100	Balloons-SbS-HQ	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5,000	0	0	0	20
102	Balloons-SbS-asym40dB	4	5	5	5	4	5	5	5	5	5	5	5	5	5	5	4	5	5	5	5	4,850	0,366348	0,159848975	0,364728	20
405	cafe-SbS-mask3	3	3	4	2	4	5	5	4	4	5	4	5	5	4	4	5	5	2	4	5	4,100	0,967906	0,41443933	0,945627	20
304	BMX-SbS-mask2	4	4	4	4	4	5	5	5	5	5	5	5	5	5	4	4	5	5	5	5	4,650	0,48936	0,211564871	0,482728	20
500	ChampTower-SbS-HQ	4	5	5	4	5	3	5	5	4	5	5	5	5	4	4	5	5	3	5	5	4,550	0,686333	0,297376228	0,678524	20
201	Bike-SbS-sym40dB	4	4	4	4	4	4	5	5	4	5	5	5	5	4	4	5	5	4	5	5	4,500	0,512989	0,219134662	0,5	20
601	Kendo-SbS-sym40dB	4	5	3	5	3	2	5	5	4	5	5	5	5	5	4	5	5	3	4	5	4,350	0,933302	0,403782784	0,921312	20
202	Bike-SbS-asym40dB	4	3	4	5	5	5	4	4	5	5	5	5	5	4	4	5	5	3	5	5	4,500	0,688247	0,29741872	0,678621	20
300	BMX-SbS-HQ	4	4	4	5	4	5	5	4	4	5	3	5	5	3	4	5	5	5	5	5	4,450	0,686333	0,295671564	0,674634	20
603	Kendo-SbS-mask1	4	4	5	2	3	2	3	4	4	5	4	5	5	4	5	5	5	4	3	5	4,050	0,998683	0,427142513	0,974612	20
705	car-SbS-mask3	4	4	4	4	4	5	4	5	5	5	4	5	5	3	4	4	5	2	4	5	4,250	0,786398	0,336303043	0,767343	20
301	BMX-SbS-sym40dB	3	3	4	5	4	5	4	4	5	5	3	5	5	4	5	4	5	2	5	5	4,250	0,910465	0,391838783	0,894059	20
204	Bike-SbS-mask2	4	5	4	4	5	5	5	4	5	5	3	5	5	4	4	5	5	3	4	5	4,450	0,686333	0,295671564	0,674634	20
205	Bike-SbS-mask3	3	4	3	5	3	5	5	4	5	5	2	4	5	5	5	4	5	2	2	5	4,050	1,145931	0,493059396	1,125015	20
502	ChampTower-SbS-asym40dB	4	5	5	5	5	2	5	5	4	5	4	5	5	4	4	5	5	4	5	5	4,550	0,759155	0,329623374	0,752102	20
404	cafe-SbS-mask2	3	5	5	4	4	3	4	4	4	5	3	5	5	3	4	5	5	3	4	5	4,150	0,812728	0,345788664	0,788987	20
104	Balloons-SbS-mask2	3	3	4	4	4	5	5	5	4	5	4	5	5	4	4	5	5	4	5	5	4,400	0,680557	0,29210265	0,666491	20
302	BMX-SbS-asym40dB	4	4	5	5	4	5	4	4	5	5	4	5	5	4	4	5	4	2	5	4	4,350	0,74516	0,324679129	0,740821	20
400	cafe-SbS-HQ	4	4	5	5	5	3	4	5	5	5	5	5	5	4	5	5	4	5	5	5	4,650	0,587143	0,254909086	0,581627	20
605	Kendo-SbS-mask3	3	3	5	1	3	1	4	4	4	5	4	5	5	4	3	4	5	4	3	5	3,750	1,208522	0,514530701	1,174006	20
802	notebook-SbS-asym40dB	4	4	5	4	5	5	5	5	4	5	5	5	5	5	4	5	5	3	5	5	4,650	0,587143	0,254909086	0,581627	20
805	notebook-SbS-mask3	4	4	5	4	3	5	4	5	5	5	3	5	5	5	3	4	4	3	4	5	4,250	0,786398	0,336303043	0,767343	20
602	Kendo-SbS-asym40dB	4	4	4	2	3	2	4	4	4	5	4	5	5	3	3	5	5	3	4	4	3,850	0,933302	0,408759497	0,932667	20
800	notebook-SbS-HQ	4	3	5	5	4	5	5	5	5	5	5	5	5	4	5	5	5	3	5	5	4,650	0,67082	0,291886261	0,665997	20

Figure E.1: Discriminated subjective scores per image, impairment and observer.

