An Architecture for 3D multi-view video transmission based on Dynamic Adaptive Streaming over HTTP (DASH)

Tianyu Su

Thesis submitted to the
Faculty of Graduate and Postdoctoral Studies
In partial fulfillment of the requirements
For the M.A.Sc. Degree in
Electrical and Computer Engineering

School of Electrical Engineering and Computer Science
Faculty of Engineering
University of Ottawa

Acknowledgements

I am highly indebted to my teacher and supervisor, Professor Shervin Shirmohammadi, for his patience, intelligent advice, encouragement during different stages of my graduate study. I thank my supervisor for providing me with the devices, facilities and, last but not least, financial support throughout my study. Dr. Shervin Shirmohammadi gives me an extraordinary opportunity to research in the emerging, prosperous and useful field of Multimedia Communication. It is my honor and lucky to work with him.

Special thanks to Dr. Abdulsalam Yassine, Postdoctoral Fellow at the Discover lab, for his invaluable help, continuous support and careful revisions at various stages of writing this thesis. And thanks to Dr. Abbas Javadtalab, now research engineer at Ericsson, for his kindly help and brilliant guidance during the first year of my study.

Also I greatly acknowledge the assistance of Ashkan Sobhani, Ph.D candidate of University of Ottawa for his continuously help to clarify my understanding of key concepts of DASH and valuable suggestions during the implementation phase.

I would like to thank the members of the DISCOVER Laboratory, for their suggestions, critical remarks, and cooperation during my research work.

A special word of gratitude should be dedicated to my fiancé, Qi Zhang, for her support and accompany.

Last but not the least, I thank to my parents, I would not be able to study in University of Ottawa without their supports, encouragement, and endless love.

Dedication

I dedicate this research to my grandmother, who raised me up and saw my growth. I hope her health whenever and wherever.

Table of Contents

List of Figures

List of Tables

Glossary of Terms

HEVC	High Efficiency Video Coding
MVV	Multi-View Video
DASH	Dynamic Adaptive Streaming over the HTTP
MVD	Multi-View plus Depth Map
AVC	Advanced Video Coding
FVV	Free Viewpoint Video
SSIM	Structural Similarity
3D	Three Dimensional
CTU	Coding Tree Unit
LCU	Large Coding Unit
CU	Coding Unit
PU	Prediction Unit
MPD	Media Presentation Description
DIBR	Depth Image Based Rendering
PSNR	Peak Signal to Noise Ratio
VBR	Variable Bit Rate
CBR	Constant Bit Rate
QP	Quantization Parameter
URL	Uniform Resource Locator
XML	Extensible Markup Language
HTTP	The Hypertext Transfer Protocol
UDP	User Datagram Protocol
TCP	Transmission Control Protocol
RTP	Real Time Transport Protocol
SVC	Scalable Video Coding

LCU	Large Coding Unit
GOP	Group of Pictures
MPEG	Motion Picture Experts Group
ITU	United Nations specialized agency for information and communication
3GPP	3rd Generation Partnership Project
SMVC	Scalable Multiview Video Coding
MSE	Mean Square Error
SAO	Sample Adaptive Offset
FPGA	Field-Programmable Gate Array
CUDA	after the Plymouth BarraCUDA
SDSCE	Simultaneous Double Stimulus for Continuous Evaluation
MOS	Mean Opinion Scores

Abstract

Recent advancement in cameras and image processing technology has generated a paradigm shift from traditional 2D and 3D video to Multi-view Video (MVV) technology, while at the same time improving video quality and compression through standards such as High Efficiency video Coding (HEVC). In multi-view, cameras are placed in predetermined positions to capture the video from various views. Delivering such views with high quality over the Internet is a challenging prospect, as MVV traffic is several times larger than traditional video since it consists of multiple video sequences each captured from a different angle, requiring more bandwidth than single view video to transmit MVV. Also, the Internet is known to be prone to packet loss, delay, and bandwidth variation, which adversely affects MVV transmission. Another challenge is that end users' devices have different capabilities in terms of computing power, display, and access link capacity, requiring MVV to be adapted to each user's context. In this paper, we propose an HEVC Multi-View system using Dynamic Adaptive Streaming over HTTP (DASH) to overcome the above mentioned challenges. Our system uses an adaptive mechanism to adjust the video bitrate to the variations of bandwidth in best effort networks. We also propose a novel scalable way for the Multi-view video and Depth (MVD) content for 3D video in terms of the number of transmitted views. Our objective measurements show that our method of transmitting MVV content can maximize the perceptual quality of virtual views after the rendering and hence increase the user's quality of experience.

Chapter 1: Introduction

1.1 3D multi-view video

Traditional 3D video representation is usually achieved with two cameras. Users can observe the 3D scene by wearing shutter goggles or polarized goggles [1]. Although, stereoscopic immersive 3D video is popular both in theaters and in home entertainment, the flexibility for the users is low [1, 6] due to the fixed conditions in which the stereo content is captured as can be shown in Figure1 (a) Recent advancement in camera and image processing technology has generated a paradigm shift from traditional 2D and 3D video to multi-view video technology. In multi-view video (MVV), cameras are placed in predetermined positions and angles to capture the video sequences. The video sequences of MVV are fixed [2]. MVV content is then compressed and transmitted in a suitable way so that the users' viewing device can easily access the relevant views to interpolate new views. Multi-view representation allows the users to freely change their viewpoints [2].

Currently, Multi-view Video Coding (MVC) for MVV or Free Viewpoint Video (FVV) is standardized as an extension of H.264/MPEG-4 Advanced Video Coding (AVC). This technology is used in several applications such as stereoscopic 3D video (temporal interleaving or spatial interleaving), FVV, and auto-stereoscopic 3D video [2]. Furthermore, the Free Viewpoint Television (FTV) is based on this concept, which allows the users to interactively control the viewpoint and generate new views of a dynamic scene from any 3D position [3] as can be seen from Figure1 (b). This means providing a realistic feeling of natural interaction to the users. A more recent Multi View plus Depth (MVD) format [4, 5] encodes a depth signal for each camera, as illustrated in Figure 1(c). The main advantage of MVD is that it allows synthesizing virtual views at

the client via the depth signals. These signals are sent along MVC by means of Depth Image Based Rendering (DIBR) that can render several virtual views based on the few real views and their associate depth map [4]. The DIBR method can convert transmitted multi-views to more virtual views for auto stereoscopic 3D display in order to efficiently represent the 3D views to transmit. The popularity of this technology has gained attention from researchers and practitioners to develop innovative 3D techniques to meet new demands.

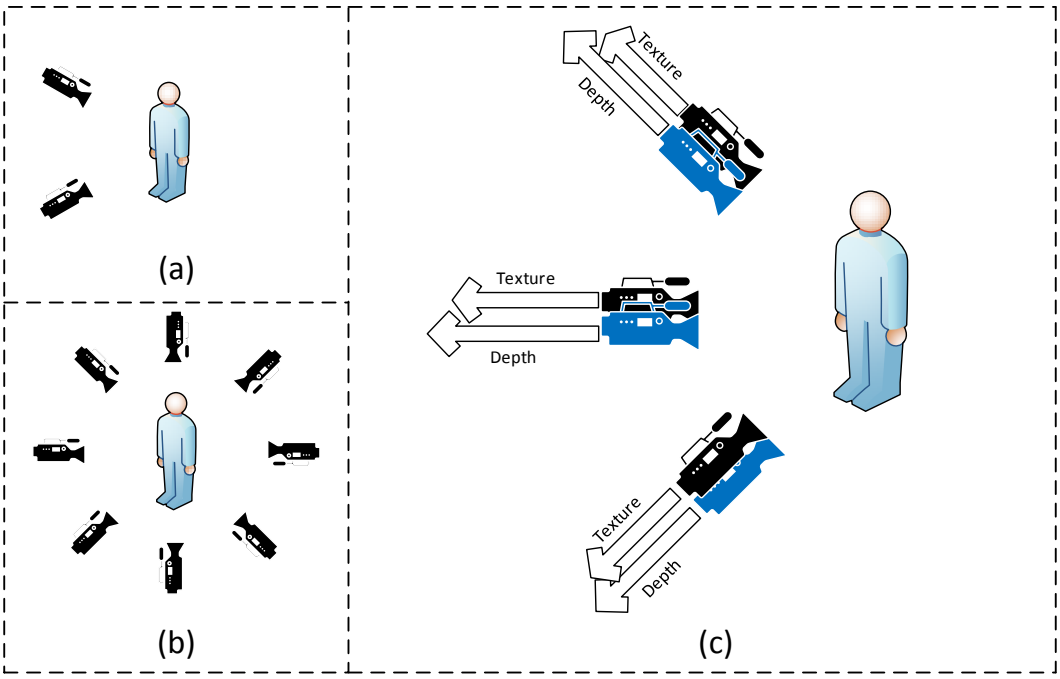


Figure 1Different camera capture ways.

(a) Stereo capture. (b) Freeviewpoint capture. (c) Multiview plus depth capture.

1.2 Associated Problems

While MVV and MVD technologies are very promising, the development of appropriate mechanisms that support the delivery of these technologies to the end user over best

effort networks is not progressing at the same speed [6]. There are three major challenges encumbering the delivery of MVV over present networks [6]:

First, MVV traffic is several times larger than traditional multimedia since it consists of the video sequences captured by multiple cameras, which means more bandwidth is required to transmit. Also, the Internet is known to be prone to packet loss, delay, and bandwidth variation, which is rather substantial in the case of MVV content. Second, end users devices have different capabilities in terms of computing power, display, and access link capacity, which requires adaptive mechanisms to adjust for the variations of bandwidth while traversing the network's paths.

Second, the trend of video development is to increase the resolution to let the view enjoy more details of the picture. However, the higher resolution means the much higher content to be delivery. Especially for the 3D multi-view display, more than two High Definition views are needed to be transmitted. Therefore, a more efficiency encoder should be considered.

Third, the speed of codec for 3D is too slow, that make the real time compress the content of 3D multiview video impossible. The 3D multi-view video still can be delivery via video on demand.

1.2 Motivation

The above challenges motivated us to propose a dynamic rate adaptation system and its associated rate-distortion model for multi-view 3D video transmission, which will address the issue of varying network bandwidth for Internet consumers of multi-view video. Our rate adaptation system is built on top of two state-of-the-art key technologies: High Efficiency Video coding (HEVC) [7, 8], and MPEG's Dynamic

Adaptive Streaming over HTTP (DASH) [11, 12]. HEVC, introduced by the ITU-T Video Coding Experts Group and the ISO/IEC Moving Picture Experts Group, provides about 50 percent bit rate reduction at the same video quality compared to H.264 [9]. Furthermore, HEVC 3D extension [10] is flexible in generating a bit stream format that is suitable for different setups: from traditional 2D to stereo 3D and Multi-view video. MPEG-DASH [11,12] divides the video contents into segments with equal length and stores them in a server. The copies of the video segments are encoded with different bit rates that represent different qualities and resolutions, as shown in Figure 2.

In MPEG-DASH, all segments that are stored in the server can be accessed via an XML-based Media Presentation Description (MPD) file. All streaming sessions are managed by the clients; i.e., a client chooses a bit rate by requesting a specific Media Presentation Description (MPD) based on the network condition and its decoding process. Our system uses an adaptive mechanism to adjust for the variations of bandwidth while traversing the paths of best effort networks

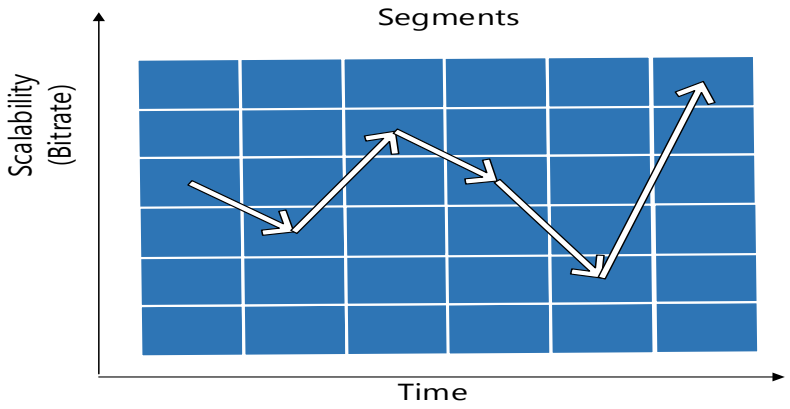


Figure 2 Scalability for video content in DASH

We will propose a novel scalable way for the Multi-view video and Depth (MVD) content for 3D video in terms of the number of transmitted views. Our Structural

Similarity (SSIM) tests show that our method of transmitting multi-view video content can maximize the perceptual quality of virtual views after the rendering and hence increase the user's quality of experience.

1.3 Research Objectives

The objective of this research is to

- Proposed an architecture of transmission system upon two state of art techniques: the High Efficiency Video Coding (HEVC) and Dynamic Adaptive Streaming over the HTTP (DASH) for the multi-view video transmission.
- Approach the effect of bitrate distribution on the final quality of multi-view video, and based on the proposed system, has a solution to deliver the maximum possible quality of multi-view video over the internet.
- Proposed a rate control system based on the architecture to adapt the scalable 3D video content over the network and have some simulation on it.
- Come up with a bandwidth prediction algorithm based on the approach above to predict the available bandwidth of the transmission. Emulate the algorithm, and analyze the emulation result on it.

1.6 Contribution

This work introduces novel ideas and concepts that are not considered by existing work, including:

- A dynamic rate adaptation system for multi-view 3D video transmission is proposed and its associate rate distortion model is established. The proposed

system addresses the issue of varying network bandwidth for Internet consumers of multi-view video. Moreover, the lower flexible nature of HEVC 3D codec system is also addressed by using the proposed smoothed bitstream selection algorithm on the Dynamic Adaptive Streaming over the HTTP (DASH). Not only the architecture is proposed, but also implementation and simulation is finished based on the proposed system.

- On the server side, in order to maximize the QoE for users, a new adaptive method is introduced to the proposed system by switching between different numbers of transmitted views on DASH. The new adaptive method is based on our experiments and investigation on the effect of different number of transmitted views and the baseline distance on the perceived quality of the final rendered result. These results are evaluated using SSIM and PSNR objective tests. The proposed adaptive way of streaming the MVD content extent current work by selecting the different video bitrate by changing the resolution of the views, the perceptual quality, and the frame rate of the temporal video.
- On the client side, a switching strategy corresponding to the adaptive transmitted views selection made by the server is applied. The strategy is based on Moving Average Smoothed selection algorithm according to which the client can adaptively download the multiview-3D content.

1.7 Publications

The research work has resulted in one paper published in conference proceeding, and one journal submitted with respect to the research motivations and contributions as described above. All papers are related to the topic of this thesis. A publication list of

research work are listed below during my Master of Applied Science program:

Conference paper

[1] **T. Su**, A. Javadtalab, A. Yassine and S. Shirmohammadi, "A DASH-Based 3D Multi-view Video Rate Control System," in IEEE international Conference on Signal Processing and Communication System, 2014.

Journal paper:

[1] **T. Su**, A. Sobhani, A. Yassine, S. Shirmohammadi and A.Javadtalab, "A DASH based 3D multview video transmission system," Journal of Real Time Image Processing. Special issue on Architectures and Algorithms of High Efficiency Video Coding (HEVC)standard for Real-Time Video Applications, 2015.

1.8 Organization of the Thesis

The rest of this thesis is organized as follows:

Chapter 1 The introduction of this thesis is proposed as well as the motivation, related problems, the goal of this research, the contributions of this thesis and the scholar achievements are highlighted in Chapter 1.

Chapter 2 Fundamentals concepts associated with this research along with the reason why use the High Efficiency Video Coding (HEVC) video format and Dynamic

Adaptive Streaming over HTTP (DASH) as transmission protocol are explained in Chapter 2. Literature Review covering 2D video adaptation, Bitrate adaptation for HTTP streaming, Dynamic Adaptive Streaming Over the HTTP for Free Viewpoint Video Streaming and Stereo 3D Streaming, Subjective Approach for the Multi-view Plus Depth content are highlighted in Chapter 2.

Chapter 3 Design of the proposed architecture of the 3D multi-view video transmission system based on Dynamic Adaptive Streaming over the HTTP (DASH) is presented in Chapter 3.

Chapter 4 The emulation of the proposed system and a performance evaluation of this system is introduced in Chapter 4.

Chapter 5 In chapter 5, the conclusion and future work which is indicating a direction for future is presented in Chapter 5.

Chapter 2: Background and related work

In this chapter, the background of the thesis is introduced first. The background includes five main parts:

1. 3D multi-view video: The stereo vision and depth map creation are explained first in section 2.1.1. Then, the development of the 3D display is introduced in sections 2.1.2 and 2.1.3. Next, the auto-stereoscopic 3D display for multi-view video is introduced. An analysis of the problems with content delivery is introduced and compared in section 2.1.4. Finally, a review of the techniques of virtual view generation, based on the multi-view plus depth format, is presented in section 2.1.6.
2. High Efficiency Video Coding (HEVC): An overall introduction of the emerging HEVC video coding standard is presented in section 2.2.1, and an explanation of the rate control technique in this new standard is introduced and explained in section 2.2.2.
3. Dynamic Adaptive Streaming over the HTTP (DASH): The reason for choosing HTTP is introduced in section 2.3.1, previous types of video progressive download are explained in section 2.3.2, and a summary of DASH concepts is presented in section 2.3.3
4. A methodology for the video evaluation metrics is introduced in section 2.4. This methodology includes two parts: subjective test metrics and objective test metrics.
5. The scalability of the video content is summarized in section 2.5.

After the background sections, related work is introduced and analyzed in sections 2.7, 2.8, and 2.9.

2.1 Multi-view plus depth

2.1.1 Stereo vision and depth map creation

Traditional stereoscopic 3D extends the visual experience from one 2D view to two 2D views in order to enable the left and right eyes to receive two slightly different view points, thereby simulating the natural eye system of human beings.

In binocular vision, the disparity between the two stereo pictures projected onto the human brain creates an illusion of depth within the human mind. Although some sensors can gather the depth information for vision views directly, researchers also seek to extract the depth information from two binocular images or a multi-view system in order to represent stereo 3D images or video efficiently. The reason for this focus is the low cost and higher resolutions of this method, in comparison to actively extracting depth information from a sensor (Kinect).

Computer vision technology allows one of two stereo views to be reconstructed from the other stereo view and depth information. There are three main steps to creating a depth map using stereo vision: epipolar rectification (calibration), disparity matching, and depth calculation [44].

Epipolar rectification calibrates the stereo views parallel to one of the image's axes (the y axis in Figure 3 in two dimensions. It has the advantage of reducing the disparity matching process to a one-dimensional search along the horizontal raster of the calibrated images [44].

The disparity matching involves seeking the difference in each stereo vision projected from the same region in the 3D space. The calculation of the depth information is shown in Figure 3. P is the object in the 3D space, and P_L and P_r are the projected coordinates of the stereo view from the 3D space. Since the rectification has

been finished before this step, we can assume the same y values for P_L and P_r. The baseline stands for the distance of two stereo views (eyes or cameras), and the focal is the focal length of the eyes or the camera. All of this information is known.

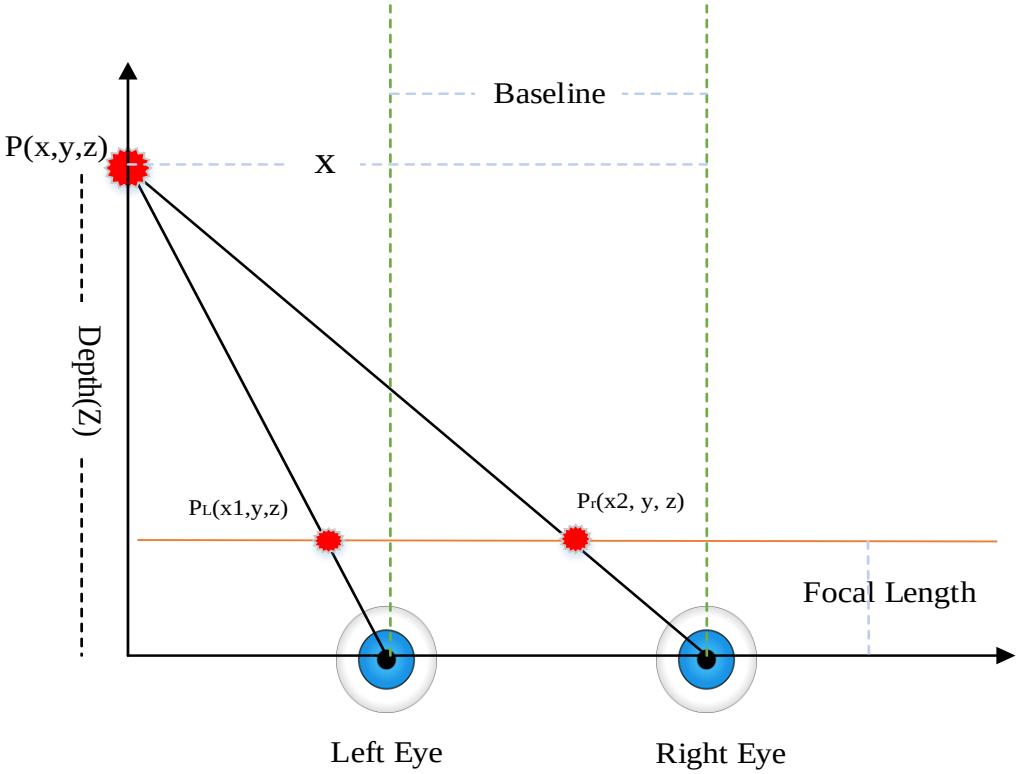


Figure 3 Calculation of depth from stereo views

From the Similar Triangulation theorem,

$$\frac{X - Baseline}{Depth} = \frac{X - Baseline - X}{Focal} \tag{2.1}$$

$$\frac{X}{Depth} = \frac{X - X_2}{Focal} \tag{2.2}$$

After combining (2.1) and (2.2), we get

$$Depth = \frac{Baseline \times Focal}{X - X}$$

$$Depth = \frac{Baseline \times Focal}{d} \tag{2.4}$$

where d is the disparity between the two coordinates from the two stereo views.

2.1.2 Glass wear 3D display system

There are two main glass wear display systems: Active and Passive, or the shutter system and the polarization system, respectively.

- In the shutter system (time multiplexed), the main idea is that each time only one of the two stereo images is present, blocking the other, each glass for one eye is only applied one time.
- In a polarization system, two stereo frames are projected onto one display through polarizing filters. Each polarizing filter can change the direction of the light..

2.1.3 Naked eye 3D with single view

Although stereoscopic glass wear 3D displays are widely established, they have some disadvantages [45]:

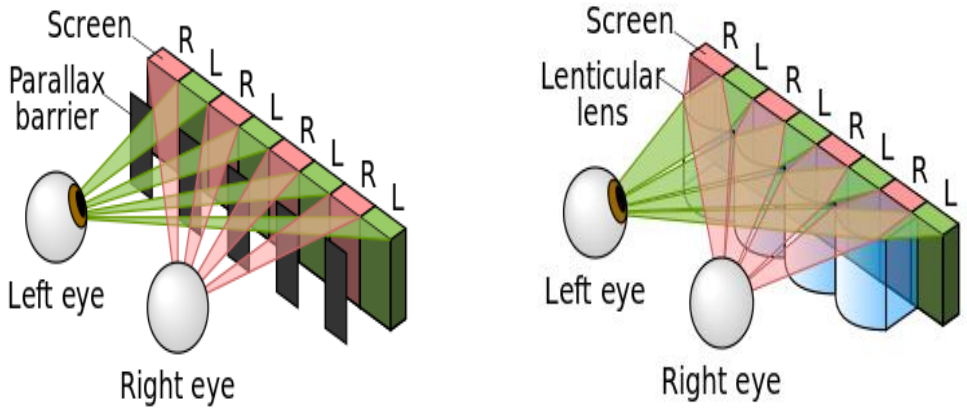
- Visual discomfort is a significant problem [45].
- The viewer can only see the stereo vision within a limited angle.
- The cumbersome nature of the glasses results in a low quality of experience.

In order to display a stereoscopic image without requiring users to wear specific glasses, auto-stereoscopic 3D was established. Auto-stereoscopic 3D solves the existing problems caused by stereo 3D. As can be seen in Figure 4, there are two types of auto-stereoscopic displays: parallax barrier-based and lenticular array-based displays.

- Parallax barrier-based: A parallax barrier is placed in front of the display to

ensure that each eye sees a different area of the screen. This allows the viewer to feel the depth caused by the binocular effect [46]. The structure can be seen in Figure 4.