Avaliador -> Sequência		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	Média	Desvio Padrão	Intervalo Conf. 95% δ_{jkr}	S_{jkr}	**N** avals
104	Balloons-SbS-eq3	4	4	5	4	4	4	4	3	3	4	2	5	3	2	5						3,733	0,96115	0,486409064	0,96115	15
802	notebook-SbS-mask4	4	4	4	5	3	4	4	3	4	4	4	4	4	3	4						3,867	0,516398	0,261333333	0,516398	15
505	ChampTower-SbS-eq5	5	5	4	5	5	5	2	4	3	5	5	5	4	4	4						4,333	0,899735	0,455328941	0,899735	15
204	Bike-SbS-eq3	5	5	5	2	4	5	4	5	4	3	4	5	5	3	3						4,133	0,99043	0,501226939	0,99043	15
301	BMX-SbS-asym40dB	5	5	3	5	5	4	3	4	3	2	5	5	5	5	5						4,267	1,032796	0,522666667	1,032796	15
102	Balloons-SbS-mask4	4	5	5	4	5	5	2	5	4	4	2	5	5	5	4						4,267	1,032796	0,522666667	1,032796	15
100	Balloons-SbS-HQ	4	5	5	5	5	4	4	5	3	5	5	5	4	5	5						4,600	0,632456	0,32006666	0,632456	15
504	ChampTower-SbS-eq3	5	5	5	5	5	5	1	4	5	4	5	5	5	5	4						4,533	1,060099	0,536484027	1,060099	15
501	ChampTower-SbS-asym40dB	5	5	5	5	5	3	3	5	4	3	5	5	3	4	3						4,200	0,941124	0,476274431	0,941124	15
302	BMX-SbS-mask4	4	5	5	3	3	3	4	4	4	3	5	5	4	4	3						3,933	0,798809	0,404252946	0,798809	15
703	car-SbS-mask5	2	4	4	4	4	3	1	3	2	3	2	4	3	2	4						3,000	1	0,506069824	1	15
803	notebook-SbS-mask5	4	3	5	5	3	3	4	4	3	3	3	4	4	4	4						3,733	0,703732	0,356137302	0,703732	15
705	car-SbS-eq5	3	4	4	4	5	2	3	4	2	4	3	4	3	2	5						3,467	0,99043	0,501226939	0,99043	15
805	notebook-SbS-eq5	4	4	5	3	4	4	3	5	4	4	4	4	3	2	3						3,733	0,798809	0,404252946	0,798809	15
601	Kendo-SbS-asym40dB	4	5	4	1	4	5	2	4	5	4	4	5	2	4	5						3,867	1,245946	0,630535575	1,245946	15
200	Bike-SbS-HQ	4	5	5	5	5	5	5	5	4	5	5	5	5	5	4						4,800	0,414039	0,209532814	0,414039	15
201	Bike-SbS-asym40dB	4	5	3	4	5	4	2	5	4	5	5	5	5	5	4						4,333	0,899735	0,455328941	0,899735	15
801	notebook-SbS-asym40dB	4	5	4	5	5	5	2	4	3	5	5	5	4	5	4						4,333	0,899735	0,455328941	0,899735	15
400	cafe-SbS-HQ	5	5	5	5	5	4	3	5	4	5	5	5	4	5	5						4,667	0,617213	0,312353077	0,617213	15
701	car-SbS-asym40dB	5	5	5	5	5	5	4	4	4	5	5	4	3	5	4						4,533	0,63994	0,323854563	0,63994	15
101	**Balloons-SbS-asym40dB**	**4**	**5**	**4**	**4**	**5**	**5**	**4**	**4**	**5**	**4**	**5**	**5**	**4**	**5**	**5**						**4,533**	**0,516398**	**0,261333333**	**0,516398**	**15**
804	notebook-SbS-eq3	5	5	5	4	4	3	3	4	4	5	5	5	5	4	4						4,333	0,723747	0,366266448	0,723747	15
105	Balloons-SbS-eq5	4	5	5	4	4	3	4	4	3	4	4	5	5	4	3						4,067	0,703732	0,356137302	0,703732	15
202	Bike-SbS-mask4	5	5	4	3	4	3	2	4	3	4	4	4	4	4	4						3,800	0,774597	0,392	0,774597	15
401	cafe-SbS-asym40dB	5	4	4	5	5	5	3	5	4	4	5	4	5	5	5						4,533	0,63994	0,323854563	0,63994	15
103	Balloons-SbS-mask5	1	5	5	2	4	5	1	5	3	3	2	5	3	3	4						3,400	1,454058	0,735855058	1,454058	15
300	BMX-SbS-HQ	5	5	3	5	5	3	2	5	3	5	5	5	5	5	4						4,333	1,046536	0,529620409	1,046536	15
503	**ChampTower-SbS-mask5**	**4**	**5**	**5**	**5**	**5**	**4**	**4**	**4**	**3**	**5**	**3**	**4**	**5**	**5**	**3**						**4,267**	**0,798809**	**0,404252946**	**0,798809**	**15**
404	cafe-SbS-eq3	5	5	5	4	4	4	2	4	4	3	3	5	3	3	4						3,867	0,915475	0,463294483	0,915475	15
700	car-SbS-HQ	5	5	5	4	5	5	2	4	5	4	5	5	5	5	4						4,533	0,833809	0,42196577	0,833809	15
402	cafe-SbS-mask4	1	4	5	1	3	2	1	2	1	3	2	5	1	1	3						2,333	1,447494	0,732532896	1,447494	15
405	cafe-SbS-eq5	1	3	4	2	3	2	1	1	2	2	2	5	1	1	4						2,267	1,279881	0,647709125	1,279881	15
800	notebook-SbS-HQ	5	5	5	5	5	5	4	5	4	5	5	4	5	5	3						4,667	0,617213	0,312353077	0,617213	15
303	BMX-SbS-mask5	4	4	3	4	5	4	2	5	3	3	3	5	4	2	3						3,600	0,985611	0,498787864	0,985611	15
500	**ChampTower-SbS-HQ**	**5**	**5**	**5**	**5**	**5**	**5**	**5**	**5**	**4**	**5**	**5**	**5**	**4**	**5**	**5**						**4,867**	**0,351866**	**0,178068651**	**0,351866**	**15**
304	BMX-SbS-eq3	5	5	3	4	2	3	4	4	4	5	3	5	5	3	4						3,933	0,96115	0,486409064	0,96115	15
203	Bike-SbS-mask5	3	5	4	3	5	5	5	4	2	4	4	5	5	3	4						4,067	0,96115	0,486409064	0,96115	15
305	BMX-SbS-eq5	3	4	3	3	2	4	4	4	4	4	3	5	4	2	3						3,467	0,833809	0,42196577	0,833809	15
403	cafe-SbS-mask5	1	4	2	1	2	2	2	2	1	2	3	5	2	1	4						2,267	1,222799	0,61882182	1,222799	15
604	Kendo-SbS-eq3	3	5	4	3	4	5	2	4	4	3	3	5	4	3	3						3,667	0,899735	0,455328941	0,899735	15
704	car-SbS-eq3	4	5	3	4	5	3	5	4	3	4	4	5	5	5	5						4,267	0,798809	0,404252946	0,798809	15
502	**ChampTower-SbS-mask4**	**5**	**5**	**5**	**4**	**5**	**4**	**4**	**3**	**3**	**3**	**5**	**5**	**5**	**5**	**4**						**4,333**	**0,816497**	**0,413204281**	**0,816497**	**15**
702	car-SbS-mask4	2	4	4	3	4	4	2	3	1	4	2	4	4	4	5						3,333	1,112697	0,563102517	1,112697	15
600	Kendo-SbS-HQ	5	5	4	5	5	5	4	5	5	4	5	5	4	5	5						4,733	0,457738	0,231647241	0,457738	15
205	Bike-SbS-eq5	3	5	5	3	4	3	5	4	2	3	4	4	4	3	4						3,733	0,883715	0,447221546	0,883715	15

Figure E.2: Discriminated subjective scores per image, impairment and observer (2nd tests session).

BALLOONS																		
Total Rate	Total PSNR		View ID	QP	Rate	SNR Y		View ID	QP_Mask	QP_View	Rate	SNR Y	SNR In	SNR Out		MOS	Desvio Padrão	Intervalo Conf	% MBs Mask
256016	41,2615		0	30	190896	41,58		1	30	30	65120	40,943	0	0		5	0	0	33,76
158272	38,05875		0	34	124368	38,3047		1	34	34	33904	37,8128	35,2647	40,0801		4,55	0,686332741	0,297376228	33,76
143184	37,60585		0	34	124368	38,3047		1	38	38	18816	36,907	34,4137	39,0909		4,85	0,366347549	0,159848975	33,76
139800	37,40245		0	34	124368	38,3047		1	38	41	15432	36,5002	34,3772	38,1832		4,7	0,470162346	0,203838016	33,76
138256	37,18175		0	34	124368	38,3047		1	38	44	13888	36,0588	34,3528	37,2804		4,4	0,680557047	0,29210265	33,76
137920	36,97505		0	34	124368	38,3047		1	38	47	13552	35,6454	34,359	36,4844		4,55	0,604805319	0,259234886	33,76

BIKE																		
Total Rate	Total PSNR		View ID	QP	Rate	SNR Y		View ID	QP_Mask	QP_View	Rate	SNR Y				MOS	Desvio Padrão	Intervalo Conf	% MBs Mask
297832	40,8075		0	28	222064	40,9		1	28	28	75768	40,715	0	0		4,85	0,366347549	0,159848975	24,18
205552	37,59695		0	31	159296	37,8627		1	31	31	46256	37,3312	33,369	40,1304		4,5	0,512989176	0,219134662	24,18
185248	37,0499		0	31	159296	37,8627		1	34	34	25952	36,2371	32,1316	39,2624		4,5	0,688247202	0,29741872	24,18
181936	36,90265		0	31	159296	37,8627		1	34	37	22640	35,9426	32,1025	38,5634		4,75	0,444261658	0,193076887	24,18
178912	36,6653		0	31	159296	37,8627		1	34	40	19616	35,4679	32,1139	37,4839		4,45	0,686332741	0,295671564	24,18
177176	36,2629		0	31	159296	37,8627		1	34	43	17880	34,6631	32,0993	35,9505		4,05	1,145931017	0,493059396	24,18

BMX																		
Total Rate	Total PSNR		View ID	QP	Rate	SNR Y		View ID	QP_Mask	QP_View	Rate	SNR Y	SNR In	SNR Out		MOS	Desvio Padrão	Intervalo Conf	% MBs Mask
444760	40,7545		0	28	325752	41,223		1	28	28	119008	40,286	0	0		4,45	0,686332741	0,295671564	29,30
357672	37,7844		0	30	274288	38,3091		1	30	30	83384	37,2597	33,4252	41,1061		4,25	0,910465468	0,391838783	29,30
324752	37,312		0	30	274288	38,3091		1	33	33	50464	36,3149	32,3635	40,4589		4,35	0,74515982	0,324679129	29,30
319512	37,1041		0	30	274288	38,3091		1	33	36	45224	35,8991	32,3402	39,1407		4,7	0,571240571	0,248533467	29,30
316608	36,9014		0	30	274288	38,3091		1	33	39	42320	35,4937	32,3245	38,0526		4,65	0,489360485	0,211564871	29,30
314328	36,48465		0	30	274288	38,3091		1	33	42	40040	34,6602	32,329	36,1746		4,35	0,812727701	0,350146601	29,30

CAFE																		
Total Rate	Total PSNR		View ID	QP	Rate	SNR Y		View ID	QP_Mask	QP_View	Rate	SNR Y	SNR In	SNR Out		MOS	Desvio Padrão	Intervalo Conf	% MBs Mask
221320	40,8475		0	29	164712	41,407		1	29	29	56608	40,288	0	0		4,65	0,587142949	0,254909086	32,86
127664	37,52405		0	33	101328	38,0085		1	33	33	26336	37,0396	34,1806	39,6823		4,7	0,470162346	0,203838016	32,86
116272	37,19745		0	33	101328	38,0085		1	37	37	14944	36,3864	33,4917	39,0916		4,35	0,74515982	0,324679129	32,86
114456	37,07985		0	33	101328	38,0085		1	37	40	13128	36,1512	33,488	38,4734		4,35	0,670820393	0,286643938	32,86
114216	36,9218		0	33	101328	38,0085		1	37	43	12888	35,8351	33,4886	37,71		4,15	0,812727701	0,345788664	32,86
113688	36,83105		0	33	101328	38,0085		1	37	46	12360	35,6536	33,4918	37,3009		4,1	0,967906042	0,41443933	32,86

Figure E.3: Complimentary data from stereoscopic still images subjective testing (session 1 - part 1).