- Lenticular array-based: This is similar to the parallax barrier, in which a lenticular array is placed in front of the display to ensure that the user can receive stereo pictures from different areas of the display [46]. The structure can be seen in Figure b. It worthwhile to note that lenticular array-based 3D displays are more expensive than parallax barrier-based.



(a) Parallax barrier-based (b) Lenticular based

Figure 4. Autostereo display for one view [47].

2.1.4 Multi-view 3D display

Despite the problems solved by the naked eye 3D with a single view display, one problem still exists: the limitation of the view angle. Thus, the multi-view 3D display was introduced in order to develop a new 3D display [48].

In a multi-view 3D display, there are lenses behind the display that are focalized in such a way that, when one is watching the screen, a different sub-pixel can be seen on

each fixed baseline through each lens. We assume that human eyes are separated by the distance of the baseline (normally 5 cm), so that each eye can see a different image at all times. Moreover, the 3D auto-stereoscopic display is a combination of several sub-pixels of different colors (YUV or RGB); therefore, the 3D stereo vision can be seen from different viewpoints via the technique of focalized lenses, as can be seen in Figure 5.

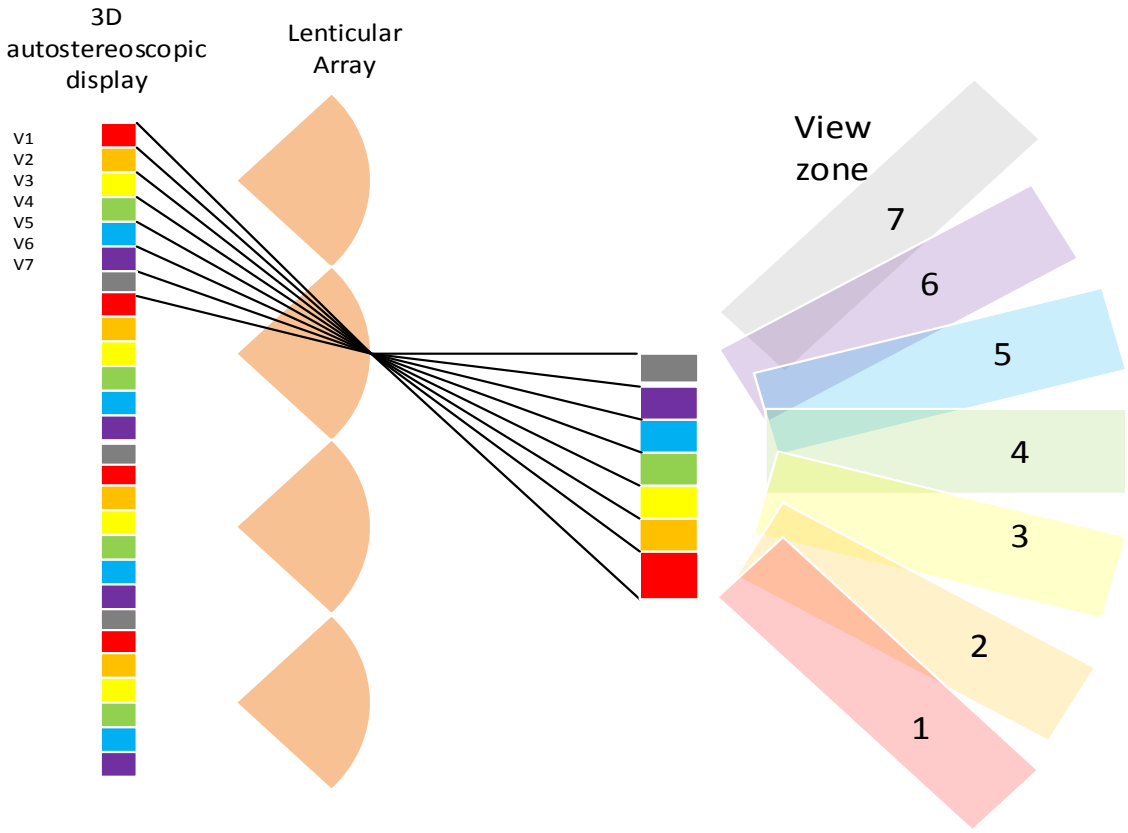


Figure 5. Lenticular-based multi-view 3D display

The advantages of the 3D multi-view display over the stereo 3D glass wear and auto-stereo one-view display are the following:

- It provides the viewer with a three-dimensional perspective without specialized eye glasses.
- There is no restriction of viewpoint. The 3D multi-view display offers a multitude of views that enables the user to view 3D at different viewpoints.

2.1.5 Multi-view plus depth format for 3D content

In order to support multi-view displays, multiple simultaneous views (at least eight views) of a scene are needed. However, it is difficult and very expensive to capture so many views. Moreover, in terms of transmission, the delivery of a large number of views is impossible over a non-dedicated network.

Since a multitude of views can be synthesized from input views with the help of texture views and their associated depth maps, there are several solutions for different types of 3D displays. In order to solve the challenges related to delivering a huge number of views over the Internet, the multiple views plus depth format for 3D content has become increasingly popular. A comparison of the different formats of 3D content is shown in Figure 6. Although the traditional stereo 3D is much easier to deliver in the network due to its low data request size, the glass wear will result in several quality of experience problems. Specifically, the discomfort caused by the lack of depth adaptations for different people, the fixed viewpoint of the 3D stereo display, and the cumbersomeness of the glass wear will restrict the development of the 3D glass wear display. The multi-view 3D display is proposed to handle the problem the occurs in the 3D stereo display; however, the problem appears during the delivery of multiple views in the non-dedicated network. The multi-view plus depth map format of the 3D content arouses more attention from both industry and research institutes around the world.

The conclusion is that, among all 3D contents, the multi-view plus depth map is the most efficient format for delivering 3D multi-view video content on non-dedicated networks. The multiple views within the fixed baseline distance can be generated from a limited number of views plus associated geometry information.

Type	Basic info	Advantages	Weakness
Stereoscopic 3D	Views for each eyes jointly or compatibly(side by side or up and down)	Very similar with existing 2D transmission systems	Glass wear and cannot adapt depth for different users
View plus depth	2 stereo views can be rendered from one view and its corresponding depth	Compared with stereo 3D, more efficiency regarding to the compression.	Still glass wear uncomfortable and a single view can render can result in some occlude area.
Multi-view 3D	Multi-views are captured from multiple angle in an array	Support for the auto-stereoscopic display without the glass wear and more views angles and depth perception for users	Huge bandwidth requirement due to the large content size in this format of multi-view 3D.
Multi-view plus depth	Multi-view from different angles plus their associated depth	Support for non- glass wear display and less views are needed	Much less bandwidth is needed compared with multi-view 3D display

Figure 6. Comparison of different formats for 3D content， adopted from [1]

As is shown in Figure 7, in order to provide the 3D multi-view display with at least nine views and a fixed baseline distance (from the leftmost camera to the rightmost camera), the MVD format can be used to render the six virtual views inside the baseline. The details of the rendering technique will be introduced in Section 2.1.6.

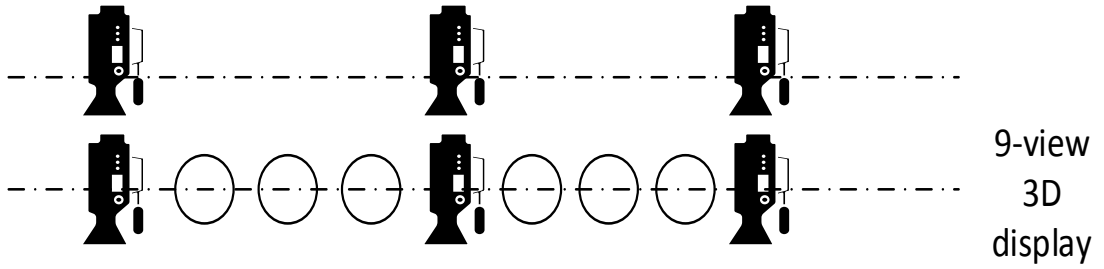


Figure 7. View generation from a three-streamed view scenario for a nine-view 3D display

2.1.6 View synthesis technique based on the MVD format

In 2.1.1, the distance related to the camera position and the scene geometry information can be considered the depth data. Moreover, we can collect depth data by finding the disparity between two cameras. The projective geometry introduced in [49] suggests that the virtual views between input views can be synthesized via three-dimensional projection theorems or two-dimensional warping. This is the so-called Depth Image-Based Rendering (DIBR).

Because the cameras capture one scene from different viewpoints, the partial occlusion of the foreground is one of the weaknesses of the view synthesis technique [47]. Therefore, data from at least two viewpoints, plus depth information, are required to tackle the problem of invisible data caused by an occlusion in one viewpoint that is visible in another viewpoint. This is accomplished by filling the visible data from the non-occluded area into another viewpoint (i.e., the reference view). Since the cameras are calibrated when the depth map is acquired, the process of depth-based image rendering can be simplified, since the horizontal-based pixel shifts from the original view (i.e., the reference view) to the synthesis view [47]. Assume that there are two reference views (view 1 and view 2) and two corresponding pixels (P1 and P2,

respectively). The corresponding pixel in virtual view p_k is in the intermediate position between P1 and P2, where k is between 0 and 1 (k represents the position parameter). Thus, the simplified view synthesis technique can be shown below:

$$p_k(x_k, y_k) = (1 - \alpha)p_1(x_1, y_1) + \alpha p_2(x_2, y_2) \quad (2.5)$$

From the above formula, we can see that the rendering process can be completed based on the view information; however, in reality, most image-based rendering techniques require depth map information for the following reasons:

- The disparity between the information associated with each view can result in additional distance from the original view at a good rendering quality, which makes a longer baseline for multi-view video possible [49].
- On the client side, if the depth map or the disparity information is provided, the time-consuming process of disparity matching of the two reference views can be avoided. This is very important for alleviating the burden on the client side, since the computational burden for decoding a multi-view plus depth map is already very high.

The basic processes of the depth image-based rendering are as follows (see Figure 8).

- Classification: The whole frame can be classified into two classes: the unreliable area and the reliable area. The unreliable area refers to those textured areas along the edge of the depth [49], while the remaining areas are the reliable areas.
- Projection and blending: The projection rules for different areas in each frame are different. The pixels in the reliable area can be shifted and blended into the synthesis view, based on formula 2.5. The unreliable areas can be divided into two areas: the foreground and the background [49]. The pixels in the foreground areas are blended with those in the reliable area, while the background area is filled with the uncovered area [49].
- Synthesized view enhancement: Due to certain problems that occur after

blending, such as blurred edges and holes in the picture, specific filters need to be implemented.

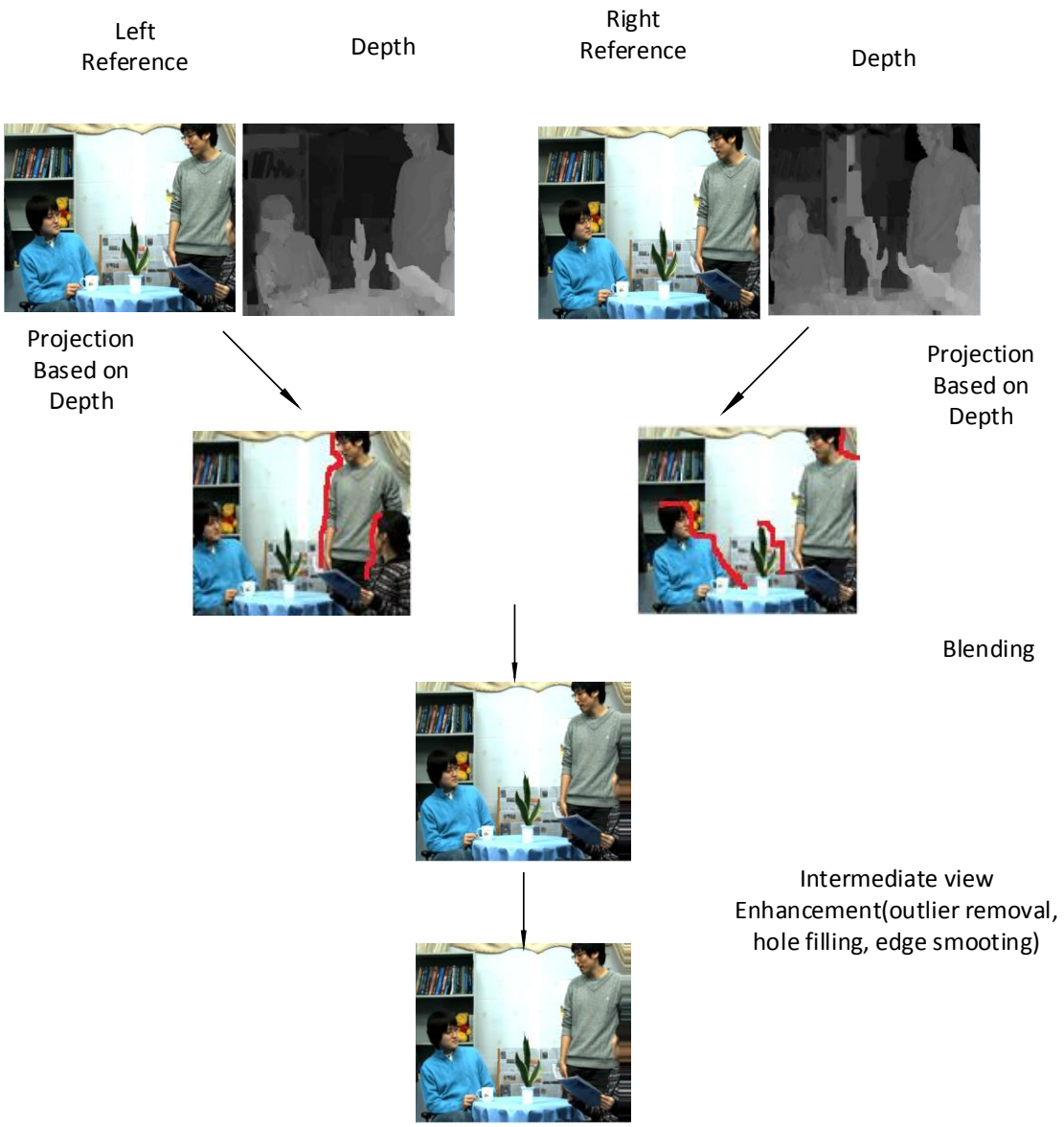


Figure 8 Steps for Depth Image-based Rendering

2.2 High Efficiency Video Coding (HEVC)

2.2.1 Introduction of HEVC

High Efficiency Video Coding (HEVC) [26], introduced by the ITU-T Video Coding Experts Group and the ISO/IEC Moving Picture Experts Group, is a hybrid video codec design. It aims at maintaining a constant perceptual quality level while reducing the bitrate by 50 percent by increasing the computational complexity. Compared to H.264/AVC [27] and MPEG-2 [28], HEVC offers the following improvements:

- Hierarchical Quad Tree Structure: HEVC offers a quad-tree coding structure within a picture, which includes a Coding Tree Unit (CTU), a smaller Coding Unit (CU) that is split from the CTU, and further Prediction Units (PU) that are split from the CU and that are used for intra- and inter-prediction inside the CU. The Transform Unit is furthered by the PUs, which define the transformation (e.g., DCT) and quantization [26].
- The larger-sized "Macroblock", the Coding Tree Unit, consists of a luma Coding Tree Block (CTB) and Chroma CTBs, which are analogous to the Macroblock in H.264/AVC [7]. The size of the CTB is selected by an encoder that can be larger (64*64) than that of traditional Macroblocks (16*16) introduced in previous standards (e.g., H.264 and MPEG2). The larger size of the CTU offers improved compression performance for high definition or 4K videos.
- Parallelization design: In order to accelerate the speed of the codec in tackling the issue of improving the computation complexity, a parallelization design for HEVC introduces tiles, which are several rectangular parts inside each frame.
- Support for 3D extension: The new HEVC standard not only tackles video compression for 2D high definition video and 4K video, but also supports the views ad associated depth formats of 3D multi-view videos [5].

2.2.2 Rate control in HEVC

The main task of rate control is to enable the codec to adjust discrete coding parameters (e.g., mode, motion, quantization parameter, etc.) to retain the expected bitrate. In particular, the QP value (i.e., how much compression the encoder needs) is the main parameter used for rate control. In H.264/AVC, the QP can be adjusted at both the picture and the macroblock level. When the QP is small, more details can be retained in each frame (and vice versa).

The simplest approach to a rate control algorithm is to start with a predefined QP value in every frame; however, this method of rate control cannot control the specific bitrate that the encoder wants to provide. The disadvantage of this is that it may pose problems when live streaming, due to the nature of instantaneous bit rates, since, under constrained network conditions, the bitrate may far exceed the available bandwidth. Thus, the transmission is totally outside the control for this method of rate control. Due to the weakness of the Variable Bit Rate (VBR), the Constant Bit Rate has been widely adopted, and the rate control algorithm for the CBR rate controller has given rise to wide interest among researchers.

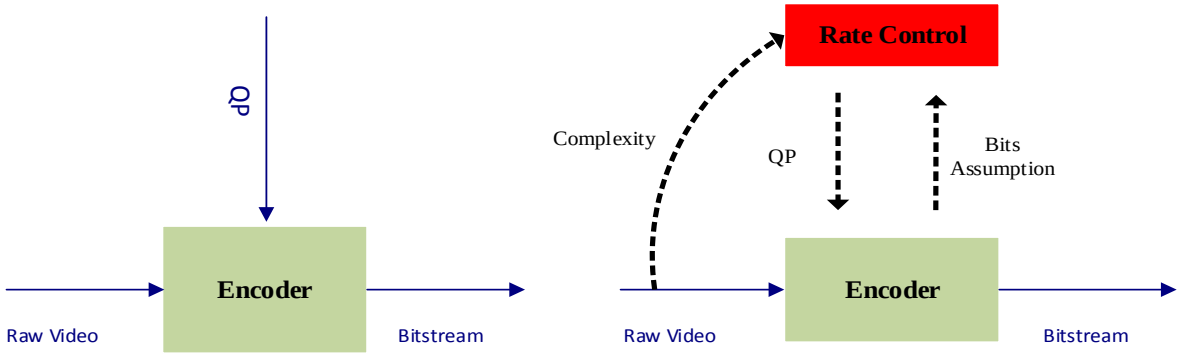


Figure 9 Video Bitrate Control

In H.264, rate-distortion performance is the fundamental consideration. The objective is to achieve a specific bitrate, while minimizing the distortion caused by the lossy compression.

$$D(R) = CR^{-k} \quad (2.6)$$

where C and K are parameters related to different video sources.

$$RD_{cost} = D + \lambda R \quad (2.7)$$

D+λR is the Rate Distortion cost. The rate-distortion-based rate control algorithm searches for the best points in the RD curve around the target bitrate.

The method of rate control used in H.264 attempts to build the relationship between the Quantization Parameter and the different specific bitrates [25], under the assumption that the QP is the only factor that determines the total bitrate.

However, in HEVC, more modes of encoding are available, and its hybrid quad-tree coding structure, makes the assumptions hard to hold. Thus, the R-Lamda model introduce in [33] is adopted, due to the HM8.0 of the HEVC reference software. In [33], Lamda is the slope of the RD model, and it is a more critical value than QP.

$$\lambda = -\frac{\partial D}{\partial R} = CK.R^{-k-1} \quad (2.8)$$

$$\lambda = \alpha R^{\beta} \quad (2.9)$$

where α and β are parameters associated with different video sources.

The bitrate allocation for HEVC can exist in three levels: that is, the Group of Picture (GOP) level, the Picture Level, and Large Coding Unit (LCU) level.

The advantage of the R-lambda model is as follows:

- Control of lamda can be more precise than QP, and lambda can be a positive continuous value, rather than a discontinuous one like QP [33].
- R and Lamda have a one-to-one relationship with one another [33].

- Lambda impacts the non-residue bits [33].
- Lamda does not execute during the rate distortion process. Adjustments to lambda adjust the optimizing target bitrate.

2.3 Dynamic Adaptive Streaming over HTTP (DASH)

2.3.1 Why HTTP

HTTP plays an important role as a protocol for the delivery for video streaming [15, 16]. HTTP is pervasive, and it can pass through all firewalls and Network Address Translators (NATs). Video streaming deployments based on HTTP do not impose considerable costs compared to other protocols. This is one of the main reasons that most popular video hosting service providers, such as Apple HTTP live steaming [17], Microsoft smooth streaming [18], Adobe HTTP dynamic streaming [19], and Akamai [20], prefer to use HTTP with TCP instead of RTP/UDP. Moreover, HTTP uses TCP as its underlying layer, allowing it to automatically achieve reliable transmissions. It also has the function of congestion avoidance.

2.3.2 Progressive download

Progressive download is a method of streaming video from server to client so that a user can enjoy a video without downloading the entirety of the video content. Although progressive download was proposed as an approach using HTTP to play the online video content [11], it does not, however, support the main aspects of real streaming, like adaptively changing the resolution and quality to match network conditions. Thus, viewers should select the most suitable representation (i.e., bit rate) before playing a

video; otherwise, they may experience interruptions and freezes if the network conditions change during play time [11].

2.3.2 Dynamic Adaptive Streaming

To overcome the limitations of progressive downloading, adaptive streaming [17-20] has been proposed as a method to resolve the drawbacks, while trying to retain the simplicity of progressive download. This approach was behind the creation of the Dynamic Adaptive Streaming over HTTP (DASH) standard, which was introduced by the Motion Picture Experts Group (MPEG) and the 3rd Generation Partnership Project (3GPP), with the goal of integrating all individual efforts [15]. Like other adaptive streaming methods, each video in DASH is encoded and compressed to a variety of video bitrates, corresponding to different resolutions and qualities. It is worth noting that DASH is encoder-agnostic, so that HEVC can easily be used along with DASH. These compressed versions are called different representations of the video. Next, all of the representations are fragmented into several segments, usually with constant durations of a few seconds each. These segments are then stored in common web servers in company with a generated XML-based file called Media Presentation Description (MPD). This file is sent to the client via the server, and it is responsible for completely determining the available representations and corresponding URLs. DASH is a pull-based method [15] that allows the client to start playing a video by asking for an MPD file using HTTP GET requests. The client becomes aware of available video by parsing the received MPD file and sending requests to fetch and download appropriate segments based on its knowledge about the conditions of its own network, such as incoming bandwidth and the status of the incoming buffer [21]. The procedure of the DASH transmission protocol can be seen in Figure 10.

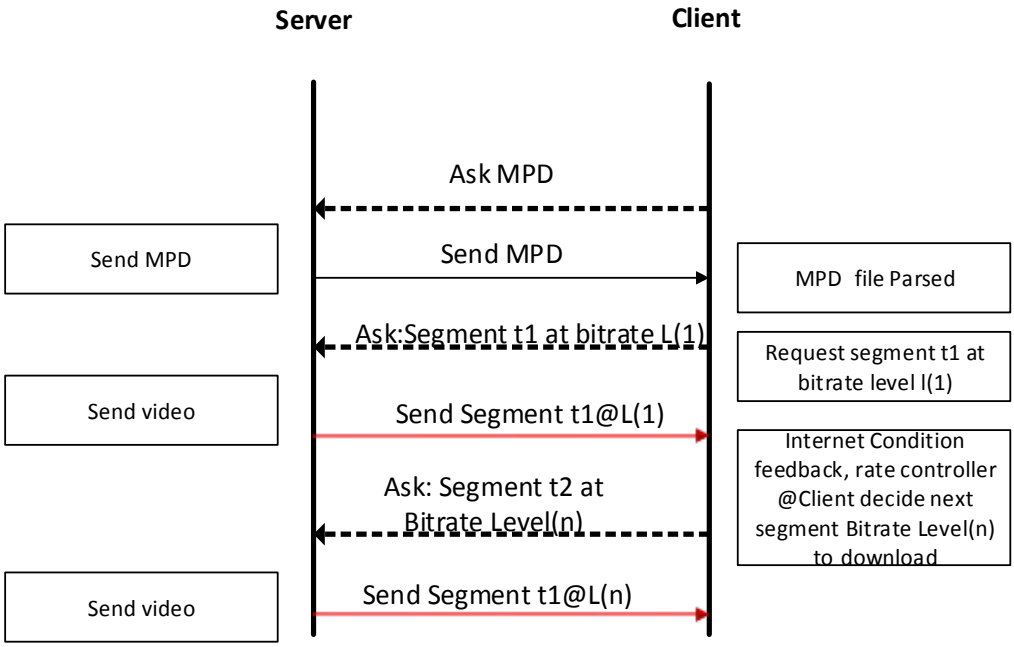


Figure 10 DASH transmission protocol

2.4 Methodology of Quality Evaluation

Several processes, such as processing, compression, transmission and reconstruction, may result in the distortion of video content. These distortions lead to the degradation of the perceptual quality of the video content. The goal of quality evaluations for a video is to estimate the user's satisfaction with the perceptual quality of the video. There are two main ways to evaluate the quality of experience of video content: a subjective test and an objective test.