CHAMPTOWER																		
Total Rate	Total PSNR		View ID	QP	Rate	SNR Y		View ID	QP_Mask	QP_View	Rate	SNR Y	SNR In	SNR Out		MOS	Desvio Padrão	Intervalo Conf	% MBs Mask
385488	40,5235		0	30	252744	40,717		1	30	30	132744	40,33	0	0		4,55	0,686332741	0,297376228	44,47
261600	37,537		0	33	183824	37,9318		1	33	33	77776	37,1422	34,2338	43,405		4,45	0,944513241	0,410240738	44,47
242504	37,09085		0	33	183824	37,9318		1	35	35	58680	36,2499	33,343	42,5033		4,55	0,759154655	0,329623374	44,47
240736	36,98715		0	33	183824	37,9318		1	35	38	56912	36,0425	33,3526	41,0845		4,7	0,470162346	0,203838016	44,47
239624	36,9271		0	33	183824	37,9318		1	35	41	55800	35,9224	33,3523	40,4369		4,75	0,550119604	0,239786638	44,47
238344	36,8061		0	33	183824	37,9318		1	35	44	54520	35,6804	33,3168	39,4391		4,75	0,550119604	0,239786638	44,47

KENDO																		
Total Rate	Total PSNR		View ID	QP	Rate	SNR Y		View ID	QP_Mask	QP_View	Rate	SNR Y	SNR In	SNR Out		MOS	Desvio Padrão	Intervalo Conf	% MBs Mask
117904	41,047		0	33	95912	41,218		1	33	33	21992	40,876	0	0		4,45	0,759154655	0,328086301	21,42
76096	37,9179		0	37	63696	37,9772		1	37	37	12400	37,8586	33,7092	40,3463		4,35	0,933302004	0,403782784	21,42
70384	37,42625		0	37	63696	37,9772		1	42	42	6688	36,8753	33,139	38,894		3,85	0,933302004	0,408759497	21,42
69064	37,16565		0	37	63696	37,9772		1	42	45	5368	36,3541	33,1238	37,9088		4,05	0,998683344	0,427142513	21,42
68800	36,80705		0	37	63696	37,9772		1	42	48	5104	35,6369	33,0904	36,7013		3,9	1,071152847	0,456239042	21,42
68320	36,3701		0	37	63696	37,9772		1	42	51	4624	34,763	33,0888	35,3588		3,75	1,208522369	0,514530701	21,42

CAR																		
Total Rate	Total PSNR		View ID	QP	Rate	SNR Y		View ID	QP_Mask	QP_View	Rate	SNR Y	SNR In	SNR Out		MOS	Desvio Padrão	Intervalo Conf	% MBs Mask
321816	40,6265		0	27	263448	40,869		1	27	27	58368	40,384	0	0		4,6	0,680557047	0,295543318	21,35
210264	37,5726		0	30	178144	37,679		1	30	30	32120	37,4662	35,1586	38,3835		4,55	0,825577947	0,358985398	21,35
192992	37,2368		0	30	178144	37,679		1	36	36	14848	36,7946	34,3029	37,8206		4,75	0,550119604	0,239786638	21,35
190816	37,1145		0	30	178144	37,679		1	36	39	12672	36,55	34,2731	37,4499		4,2	0,894427191	0,38365862	21,35
189656	36,88965		0	30	178144	37,679		1	36	42	11512	36,1003	34,2549	36,7743		4,5	0,688247202	0,29741872	21,35
188496	36,6296		0	30	178144	37,679		1	36	45	10352	35,5802	34,2822	36,0117		4,25	0,786397516	0,336303043	21,35

NOTEBOOK																		
Total Rate	Total PSNR		View ID	QP	Rate	SNR Y		View ID	QP_Mask	QP_View	Rate	SNR Y	SNR In	SNR Out		MOS	Desvio Padrão	Intervalo Conf	% MBs Mask
301088	41,019		0	28	203760	41,545		1	28	28	97328	40,493	0	0		4,65	0,670820393	0,291886261	28,99
205312	37,92905		0	31	148592	38,5141		1	31	31	56720	37,344	35,1005	38,7475		4,55	0,759154655	0,329623374	28,99
181104	37,5099		0	31	148592	38,5141		1	35	35	32512	36,5057	34,1436	38,0244		4,65	0,587142949	0,254909086	28,99
175896	37,29125		0	31	148592	38,5141		1	35	38	27304	36,0684	34,0787	37,2448		4,55	0,686332741	0,297376228	28,99
173304	37,0478		0	31	148592	38,5141		1	35	41	24712	35,5815	34,0814	36,3827		4,35	0,87509398	0,377917437	28,99
171800	36,75705		0	31	148592	38,5141		1	35	44	23208	35	34,0962	35,431		4,25	0,786397516	0,336303043	28,99

Figure E.4: Complimentary data from stereoscopic still images subjective testing (session 1 - part 2).

BALLOONS																		
Total Rate	Total PSNR		View ID	QP	Rate	SNR Y		View ID	QP_Mask	QP_View	Rate	SNR Y	SNR In	SNR Out		MOS	Desvio Padrão	Intervalo Conf	% MBs Mask
256016	41,2615		0	30	190896	41,58		1	30	30	65120	40,943	0	0		4,6	0,632455532	0,32006666	33,76
143184	37,60585		0	30	124368	38,3047		1	38	38	18816	36,907	34,4137	39,0909		4,533333	0,516397779	0,261333333	33,76
134584	37,08215		0	30	124368	38,3047		1	42	42	10216	35,8596	33,4761	37,8842		3,733333	0,961150105	0,486409064	33,76
130152	36,25285		0	30	124368	38,3047		1	48	48	5784	34,201	31,763	36,3034		4,066667	0,703731551	0,356137302	33,76
137672	36,51975		0	30	124368	38,3047		1	38	50	13304	34,7348	34,3427	34,9491		4,266667	1,032795559	0,522666667	33,76
137560	36,21015		0	30	124368	38,3047		1	38	51	13192	34,1156	34,354	33,999		3,4	1,45405836	0,735855058	33,76

BIKE																		
Total Rate	Total PSNR		View ID	QP	Rate	SNR Y		View ID	QP_Mask	QP_View	Rate	SNR Y	SNR In	SNR Out		MOS	Desvio Padrão	Intervalo Conf	% MBs Mask
297832	40,8075		0	28	222064	40,9		1	28	28	75768	40,715	0	0		4,8	0,414039336	0,209532814	24,18
185248	37,0499		0	28	159296	37,8627		1	34	34	25952	36,2371	32,1316	39,2624		4,333333	0,899735411	0,455328941	24,18
168296	36,24675		0	28	159296	37,8627		1	42	42	9000	34,6308	30,966	37,0156		4,133333	0,990430402	0,501226939	24,18
164272	35,10395		0	28	159296	37,8627		1	46	46	4976	32,3452	30,2757	33,2852		3,733333	0,883715102	0,447221546	24,18
175256	35,50965		0	28	159296	37,8627		1	34	46	15960	33,1566	32,0972	33,557		3,8	0,774596669	0,392	24,18
174688	35,3195		0	28	159296	37,8627		1	34	49	15392	32,7763	32,0852	33,022		4,066667	0,961150105	0,486409064	24,18

BMX																		
Total Rate	Total PSNR		View ID	QP	Rate	SNR Y		View ID	QP_Mask	QP_View	Rate	SNR Y	SNR In	SNR Out		MOS	Desvio Padrão	Intervalo Conf	% MBs Mask
444760	40,7545		0	28	325752	41,223		1	28	28	119008	40,286	0	0		4,333333	1,046536237	0,529620409	29,30
324752	37,312		0	28	274288	38,3091		1	33	33	50464	36,3149	32,3635	40,4589		4,266667	1,032795559	0,522666667	29,30
299192	36,5044		0	28	274288	38,3091		1	38	38	24904	34,6997	30,9462	38,3553		3,933333	0,961150105	0,486409064	29,30
286272	35,33415		0	28	274288	38,3091		1	44	44	11984	32,3592	28,8838	35,4399		3,466667	0,833809388	0,42196577	29,30
312408	35,93305		0	28	274288	38,3091		1	33	45	38120	33,557	32,3131	34,6035		3,933333	0,798808637	0,404252946	29,30
311224	35,47715		0	28	274288	38,3091		1	33	48	36936	32,6452	32,3087	32,7926		3,6	0,985610761	0,498787864	29,30