2.4.1 Subjective Test

One clear and direct way of evaluating the perceptual quality of a video is the subjective test. In the subjective test, users are asked to rate different videos to determine the videos' levels of perceptual quality.

In [17], a Simultaneous Double Stimulus for Continuous Evaluation (SDSCE) and Double-Stimulus Continuous Quality-Scale (DSCQS) were used according to ITU-R recommendation BT.500-13.

In DSCQS, the purpose of this methodology is to measure the quality of a group of videos in relation to a reference. The viewer is presented with raw and distorted videos in two screens. The order of different levels of distortion sequences is random. Between each video, there is a three-second gray display. At the end, the participants score the quality of both videos on a scale from 1 to 5 (1 is bad, 2 is poor, 3 is fair, 4 is good, and 5 is excellent).

Despite the accuracy of this test for evaluating video quality based on the human visual system, this methodology is inconvenient and time consuming. This is especially true for multi-view video or 3D video, in which case more than nine views needed to be tested for every participant in order to get a sufficient evaluation of the video.

2.4.2 Objective Test

The objective metric of video quality plays an important role in video applications.The simplest method of measuring video quality is the Mean Square Error (MSE) test. The MSE calculates the absolute difference between each counterpart pixel in the test frames and the reference frame, as can be seen in the following:

$$\mathrm{MSE} = \frac{1}{mn}\sum_{i=0}^{m-1}\sum_{j=0}^{n-1}[X(i,j) - Y(i,j)] \tag{2.10}$$

where m and n refer to the length and width of each frames and (i,j) are the corresponding coordinates of each pixel in the frame.

Peak signal-to-noise ratio (PSNR). The PSNR is a widely used objective test metric for video and image processing. It is most commonly used to measure the quality of lossy compression in digital transmission.

$$\text{PSNR} = 10log_{10}\frac{255^2}{\frac{1}{mn}\sum_{i=0}^{m-1}\sum_{j=0}^{n-1}[X(i,j)-Y(i,j)]} \qquad (2.11)$$

where m and n refer to the length and width of each frame and (i,j) are the corresponding coordinates of each pixel in the frame. As can be seen from the formula, PSNR is derived from MSE, and it is also a pixel-error-sensitive objective metric.

This model assumes that the human visual system is linear and pixel-error sensitive. However, the human visual system is a complex and nonlinear system. The main assumption that this model holds is that the quality of each frame is determined by the error visibility, as determined by a comparison of every pixel in every frame of the original and distorted videos.

With the development of the objective metric for both images and videos, researchers have found that the human visual system is easier than the pixel-based error test in terms of extracting structural information [13].

The structural similarity (SSIM) test is a measurement of structural information for two pictures: a reference (which is considered the distortion-free one) and a distorted one, as follows:

$$SSIM(x,y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)} \qquad [13] \qquad (2.12)$$

where x and y refer to two images (or frames in two videos). μ_x and μ_y are the mean intensities of the two comparison signals and σ_x and σ_y are the standard deviations for estimating the contrasts of the two signals. It is worth noting that μ_x and μ_y and σ_x

and σ_y are computed within an 8 × 8 square window, which moves pixel by pixel through each frame in order to retrieve local window-based structural statistics.

$$\mu_x = \frac{1}{N}\sum_{i=0}^{n} x_i \tag{2.13}$$

$$\sigma_x = \sqrt{\frac{1}{N-1}\sum_{i=0}^{n}(x_i - \mu_x)^2} \tag{2.14}$$

The measurement of the SSIM index includes three levels. From the lowest hierarchical level to the highest hierarchical level, they are: luminance, contrast and structure ((2.15), (2.16) and (2.17), respectively).

$$Luminance(x,y) = \frac{(2\mu_x\mu_y + C_1)}{(\mu_x^2 + \mu_y^2 + C_1)} \tag{2.15}$$

$$contrast(x,y) = \frac{(2\sigma_x\sigma_y + C_2)}{(\sigma_x^2 + \sigma_y^2 + C_2)} \tag{2.16}$$

$$Structure(x,y) = \frac{(\sigma_{xy} + \frac{1}{2}C_2)}{(2\sigma_x\sigma_y + \frac{1}{2}C_2)} \tag{2.17}$$

2.5 Scalability of Video Coding

In video communication systems, there are two types of delivery systems for video transmission. The first is the push-based streaming system, in which a server controls the transmission and pushes the video packets to the client. In the push-based

transmission system, H.264/SVC (Scalable Video Coding) is widely used to encode video once at several different layers and to decode in different scalable ways (e.g., resolution, temporal, and quality). In scalable video coding, a high-quality video bit stream consists of hierarchical sub-bit streams. The sub-bit streams are obtained by dropping packets from the higher hierarchical bit streams. RTP and UDP are common push-based protocols. The second is the pull-based streaming system. In a pull-based streaming system, the client makes a request from the server based on different available bandwidths, and the server provides scalable content in order to offer clients choices related to scalability.

Despite the different concepts of the two types of video delivery systems, the scalability methods are the same. The three main scalabilities are: temporal scalability, spatial scalability, and quality scalability.

- Temporal Scalability: The frame rate of the video is the main temporal scalability method, such that, the higher frame rate is, the more continuous perceptual quality the user can feel.
- Spatial Scalability: Videos are encoded with different resolutions in order to make them compatible with different display sizes on the client.
- Quality Scalability: The video can be encoded at different quality levels when the spatial and temporal scalability are at the same level. The goal is to provide different levels of video content distortion to compare with the raw video.

Temporal Scalability

15 frames per second

30 frames per second

Spatial Scalability

256*192

512*384

1024*768

Quality Scalability

Low Quality

High Quality

Figure 11 Types of scalability in video coding

2.6 Dynamic Adaptive Streaming Over HTTP for Free-Viewpoint Video Streaming and Stereo 3D Streaming

DASH-based stereoscopic 3D video is proposed in [23], which proposes a method of encoding the scalability of one of the stereo views on the server to adaptively stream the 3D stereo content using DASH. In [24], a DASH-based free-viewpoint video streaming system is proposed. The adaptation mechanism is used to maximize the virtual views based on the rate distortion model, which is rendered from the texture reference views and their associate depth maps when the tracking information from the user's head is retrieved at the client side. The authors' methodology is to choose the best quality of synthesized views between the two reference views in one fixed baseline distance. However, even though the final goal of this system is to provide the best virtual viewing quality to users on the stereo 3D display, the system ignores the fact that it does not consider the users' experiences of the depth (or, in other words, whether rendered views allow different users to feel comfortable). Moreover, even though this system is based on a stereo display, another weakness is that the head tracking may introduce system delay, which would negatively affect DASH streaming.

In [25], the design of a DASH-based stereo 3D video system is introduced. The main issue is that stereo 3D provides only two views to viewers each time, and different viewers have different depth preferences that make them feel comfortable.

2.7 Objective Test Approach for the Multi-view Video

In [29] and [30], an objective methodology and approach for the multi-view video were proposed to build suitable, objective quality assessment metrics for different scalable

modalities in multi-view 3D videos. The proposed methodology aims at building quality evaluations of different layers in Scalable Multi-view Video Coding (SMVC) in H.264. Moreover, the proposed methodology considers each quality of experience layer that contributes to overall quality.

2.8 Subjective Approach for the Multi-view Plus Depth content

The previous work that is most related to ours is [29], a subjective test approach for streaming MVD content that examines the effect of the number of views on the quality of synthesized views. The subjective study shows that, by decreasing the distance between the baseline and the number of transmitted views, one can maintain satisfactory subjective quality. The study uses the Constant Quantizer Parameter (CQP) method to encode different perceptual content qualities by setting one specific QP. This method is not, however, part of the MPEG-DASH standard [15, 16]. Compared to the Constant Bit Rate (CBR) controller, which is based on the R-Lambda Model [9], it cannot guarantee the best quality at one specific bit rate, given bandwidth fluctuations.

In [30], the authors seeks the best bit rate allocation ratio in terms of rendered synthesized views between the depth map and the texture views in the multi-view plus depth map compression. They do so using the depth map image rendering technique. The experiments are based on H.264/MVC and the HEVC 3D extension. The results show that, even though the optimal ratio varies across different test sequences, the best depth-to-texture ratio for the PSNR value is between 30 and 60 in terms of percentage. This conclusion is used in our experiment to set the parameters for the DASH server.

Chapter 3: The Proposed Architecture and Design

This chapter attempts to explain the architecture of DASH-based HEVC content streaming over the HTTP system. The explanation is divided into two parts: the HTTP server and the DASH client. The HTTP server part introduces the segmentation schema, the view scalability scenario for the multi-view plus depth format, and the responsibility of Multimedia Presentation Description (MPD). In the adaptation client subsection, the bit stream selection based on different numbers of streamed views and the available bandwidth predictions based on exponential moving averages of instant throughput are explained.