CAFE																		
Total Rate	Total PSNR		View ID	QP	Rate	SNR Y		View ID	QP_Mask	QP_View	Rate	SNR Y	SNR In	SNR Out		MOS	Desvio Padrão	Intervalo Conf	% MBs Mask
221320	40,8475		0	29	164712	41,407		1	29	29	56608	40,288	0	0		4,666667	0,6172134	0,312353077	32,86
116272	37,19745		0	29	101328	38,0085		1	37	37	14944	36,3864	33,4917	39,0916		4,533333	0,639940473	0,323854563	32,86
110784	36,7239		0	29	101328	38,0085		1	42	42	9456	35,4393	32,6308	37,9956		3,866667	0,915475416	0,463294483	32,86
106760	35,3519		0	29	101328	38,0085		1	49	49	5432	32,6953	30,9817	33,8685		2,266667	1,279880947	0,647709125	32,86
113984	35,6859		0	29	101328	38,0085		1	37	49	12656	33,3633	33,4643	33,3148		2,333333	1,447493729	0,732532896	32,86
113704	35,29405		0	29	101328	38,0085		1	37	51	12376	32,5796	33,4573	32,2072		2,266667	1,222799287	0,61882182	32,86

Figure E.5: Complimentary data from stereoscopic still images subjective testing (session 2 - part 1).

CHAMPTOWER																		
Total Rate	Total PSNR		View ID	QP	Rate	SNR Y		View ID	QP_Mask	QP_View	Rate	SNR Y	SNR In	SNR Out		MOS	Desvio Padrão	Intervalo Conf	% MBs Mask
385488	40,5235		0	30	252744	40,717		1	30	30	132744	40,33	0	0		4,866667	0,351865775	0,178068651	44,47
242504	36,2499		0	30	183824	36,2499		1	35	35	58680	36,2499	33,343	42,5033		4,2	0,941123948	0,476274431	44,47
229080	35,8593		0	30	183824	36,2499		1	37	37	45256	35,4687	32,6143	41,3904		4,533333	1,060098827	0,536484027	44,47
222016	35,6239		0	30	183824	36,2499		1	38	38	38192	34,9979	32,1084	41,1377		4,333333	0,899735411	0,455328941	44,47
238080	35,8434		0	30	183824	36,2499		1	35	47	54256	35,4369	33,3291	38,4489		4,333333	0,816496581	0,413204281	44,47
237544	35,72235		0	30	183824	36,2499		1	35	50	53720	35,1948	33,3374	37,6157		4,266667	0,798808637	0,404252946	44,47

KENDO																		
Total Rate	Total PSNR		View ID	QP	Rate	SNR Y		View ID	QP_Mask	QP_View	Rate	SNR Y	SNR In	SNR Out		MOS	Desvio Padrão	Intervalo Conf	% MBs Mask
117904	41,047		0	33	95912	41,218		1	33	33	21992	40,876	0	0		4,733333	0,457737708	0,231647241	21,42
70384	36,8753		0	33	63696	36,8753		1	42	42	6688	36,8753	33,139	38,894		3,866667	1,245945806	0,630535575	21,42
66952	35,8799		0	33	63696	36,8753		1	49	49	3256	34,8845	31,6178	36,4693		3,666667	0,899735411	0,455328941	21,42

CAR																		
Total Rate	Total PSNR		View ID	QP	Rate	SNR Y		View ID	QP_Mask	QP_View	Rate	SNR Y	SNR In	SNR Out		MOS	Desvio Padrão	Intervalo Conf	% MBs Mask
321816	40,6265		0	27	263448	40,869		1	27	27	58368	40,384	0	0		4,533333	0,833809388	0,42196577	21,35
192992	36,7946		0	27	178144	36,7946		1	36	36	14848	36,7946	34,3029	37,8206		4,533333	0,639940473	0,323854563	21,35
186976	36,3106		0	27	178144	36,7946		1	42	42	8832	35,8266	33,3847	36,8223		4,266667	0,798808637	0,404252946	21,35
183792	34,9886		0	27	178144	36,7946		1	49	49	5648	33,1826	30,6812	34,2144		3,466667	0,990430402	0,501226939	21,35
188160	35,7541		0	27	178144	36,7946		1	36	48	10016	34,7136	34,2382	34,8522		3,333333	1,112697281	0,563102517	21,35
187592	34,9939		0	27	178144	36,7946		1	36	51	9448	33,1932	34,2397	32,9477		3	1	0,506069824	21,35

NOTEBOOK																		
Total Rate	Total PSNR		View ID	QP	Rate	SNR Y		View ID	QP_Mask	QP_View	Rate	SNR Y	SNR In	SNR Out		MOS	Desvio Padrão	Intervalo Conf	% MBs Mask
301088	41,019		0	28	203760	41,545		1	28	28	97328	40,493	0	0		4,666667	0,6172134	0,312353077	28,99
195312	36,5057		0	28	162800	36,5057		1	35	35	32512	36,5057	34,1436	38,0244		4,333333	0,899735411	0,455328941	28,99
181368	35,8348		0	28	162800	36,5057		1	40	40	18568	35,1639	32,8616	36,6237		4,333333	0,723746864	0,366266448	28,99
173736	34,85915		0	28	162800	36,5057		1	46	46	10936	33,2126	30,8862	34,6959		3,733333	0,798808637	0,404252946	28,99
184872	35,3577		0	28	162800	36,5057		1	35	47	22072	34,2097	34,0318	34,2845		3,866667	0,516397779	0,261333333	28,99
183952	34,75765		0	28	162800	36,5057		1	35	50	21152	33,0096	34,0413	32,65		3,733333	0,703731551	0,356137302	28,99

Figure E.6: Complimentary data from stereoscopic still images subjective testing (session 2 - part 2).

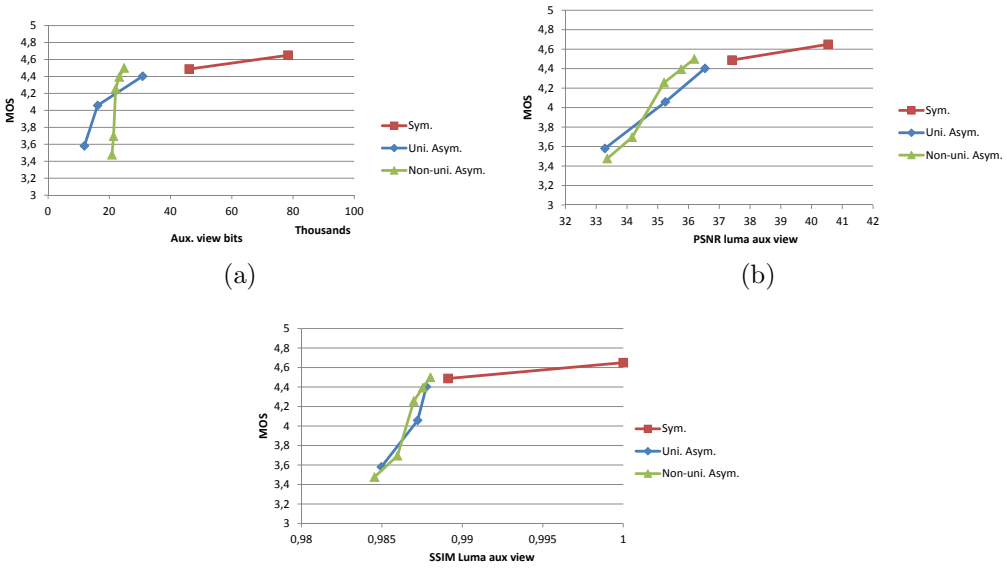


Figure E.7: MOS scores across all images according to rate (a), PSNR (b) and SSIM (c) of the auxiliary view.

Complete results for eye dominance:

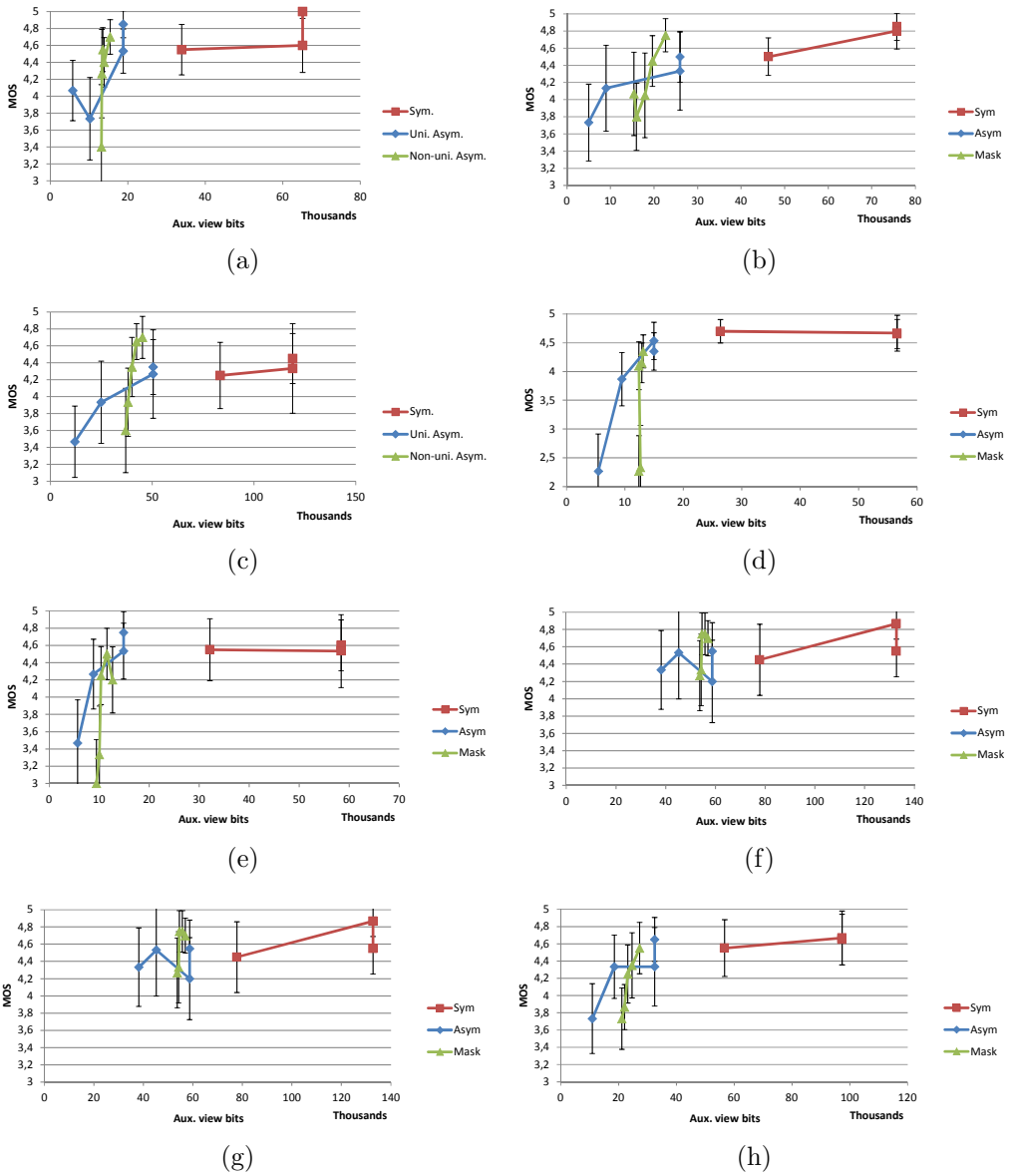


Figure E.8: Individual MOS scores versus auxiliary view size for images Balloons (a), Bike (b), BMX (c) and Cafe (d) , Car (e), Champagne Tower (f), Kendo (g) and Notebook (h).

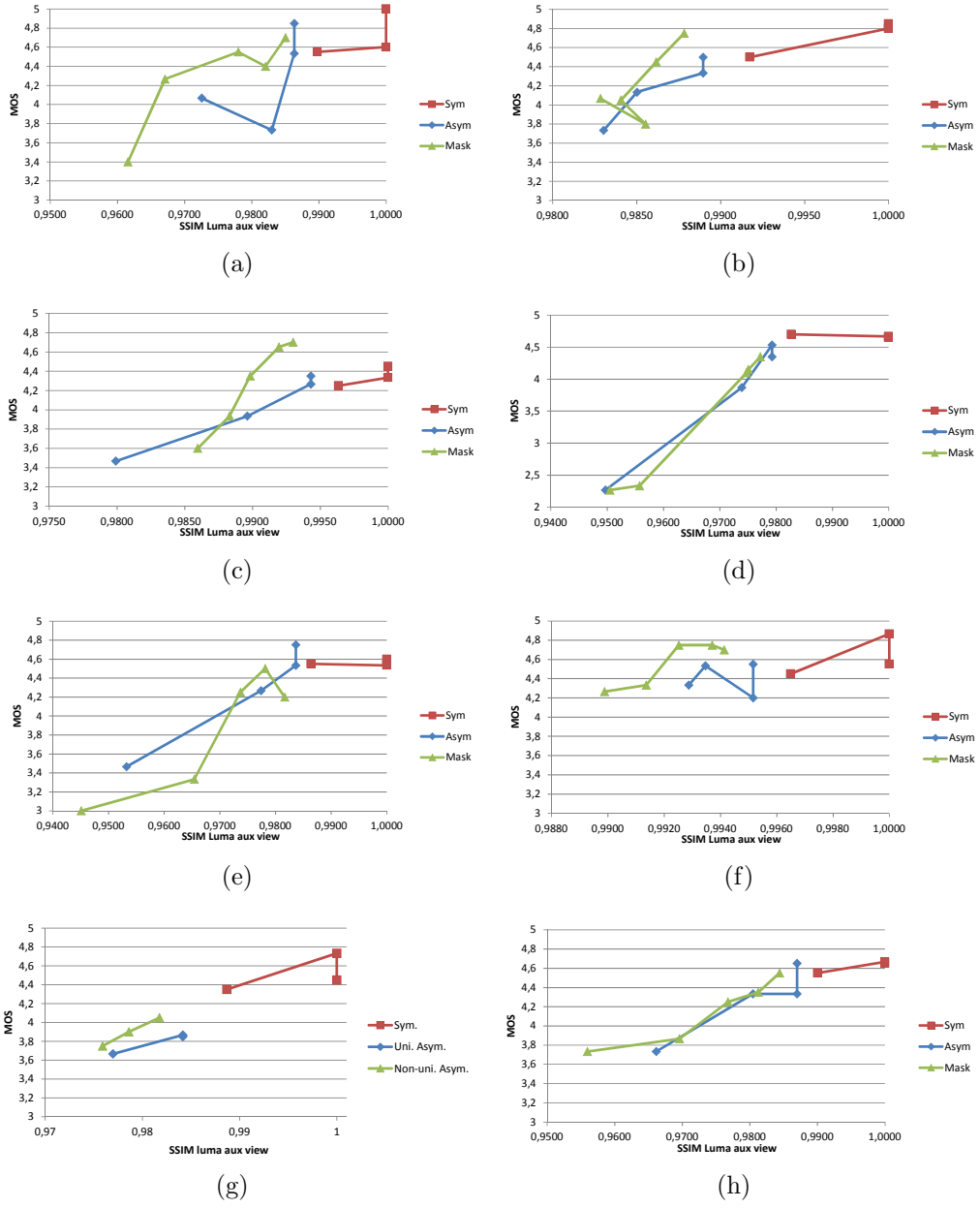


Figure E.9: Individual MOS scores versus SSIM for images Balloons (a), Bike (b), BMX (c), Cafe (d), Car (e), Champagne Tower (f), Kendo (g) and Notebook (h).