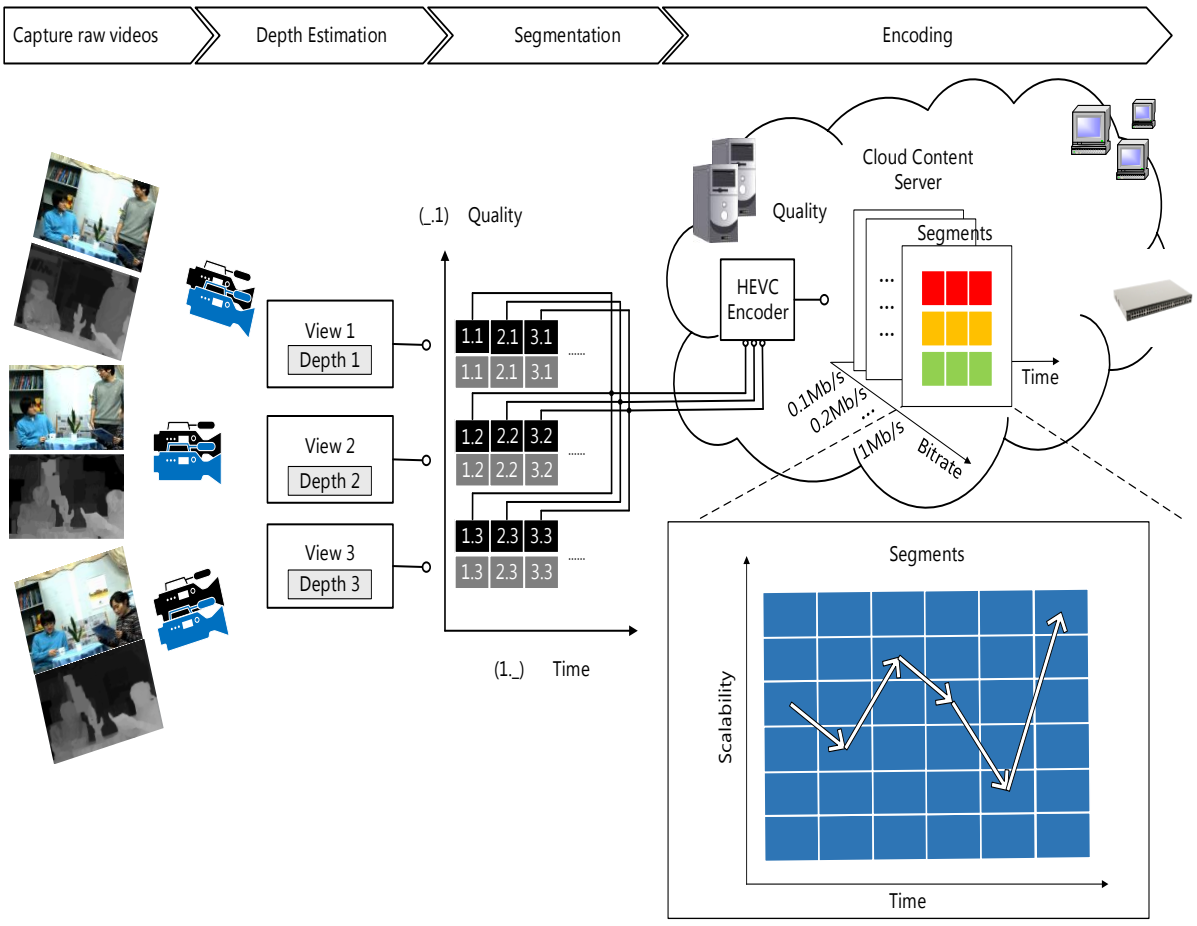


Figure 12 DASH-based multi-view video transmission system on the server

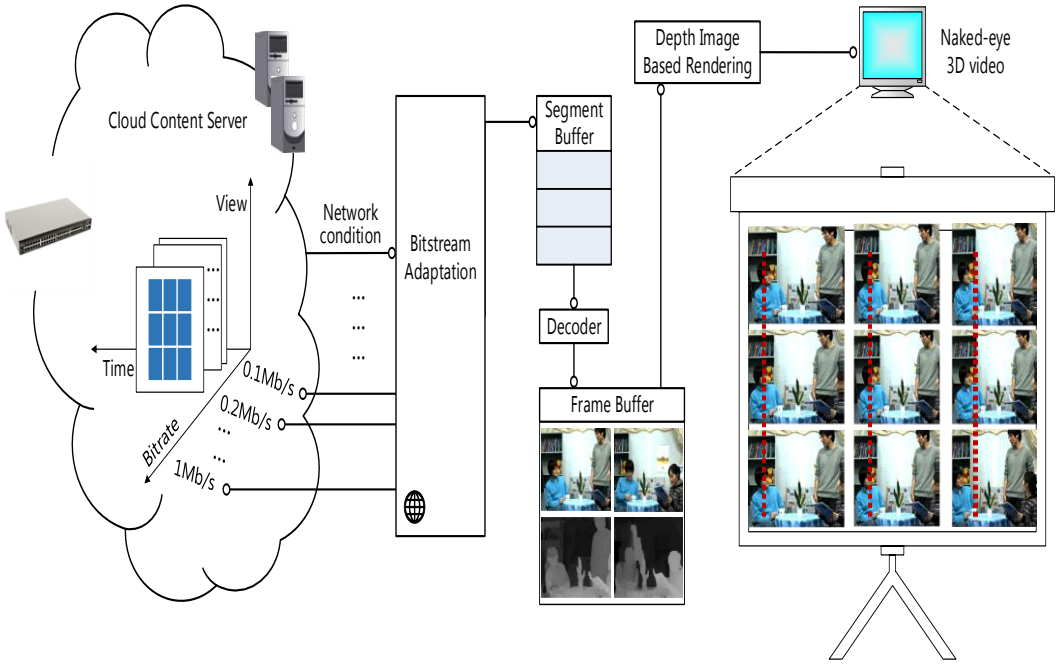


Figure 13 DASH-based multi-view video transmission system on the client side

The architecture of our HEVC multi-view video system using dynamic adaptive streaming over HTTP is shown in Figure 11 and Figure 12. The architecture shows the cloud-based DASH server side and the client side of the proposed multi-view video transmission system. The main goal of the proposed architecture is to adaptively transmit multi-view plus depth map content for 3D auto-stereoscopic video over the Internet. Both the server side and the client side of the architecture are described in detail in the following subsections.

3.1 DASH Server

The responsibility of the DASH server is to offer various versions of video so that each client can adaptively select its video representation segment (i.e., the video bit rate) according to network conditions. There are several factors that contribute to a server's selection strategy, including network congestion, available bandwidth, buffer size capacities on the client side, and the resolution of the client displayer. It must be noted that the choice of the video segment cannot be decided at the very beginning of the transmission period, due to the diversity of video content and the stochastic nature of bandwidth in best-effort networks. Therefore, the DASH server divides the entire video into temporal segments, each of which contains a video ranging between 2 seconds and 10 seconds. Then, it encodes each segment into different video bit rates and groups the segments into adaptation sets. The ultimate goal is to allow the DASH client to switch among different video bit rates according to the available bandwidth. In our DASH server, we use MVD to represent multi-view video (in other words, a scene captured from a series of cameras at various viewpoints) and the associated depth information (i.e., information captured using depth cameras, such as Kinect, or depth map creation methods, as in [1]). Furthermore, we use the HEVC 3D extension encoder in our cloud server encoding engine [10] because it provides better compression efficiency than H.264/AVC, as shown in [9]. The Rate-Lamda model in HEVC [31] provides the highest compression quality of bit streams at the target bitrate. This model can allocate bits at the Group of Picture (GOP) level, as well as at the Picture Level and Large Coding Unit (LCU) level. We revise the HEVC 3D extension encoder to encode multiple versions of segments in specific target bitrates and then store these versions in the cloud server. The raw video sequences are segmented into pieces with equal time durations. Then, these segments are fed to the encoding engine to produce different bit

rates of MVD video segments. These segments are then stored, along with their Multimedia Presentation Description (MPD) files, in the cloud.

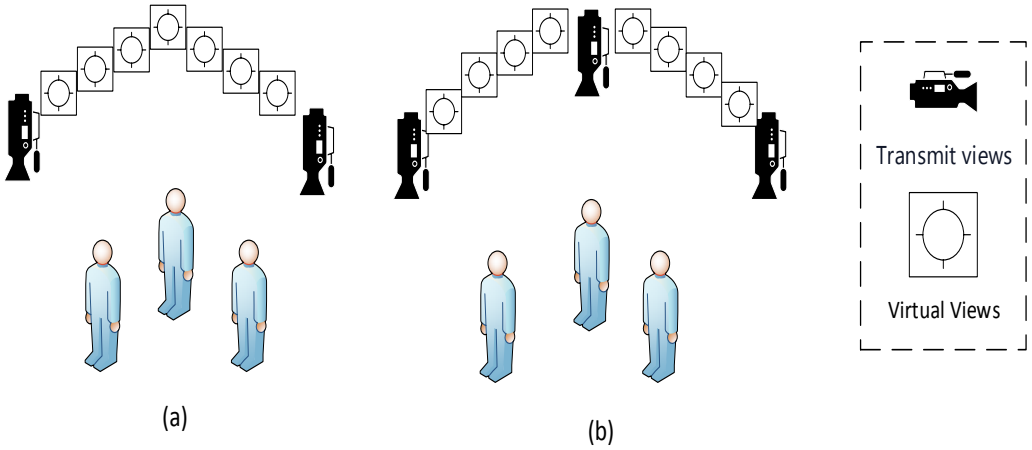


Figure 14 View Scalability Scenarios:

(a) Two-View Plus Depth with Larger Baseline. (b) Three-View Plus Depth with Narrow Baseline

Currently, there are several scalability options for 2D video, including quality, temporal and spatial scalability. These produce variant video bit rates based on priorities of perceptual quality [15].

For multi-view video, we propose adding the number of transmitted MVDs as a new scalability option. Our proposal considers that view scalability, as well as other mentioned scalabilities, should be able to adapt to the available bandwidth. By increasing the number of transmitted views (or decreasing the distance between the cameras), as can be shown in Figure 14b, we will be able to produce virtual views with higher quality than views produced using inter-camera distances, as shown in Figure 14b. As a result, the client can adapt its requests for video segments according to different numbers of views, based on the available bandwidth. This will, in turn, have a significant effect on the perceived quality of experience, as we will show in our results.

3.2 Adaptation Client

The adaptive client manages all transmission sessions when a client selects a video bit stream that is scalable for each temporal segment, such that the user can perceive the maximum quality of the 3D content given the available bandwidth. The DASH client starts playing a certain video by asking for the Media Presentation Description (MPD) file using HTTP GET requests. After parsing the MPD file (as can be seen in Figure 15), the client knows the representation of the content in the server. Then, it makes a decision about which segment version to download. In DASH-based systems, the client controls the streaming session and manages the adjustment of video bit rates in reaction to network conditions, incoming bandwidth, and the status of the playback buffer. The client-side adaptation algorithm switches among different versions of the temporal segments.

In Figure 12, we show the segment buffer, which provides a safe margin in the case of sudden decreases in network bandwidth [39], but which can also be useful for predicting the available bandwidth, as well as the throughput of the downloaded segment. The following section 3.2.1 will show the algorithm of the bitrate selection in the adaptation client.

```
</BaseURL>
<Period duration="PT1M">
<AdaptationSet segmentAlignment="true" group="1">
<Representation id="1" mimeType="video/hevc" codecs="HEVC3D" width="1024" height="768" frameRate="
30" numberofViews=3 sar="1:1" startWithSAP="1" bandwidth="1224800">
<SegmentList timescale="1000" duration="10001">
<Initialization sourceURL="
video/newspaper_10sec/300kbps/newspaper_720p_300kbps_3view_10sec_segmentinit.hevc"/>
<SegmentURL media="video/newspaper_10sec/300kbps/newspaper_720p_1200kbps_3views_10sec_segment1.hevc
"/>
<SegmentURL media="video/newspaper_10sec/300kbps/newspaper_720p_1200kbps_3views_10sec_segment2.hevc
"/>
<SegmentURL media="video/newspaper_10sec/300kbps/newspaper_720p_1200kbps_3views_10sec_segment3.hevc
"/>
<SegmentURL media="video/newspaper_10sec/300kbps/newspaper_720p_1200kbps_3views_10sec_segment4.hevc
"/>
<SegmentURL media="video/newspaper_10sec/300kbps/newspaper_720p_1200kbps_3views_10sec_segment5.hevc
"/>
....
....
....
</SegmentList>
</Representation>
<Representation id="4" mimeType="video/hevc" codecs="HEVC3D" width="1024" height="768" frameRate="
30" numberofViews=2 sar="1:1" startWithSAP="1" bandwidth="1896448">
<SegmentList timescale="1000" duration="10001">
<Initialization sourceURL="
video/newspaper_10sec/300kbps/newspaper_720p_300kbps_2views_10sec_segmentinit.hevc"/>
<SegmentURL media="video/newspaper_10sec/300kbps/newspaper_720p_1800kbps_2views_10sec_segment1.hevc
"/>
<SegmentURL media="video/newspaper_10sec/300kbps/newspaper_720p_1800kbps_2views_10sec_segment2.hevc
"/>
<SegmentURL media="video/newspaper_10sec/300kbps/newspaper_720p_1800kbps_2views_10sec_segment3.hevc
"/>
<SegmentURL media="video/newspaper_10sec/300kbps/newspaper_710p_1800kbps_2views_10sec_segment4.hevc
"/>
<SegmentURL media="video/newspaper_10sec/300kbps/newspaper_710p_1800kbps_2views_10sec_segment5.hevc
```

Figure 15 A template of a Multimedia Presentation Description (MPD) for a newspaper sequence

3.2.1 Bitstream Selection

The client is responsible for deciding which bit stream to use for different available bandwidths. In our system, the selection of bit stream is based on current network conditions, such that the perceptual quality of the average rendered views is maximized.

Section 4.3 will introduce our objective measurement test results. Specifically, it will show the effect of different numbers of transmitted views at the same total bitrate level on the final perceptual quality of rendered views and transmitted views. The result will lead us to set a policy in which, in order to switch the bitrate selection to more or fewer views, the selected fixed total bitrate of the generated bit stream should be lower than the available bandwidth. At this stage of research, we assume that each of the views shares total bitrate equally. Moreover, since, in MVD, the encoder does not seek similarities among different views, we have a linear relationship between the total bitrate and the number of views. Hence, the total bitrate used in MVD can be represented as:

$$\text{TotalBitrate} = Bitrate_{view} \times i \qquad (3.1)$$

where $\text{Bitrate}_{\text{view}}$ represents the bitrate per view, TotalBitrate represents the total bitrate of the bit stream, and i represents the number of transmitted views.

Although extensive adaptation logic is required in conventional DASH players to select the most appropriate video bitrate among the available representations, most of them look for the largest total video bitrate that is lower than the available network bandwidth. It is worth mentioning that our method can be used as a complement to conventional methods. In this approach, first, we would use conventional methods to select the largest representations that are less than the available network bandwidth. Then, since MVD makes it possible to have the same video stream (in terms of bitrate) with different numbers of views, we can use equation 3.2 to select the most suitable video segment, according to the corresponding computed SSIM.

$$\text{TotalBitrate} = \begin{cases} Bitrate_{view} \times i & \text{if } AvSSIM(i) > AvSSIM(i-1) \\ Bitrate_{view} \times (i-1) & \text{if } AvSSIM(i-1) > AvSSIM(i) \end{cases} \qquad (3.2)$$

where $AvgSSIM(i)$ represents the average of the SSIM values of all rendered views compared to raw views, without any quality loss from compression. We show that, in certain total bitrates or available bandwidths, the user can request more or fewer transmitted views from the server.