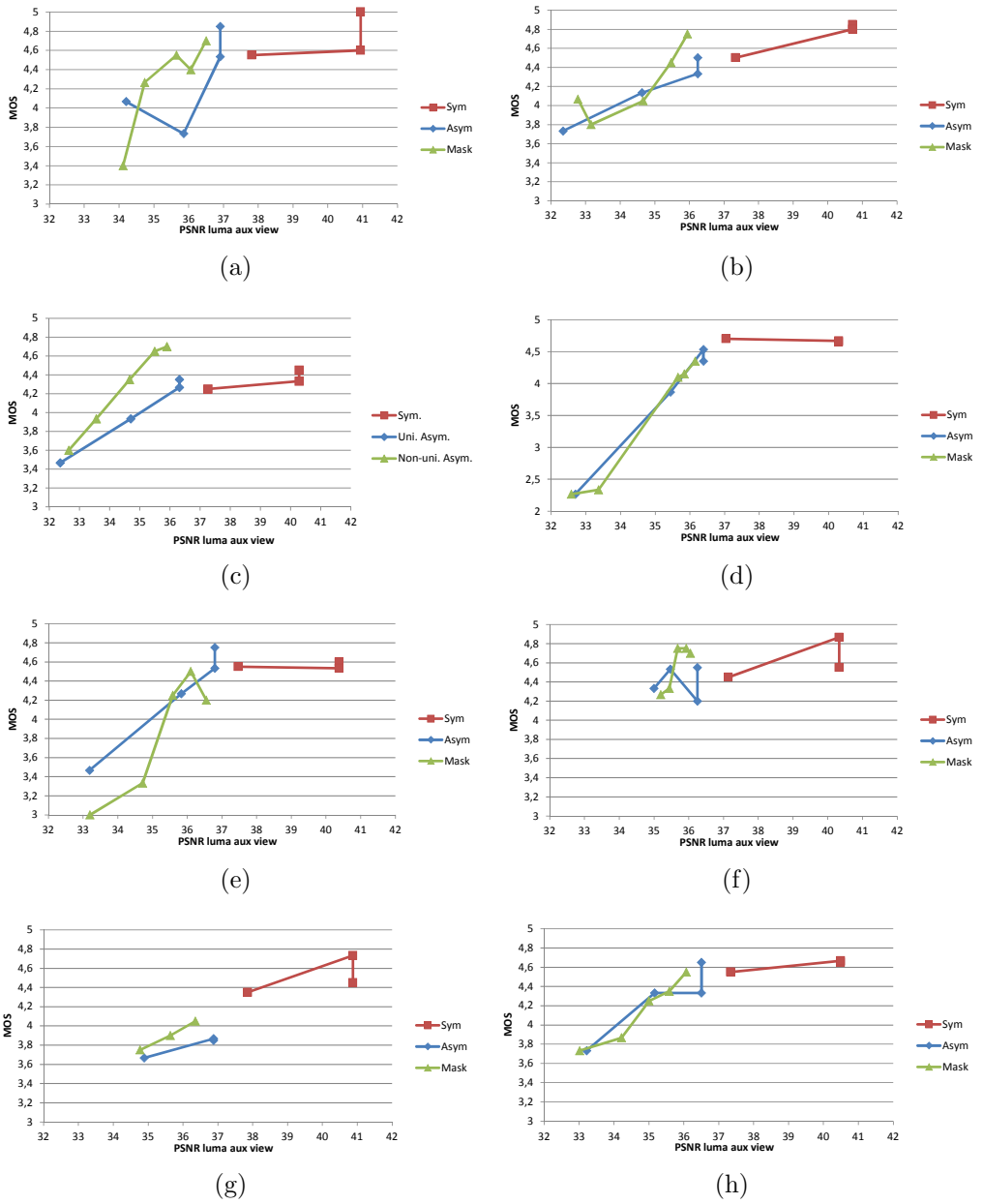


Figure E.10: Individual MOS scores versus PSNR for images Balloons (a), Bike (b), BMX (c), Cafe (d), Car (e), Champagne Tower (f), Kendo (g) and Notebook (h).

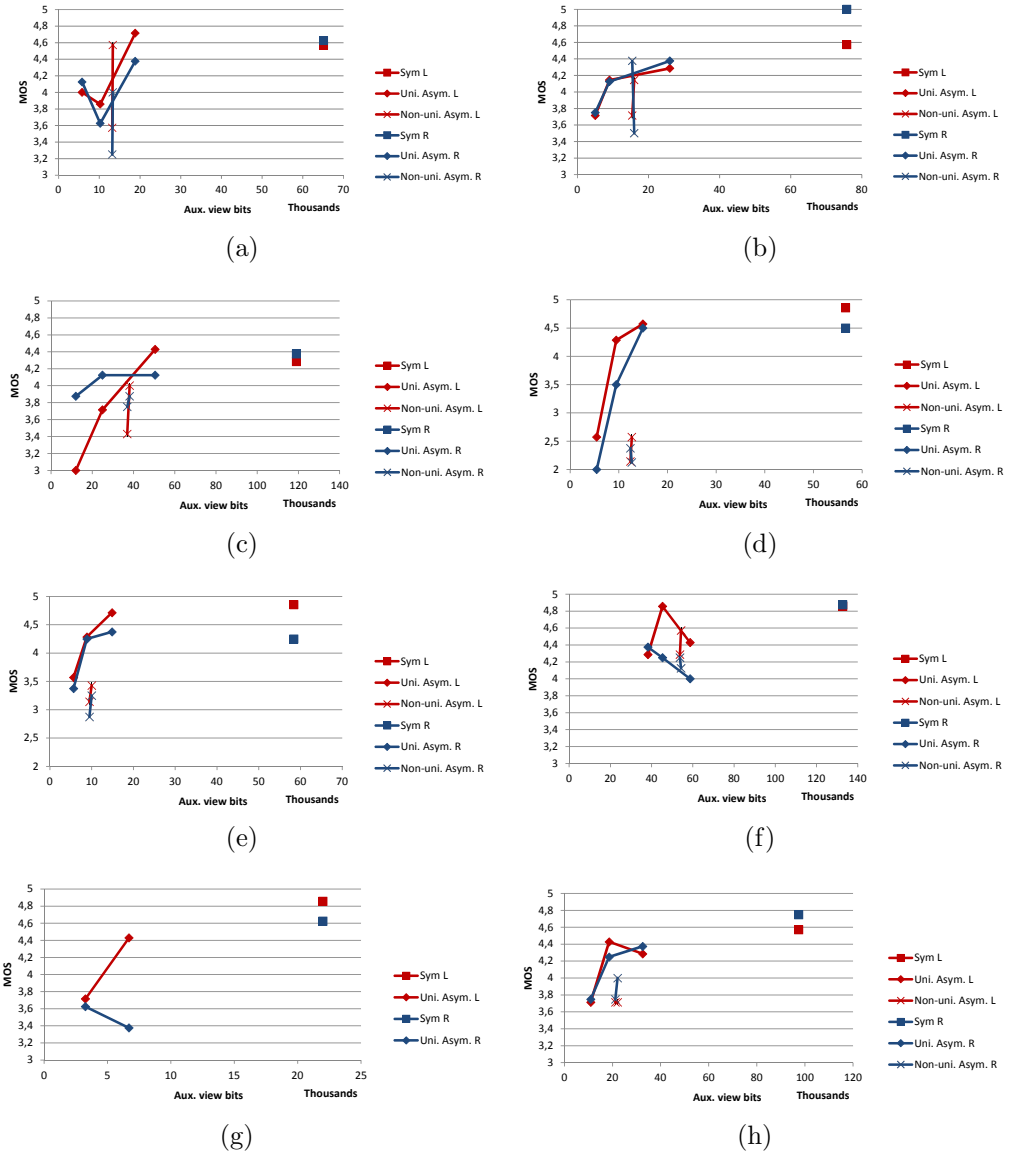


Figure E.11: Individual MOS results for left and right eye dominant observers for images Balloons (a), Bike (b), BMX (c), cafe (d), car (e), Champagne Tower (f), Kendo (g) and notebook (h).

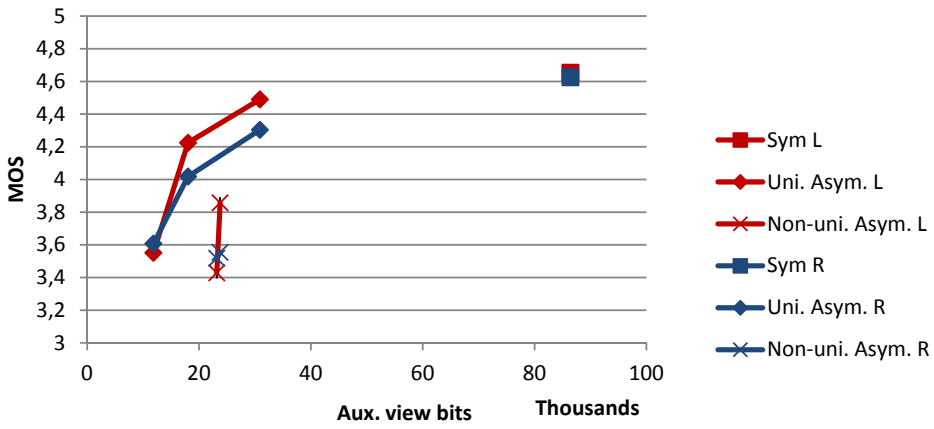


Figure E.12: Mean MOS results for left and right eye dominant observers across all images.

Appendix F

Stereoscopic Video - Complimentary Data and Results

Avaliador -> Sequência	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	Média	Desvio Padrão	Intervalo Conf. 95% δ_{jkr}	S_{jkr}
Kendo_Aval4	3	3	2	3	2	3	3	2	2	4	4	3	4	4	3	2	2	1	4		2,842	0,898342	0,403943625	0,898342
ChampTower_Aval2	4	4	4	4	3	5	5	5	5	4	5	4	5	5	4	4	4	4	5		4,368	0,597265	0,268562971	0,597265
Balloons_Aval3	2	5	3	3	2	3	1	2	4	2	4	3	3	2	3	3	1	3	4		2,789	1,031662	0,463892028	1,031662
Kendo_Aval2	3	3	2	4	3	5	4	3	4	4	5	4	4	5	4	5	3	4	4		3,842	0,83421	0,375106592	0,83421
ChampTower_Aval1	4	3	5	5	3	5	5	5	4	5	5	4	5	5	5	4	4	4	4		4,421	0,692483	0,311378156	0,692483
ChampTower_Aval3	3	4	2	4	3	3	4	4	3	3	5	3	4	5	3	2	2	3	4		3,368	0,895081	0,4024774	0,895081
Kendo_REF	5	5	5	5	5	5	4	5	3	5	5	5	4	5	5	5	5	5	4		4,737	0,561951	0,252684205	0,561951
Balloons_REF	5	5	5	5	4	5	5	5	5	5	4	4	5	4	5	5	5	4	5		4,737	0,452414	0,203430112	0,452414
ChampTower_Aval4	3	2	2	2	3	2	3	2	2	1	4	2	2	5	2	1	1	1	2		2,211	1,031662	0,463892028	1,031662
Balloons_Aval2	2	3	3	3	2	4	2	3	4	2	4	3	2	5	2	1	2	2	2		2,684	1,00292	0,450967702	1,00292
Kendo_Aval3	2	4	3	3	1	2	2	1	2	3	4	3	5	5	3	3	3	1	1		2,684	1,249561	0,561871296	1,249561
ChampTower_REF	5	4	5	5	4	4	5	5	5	5	4	5	5	5	5	5	5	5	5		4,789	0,418854	0,188339686	0,418854
Kendo_Aval1	4	3	3	4	3	4	4	4	2	3	4	4	3	5	4	4	4	2	5		3,632	0,830698	0,373527189	0,830698
Balloons_Aval4	2	2	1	2	2	1	1	1	1	2	3	2	1	2	1	1	1	1	2		1,526	0,611775	0,275087719	0,611775
Balloons_Aval1	4	3	4	3	2	4	3	4	2	4	3	3	3	4	4	2	3	3	3		3,211	0,713283	0,320730929	0,713283

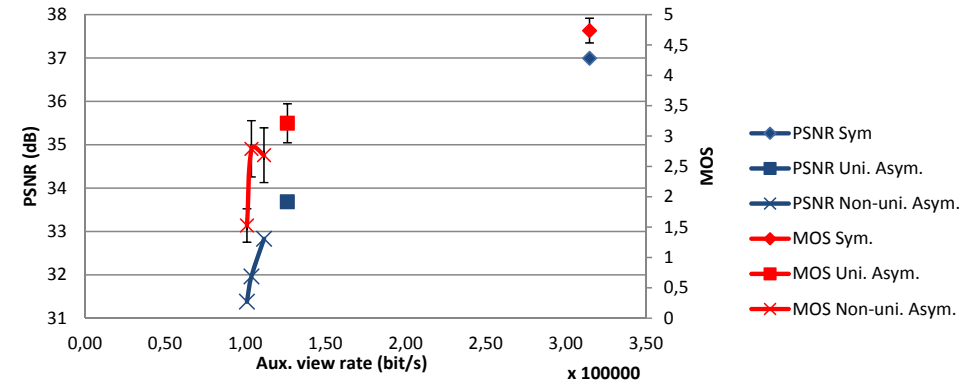
Figure F.1: Discriminated subjective scores per sequence, impairment and observer.

BALLOONS																				
Total Rate	Total PSNR	Rate/s Total		View ID	QP	Rate	SNR Y	Rate/s		View ID	QP_Mask	QP_View	Rate	SNR Y	Rate/s		MOS	Desvio Padrão	% MBs Mask	% rate save aux view	ΔQP mask
10524216	37,053	877018,00		0	35	6746600	37,112	562216,67		1	35	35	3777616	36,993	314801,33		4,736842	0,452413928	25,18		
8607176	35,5695	717264,67		0	35	7092008	37,456	591000,67		1	43	43	1515168	33,683	126264,00		3,210526	0,713282504	25,18	0,00%	0
8430248	35,1435	702520,67		0	35	7090248	37,457	590854,00		1	43	46	1340000	32,83	111666,67		2,684211	1,002919714	25,18	13,07%	3
8327064	34,7075	693922,00		0	35	7081432	37,453	590119,33		1	43	49	1245632	31,962	103802,67		2,789474	1,031662486	25,18	21,64%	6
8302200	34,4195	691850,00		0	35	7090048	37,458	590837,33		1	43	51	1212152	31,381	101012,67		1,526316	0,61177529	25,18	25,00%	8

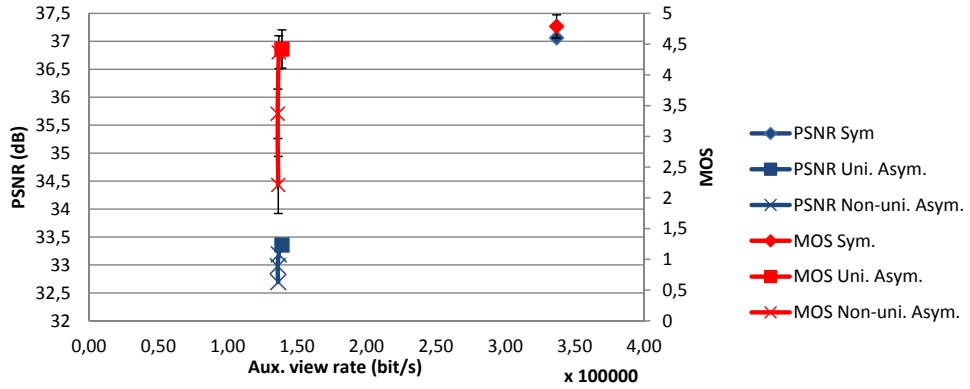
CHAMPAGNE TOWER																				
Total Rate	Total PSNR	Rate/s Total		View ID	QP	Rate	SNR Y	Rate/s		View ID	QP_Mask	QP_View	Rate	SNR Y	Rate/s		MOS	Desvio Padrão	% MBs Mask	% rate save aux view	ΔQP mask
10547472	36,951	878956,00		0	35	6505088	36,84	542090,67		1	35	35	4042384	37,062	336865,33		4,789474	0,418853908	39,96		
8408904	35,185	700742,00		0	35	6741184	37,012	561765,33		1	42	42	1667720	33,358	138976,67		4,421053	0,692482609	39,96	0,00%	0
8374864	35,1015	697905,33		0	35	6734296	37,012	561191,33		1	42	45	1640568	33,191	136714,00		4,368421	0,59726472	39,96	1,66%	3
8371336	34,9955	697611,33		0	35	6739936	37,013	561661,33		1	42	48	1631400	32,978	135950,00		3,368421	0,895080773	39,96	2,23%	6
8377752	34,8535	698146,00		0	35	6742040	37,014	561836,67		1	42	51	1635712	32,693	136309,33		2,210526	1,031662486	39,96	1,96%	9

KENDO																				
Total Rate	Total PSNR	Rate/s Total		View ID	QP	Rate	SNR Y	Rate/s		View ID	QP_Mask	QP_View	Rate	SNR Y	Rate/s		MOS	Desvio Padrão	% MBs Mask	% rate save aux view	ΔQP mask
7953432	36,983	662786,00		0	37	5011288	37,022	417607,33		1	35	36	2942144	36,945	245178,67		4,736842	0,561951487	18,67		
6729368	35,488	560780,67		0	37	5384440	37,401	448703,33		1	45	45	1344928	33,575	112077,33		3,631579	0,830697586	18,67	0,00%	0
6657120	35,2195	554760,00		0	37	5380160	37,393	448346,67		1	45	47	1276960	33,046	106413,33		3,842105	0,834210065	18,67	5,32%	2
6621616	34,933	551801,33		0	37	5379952	37,396	448329,33		1	45	49	1241664	32,47	103472,00		2,684211	1,249561327	18,67	8,32%	4
6596296	34,6265	549691,33		0	37	5382624	37,402	448552,00		1	45	51	1213672	31,851	101139,33		2,842105	0,898341552	18,67	10,81%	6