3.2.2 Available Bandwidth Prediction

We implemented a smoothed throughput based available bandwidth prediction method, which predicts the available bandwidth using a Moving Average (MA) of the observed throughputs.

This algorithm determines the optimal quality level, considering the moving average of the throughput of downloading segments (measured as Th_{inst}). The estimated throughput can be represented as:

$$Th_{est}(\mathrm{t}+1) = \begin{cases} (1-\alpha) \times Th_{est}(\mathrm{t}) + \alpha \times Th_{inst}(\mathrm{t}), & \text{if } \mathrm{t} > 0 \\ Th_{inst}(\mathrm{t}) & \text{if } \mathrm{t} = 0 \end{cases} \quad (3.3)\ [43]$$

where $Th_{inst}(\mathrm{t})$ represents the instant throughput measurement, t represents the order of the segment sequence downloaded, and Th_{est} is the estimated throughput or available network bandwidth. The α is a weighting value.

Our resulting algorithm for bit stream selection is as follows:

Algorithm: MA throughput smoothed bitstream selection Algorithm

Input: Instant throughput Th_{inst}, playlist t, counter, level of video Representation S_n (1 representates the lowest quality level), number of transmitted views in each video representation $S_n(i)$, Bitrate of representation for Nth video representations $Totalbitrate_n$

Begin

if counter>0

Download from the minimum video bitrate:

$$S_1$$

Update the estimated throughput:

$$Th_{est}(t+1) \leftarrow Th_{inst}(t)$$

Counter--

else

Calculate the available bandwidth based on the Lookup table:

$$Th_{est}(t+1) \leftarrow (1-\alpha) \times Th_{est}(t) + \alpha \times Th_{inst}(t)$$

Find the suitable representations in server for $Th_{est}(t+1)$:

$$Totalbitrate_{n-1} \leq Th_{est}(t+1) \leq Totalbitrate_n$$

$$Download\ S_n$$

While

Number of candidates for Nth representations

Number ($Totalbitrate_n$)> 1

do

Decide number of transmitted views based on (2)

if $AvSSIM(i) > AvSSIM(i-1)$

$$Download\ S_n(i)$$

else

$$Download\ S_n(i-1)$$

end if

Counter --

end if

- At the beginning, start with the lowest quality segments.
- The moving average of the throughput from the last five downloaded segment is used as the prediction of the next 10 seconds available network bandwidth.

- The selected quality level can be adjusted up and down based on the moving average of estimated throughput). For certain specific total video bitrates provided by the server, the client can decide on different numbers of transmitted views to download, based on the computed SSIM values. The switching operations are contained in the algorithm's while loop.

Two main advantages of this algorithm are that it efficiently utilizes available bandwidth and that it is sensitive to changes in estimated available bandwidth. Moreover, it uses a new scalable method to transmit Multi-view and Depth (MVD) content for 3D videos in terms of the number of transmitted views. This will further maximize the perceptual quality of virtual views following rendering, thereby increasing the user's quality of experience.

3.2.3 Reconstruction Based on MVD Format

The reconstruction of the MVD representation for potential 3D content is acquired after the decoding process, as can be seen in Figure 16. The rendering software introduced in [10] and [32] has already proven to be better than MPEG VSRS with regard to both SSIM and PSNR for rendering synthesised views from MVD videos. Following the rendering process, the MVD video representation can produce multiple virtual views for the requirement of an auto-stereoscopic 3D display. The reason we render virtual views on the client side is to avoid the transmission of a large number of virtual views, which might not be optimal in the case of best-effort networks. The number of virtual views plus multi-views ranges from 9 to 27 [33, 34]. This will linearly increase the bandwidth burden.

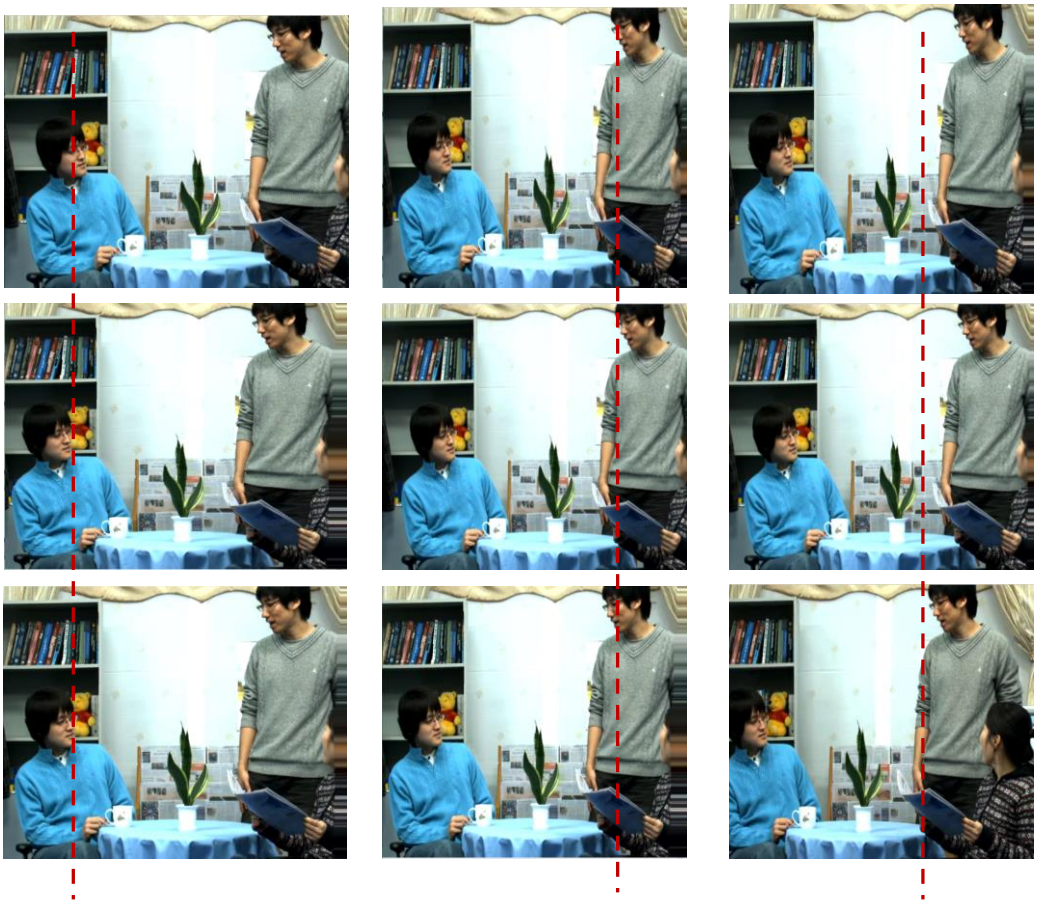


Figure 16 Nine rendered views from 3V+D (Number 300 frame of the newspaper sequence)

In Figure 16, the results of nine rendered views from three transmitted views, plus associated depth maps, are shown. These images are drawn from the 300th frame of the newspaper test sequences. Even though all views represent the same scene, slight differences exist between each in terms of the perspectives and positions that can be seen with regard to the figures.

Chapter 4: Implementation and Evaluation of the Proposed System

In this chapter, several experiments are proposed for the evaluation of the performance of the system introduced in Chapter 4. The simulation and emulation environments are introduced in Section 4.1, including the introduction of the structure of the test bed, the simulation setup that introduces the experimental devices, the content preparation, and the network simulator. In section 4.3, the objective measurement results are presented. The bit stream selection algorithm introduced in Chapter 3 is based on the objective measurement results in order to maximum the quality of multi-view content at each available video bitrate level. In Section 4.4, the subjective test is introduced for the newspaper test sequence, and some conclusions and analyses are presented, including the final emulation test results. In Section 4.5, the system behaviours under different available bandwidths are presented in order to show our policy regarding bit stream selection based on the average SSIM introduced in the previous section. In Section 4.6, the emulation of 3D multi-view video content transmission is presented.

4.1 Experimental Prototype

In order to implement and evaluate the proposed system, the prototype consists of two main parts: the scalable server and the adaptation client. Figure 17 introduces the experiment environment, which includes three parts within the evaluation system,

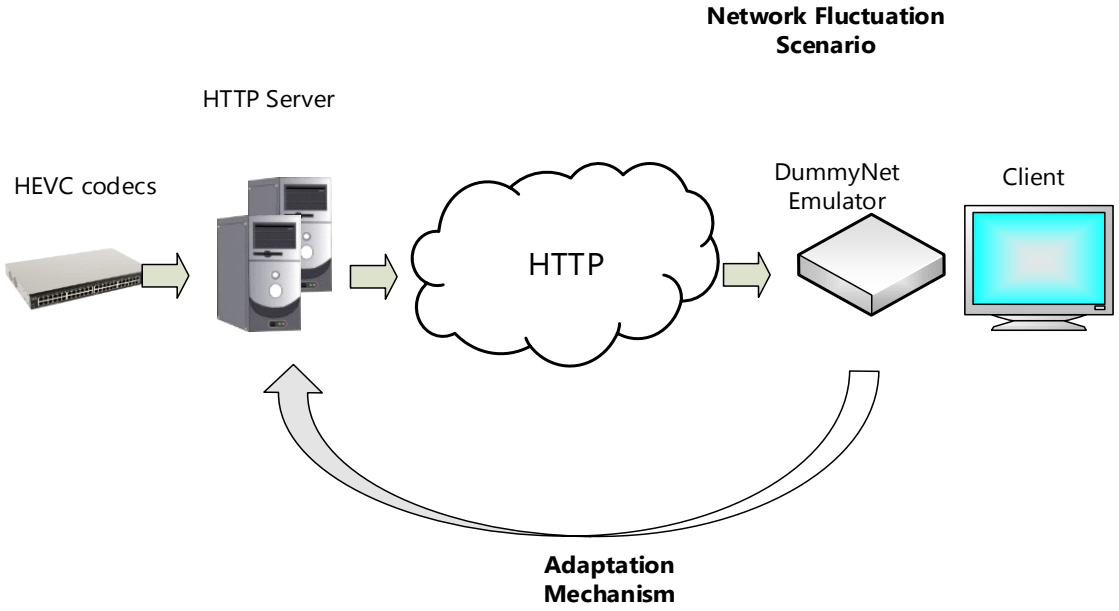


Figure 17 Experimental Environment for DASH-based 3D video streaming system

- HTTP server: Store DASH content segments as segments of different available video bitrates, with corresponding MPD files.
- DASH client: This part is implemented based on the algorithm introduced in algorithm one. It provides the rate adaptation mechanism of the proposed multi-view video streaming system.
- Scenarios for network fluctuation: A set of network fluctuation scenarios between the HTTP server and the DASH client is emulated through network simulator tools.

4.2 Simulation Setup

4.2.1 Experimental Device

We use two PCs as the emulation devices for transmission. One is used as the server, and the other is used as the client. The specification of devices is shown in Table 1.

Table 1. Specifications of devices for the transmission emulation

	Client and Server
Name	Dell Optiplex 7010
CPU	Intel(R) Core(TM) i7-3770@3.4GHz
OS	Windows 7
Network	Wire Gigabit Network

4.2.2 Content Preparation

In order to test and evaluate our proposed system, we use the following test sequences: Kendo [35], Newspaper [36], and Balloons [56], which are recommend by MPEG [37]. The properties of the test sequences are listed in Table 2. We set the picture group length and the intra-period to 8 and 24, respectively, for all test sequences, and the segment duration is considered to be 10 seconds for the Newspaper and Kendo test sequences.

Table 2 Properties of the test sequences

Test Sequences Name	Frame Rate per Second	Resolution Width*Height	Views	Length of the Sequence	Distance of Baseline
Newspaper	30	1024*768	2,4,6	300	5cm
Kendo	30	1024*768	1,3,5	300	5cm
Balloons	30	1024*768	1,3,5	300	5cm

4.2.3 Segmentation Scenario

Since longer sequences are not possible with the above-mentioned sequences, we had to repeat each of the test sequence ten times so that each could be considered a segment. That is to say, for each segment, the duration was 10 seconds. We provide 25 segments for each test sequence at each bitrate level.