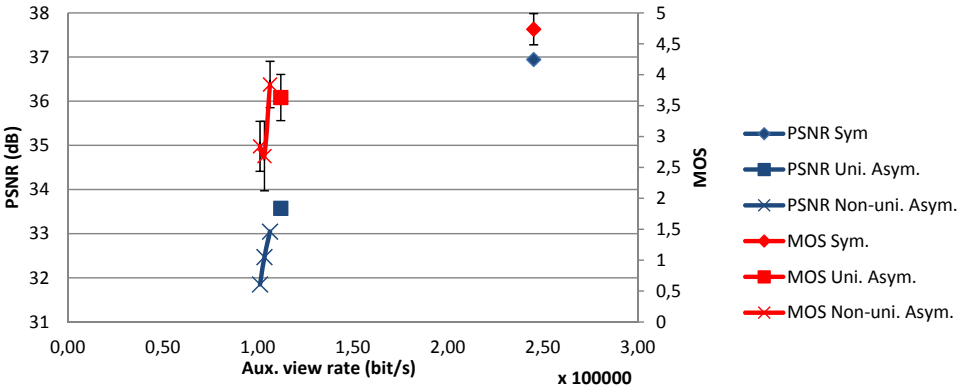
Figure F.2: Complimentary data from stereoscopic video subjective testing.



(a)

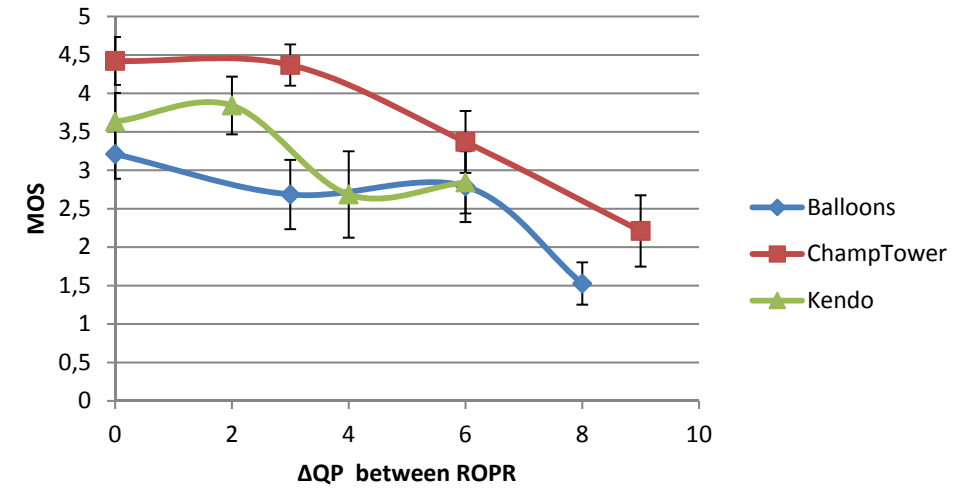


(b)

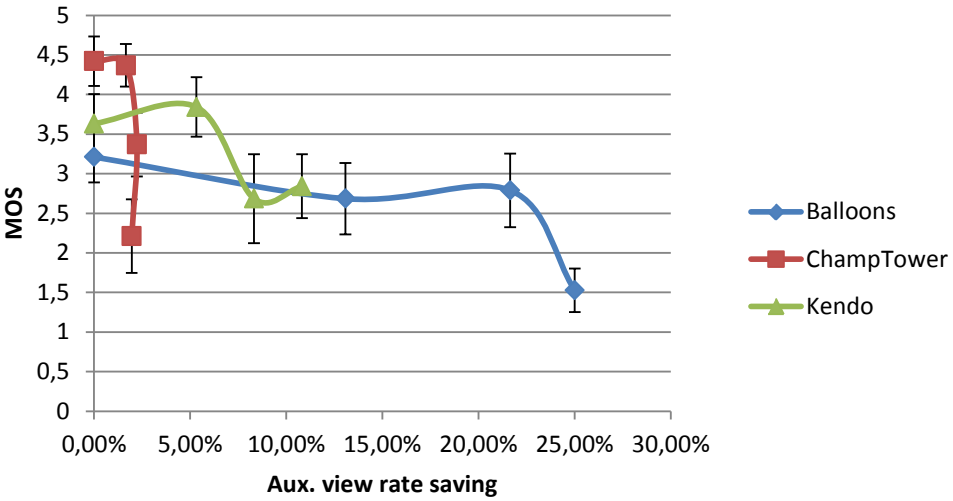


(c)

Figure F.3: R-D and MOS results for the individual videos Balloons (a), Champagne Tower (b) and Kendo (c).



(a)



(b)

Figure F.4: MOS as function of: ΔQP between regions of perceptual relevance (a) and rate saving in the auxiliary view (b).

Appendix G

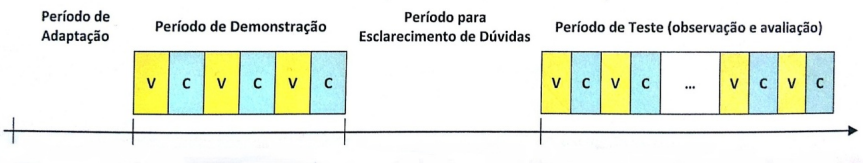
Subjective tests observer's instructions and rating sheets

Testes de Avaliação Subjectiva de Imagens 3D

1. Obrigado pela disponibilidade para efectuar testes subjectivos de imagem 3D!
2. Nestes testes você irá ver diversas imagens 3D com duração de 10-12s cada.
3. No final da visualização de cada par de imagens **deverá avaliar a qualidade da 2ª imagem em comparação com a 1ª**.
4. Deve observar atentamente cada imagem de modo a avaliar a qualidade, atribuindo uma classificação dentro da seguinte escala, que deve reflectir a sua opinião sobre as diferenças que verificar entre as duas imagens.

   5 - Imperceptível
   4 - Perceptível
   3 - Ligeiramente irritante/perturbadora
   2 - Irritante/perturbadora
   1 - Muito irritante/perturbadora

5. Após a visualização de cada par de imagens, das quais a primeira é a referência e a segunda a imagem sob teste, deverá decidir sobre a classificação a atribuir dentro de 10s. A exibição de cada par **NÃO** será repetida antes da avaliação.
6. A classificação é atribuída colocando uma cruz na ficha de avaliação, no campo correspondente à classificação que quer atribuir.
7. A sessão de avaliação é constituída pelas seguintes fases da figura seguinte, sendo que em cada visualização verá uma imagem de boa qualidade (referência) seguida da imagem a avaliar.



8. No período de adaptação serão apresentadas diversas imagens 3D, incluindo imagens das sequências que irão ser avaliadas. Este período tem por objectivo familiarizar o seu sistema de percepção 3D com a tecnologia e os conteúdos que lhe irão ser apresentados.
9. Segue-se um período de demonstração onde serão visualizados exemplos de imagens com o tipo de erros cujo impacto perceptual se pretende avaliar (tempo: V). Após cada visualização deverá ser efectuada a respectiva classificação (tempo: C).
10. O período de esclarecimento de dúvidas serve para colocar todas as questões que considerar pertinentes.
11. O período de teste consiste em visualizar diversas imagens (V) seguidas de avaliação e classificação (C). Este período terá uma duração máxima de 30 min.

Obrigado pela sua colaboração!

Projecto AV3DBMP (IPLeiria/IT), Testes de avaliação subjectiva - informação para observadores, Fevereiro 2013

Figure G.1: Instructions provided to the observers during the still images subjective assessment sessions.

Avaliação Subjectiva de Vídeo 3D

Ficha de Avaliação

Sequências para Avaliação Subjectiva

Classifica. -> Núm.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45
Imperceptível (5)																																													
Perceptível (4)																																													
Lig. Pertubador (3)																																													
Pertubador (2)																																													
Muito Pertubador (1)																																													

Nome: _____ Sexo: F _____ M _____ Idade: _____ Olho dominante: E ☐ D ☐ N ☐

Projecto AV3DBMP - Fevereiro 2013

instituto de telecomunicações

Figure G.2: Observer's rating sheet for the still images subjective assessment.