4.2.5 Selection of the bitrate levels for each test sequence

We prepared two types of streams. The first consisted of two views, plus their corresponding depths (2 V+D), while the second included three views, plus corresponding depths (3 V+D). To prepare the first stream types, 2V+D, we used views number 2 and 6 in the Newspaper test sequence and views number 1 and 5 in the Kendo test sequence. For the second type of stream, 3V+D, we used the rest of the views in the

aforementioned test sequences, plus their related depths, to emulate the streaming of a higher number of views.

The segments in the 2 V+D streams are encoded as follows using Constant Bit Rate (CBR) per view: 300kbps, 500kbps, 800kbps, 1000kbps, 1200kbps, 1500kbps, and 2000kbps. Thus, the total bit rates of the different 2V+D streams are 1200kbps, 2000kbps, 3200kbps, 4000kbps, 4800kbps, and 6000kbps, respectively. As pointed out in [4], the view and its depth have equal bit rates. For example, we have two views, each at 300kbps, and two depths at 300kbps to produce a stream at a 1200kbps bitrate. In a similar way, we used the same encoded video bitrates for 3V+D, such that the total 3V+D streams range from 1800kbps to 9000kbps. The segments belonging to all representations of 2V+D streams and 3V+D streams, as well as the MPD file, are stored in the IIS HTTP server, as described in Figure 15.

4.2.4 HTTP Server

The HEVC 3D extension Encoder HM 11.0 [10] has been chosen as the codec because of its high compression performance compared to the H.264 encoder. Sample Adaptive Offset (SAO) is enabled. Our policy of downloading segments is discussed in the system behavior section. The segments belonging to all representations of 2V+D streams and 3V+D streams, as well as the MPD file, are stored in the Internet Information Service (IIS) HTTP server, as described in Figure 18.

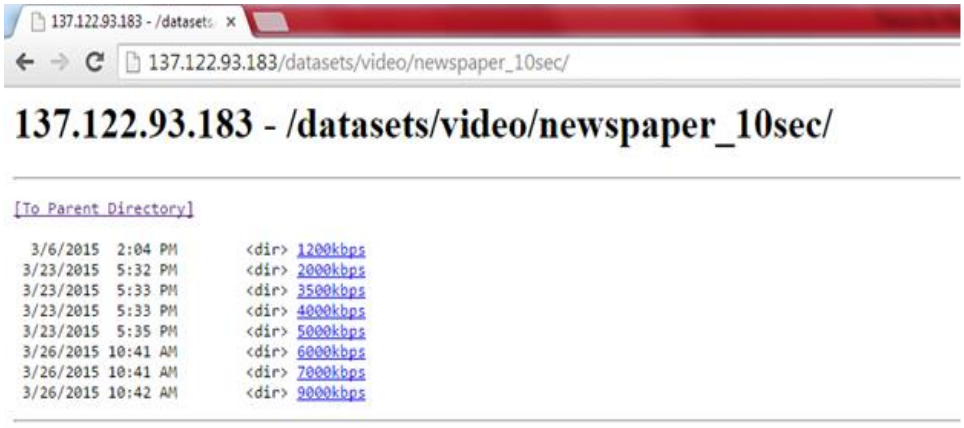


Figure 18 Sample of HTTP server for the HEVC-based 3D multi-view video streaming (Newspaper)

4.2.5 DummyNet

We used the DummyNet tool [38] on the client side, setting initial bandwidth to 2.0Mbps. This is increased by 1.0Mbps after every two segments.

In order to simulate the live streaming of multi-view content over the Internet, we use the DummyNet emulation tool in our live emulation. The DummyNet is installed between the client and server. As can be seen in Figure 19, it works by intercepting and reordering packets in a stack to simulate the restriction of available bandwidth, delays and package losses. DummyNet works between the application layer and network layer.

In brief, can be seen in Figure 20, transmitted packages can be sent through several pipes and queues to decrease the speed of outgoing packages in order to restrict outgoing packages. The incoming packets go firstly to one queue (as long as it is not full), and this queue sends the packages out according to requests for different bandwidths.

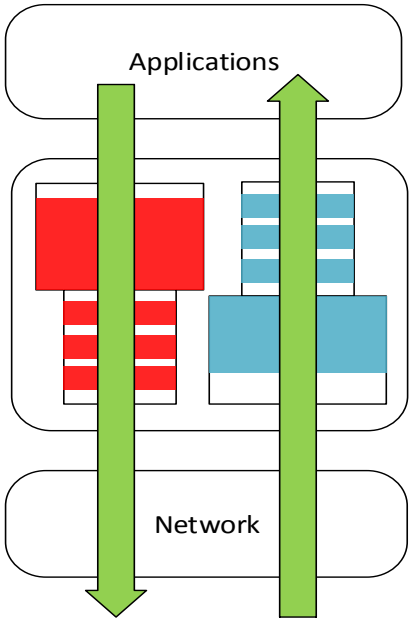


Figure 19 DummyNet restricts packages in the protocol stack. Adopted from [50].

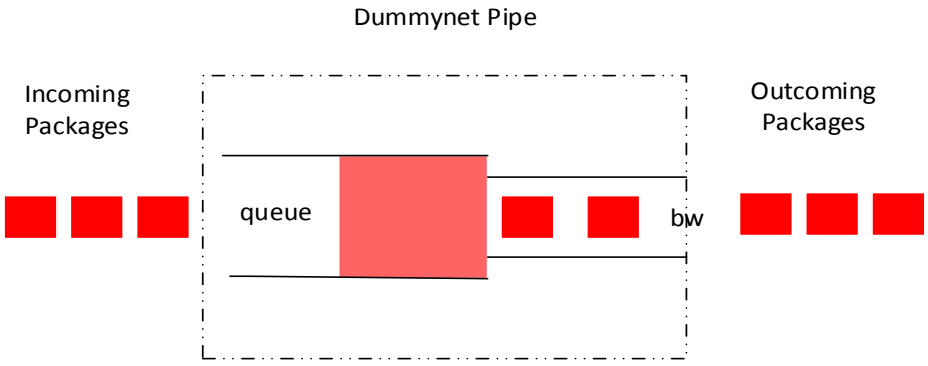


Figure 20 Queue in one pipe of DummyNet to limit available bandwidth. Adopted from [50].

We can use this tool to set up the scenarios needed to emulate different network conditions in order to evaluate the proposed system.

4.2.6 Renderer on the client side

The software introduced in [32] was chosen for the purposes of rendering to produce virtual viewpoints based on the MVD content received by the client. The BlendMode and HoleFillingMode parameters are enabled. It is worth noting that, for both the 2V+D and 3V+D experiments, the total number of views, including virtual views and transmitted views, are the same. For 2V+D, we rendered seven virtual viewpoints between transmitted views, while, in the 3V+D experiment, six virtual viewpoints were rendered, such that there were three viewpoints between the two transmitted views (as can be seen in Figure 5).

4.3 Objective Quality Measurement

In this section, by using the objective metrics PSNR and SSIM, we wanted to show how different numbers of transmitted views (i.e., decreasing or increasing the distance of the cameras) affected the quality of virtual views—and, eventually, the quality of the user experience. The results of this experiment were used to decide which policy of transmitting MVD content for auto-stereoscopic 3D displays would be best in terms of QoE under different circumstances.

We use a power curve fitting the $\mathrm{f(bitrate)}$ curve, which is computed in Equation (4.1), in order to linearly predict the perceptual quality in terms of the Peak Signal to Noise Ratio (PSNR) and the Structural Similarity (SSIM).

$$\mathrm{f(bitrate)} = \mathrm{a}\ bitrate^{b} + \mathrm{b} \qquad (4.1)$$

The inverse function $f^{-1}(\text{Quality}) = \sqrt[b]{\frac{Quality-c}{a}}$ (4.2)

Based on the inverse function, the possible qualities of different bitrates can be predicted linearly.

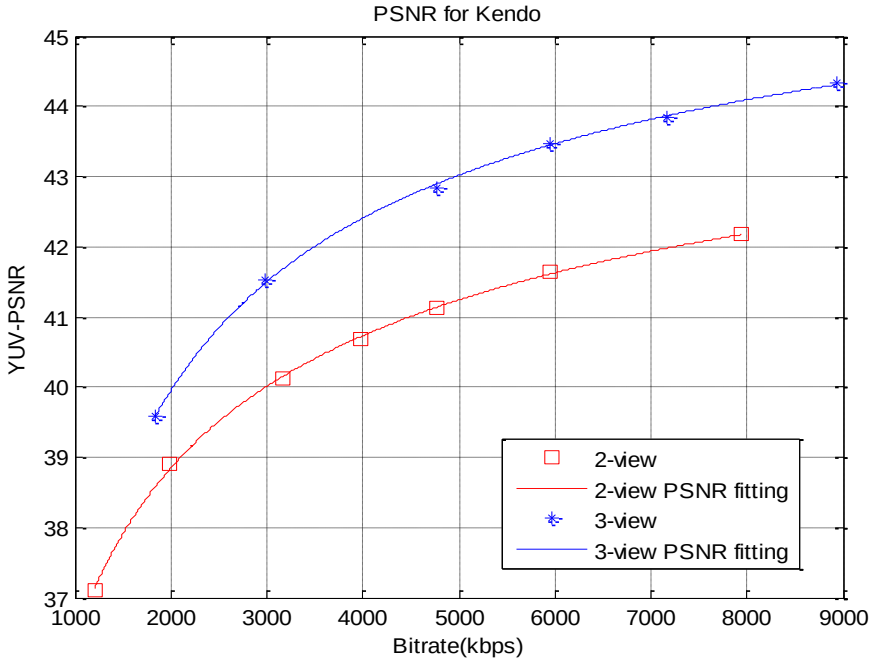


Figure 21 PSNR for Kendo

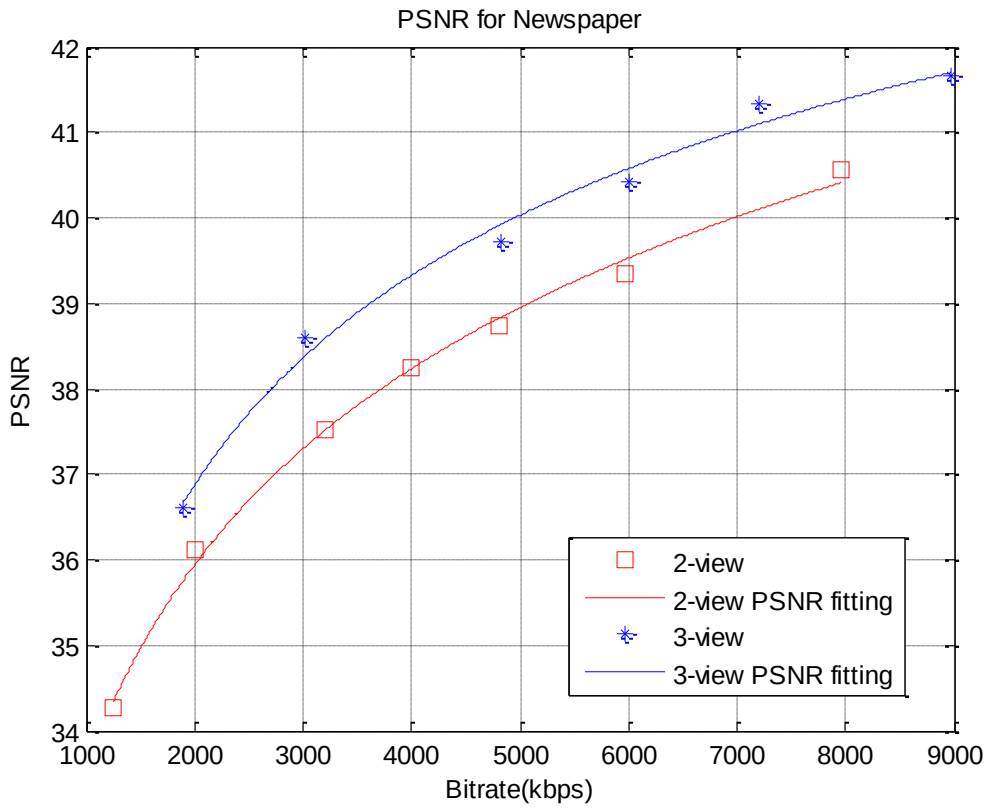


Figure 22 PSNR for Newspaper

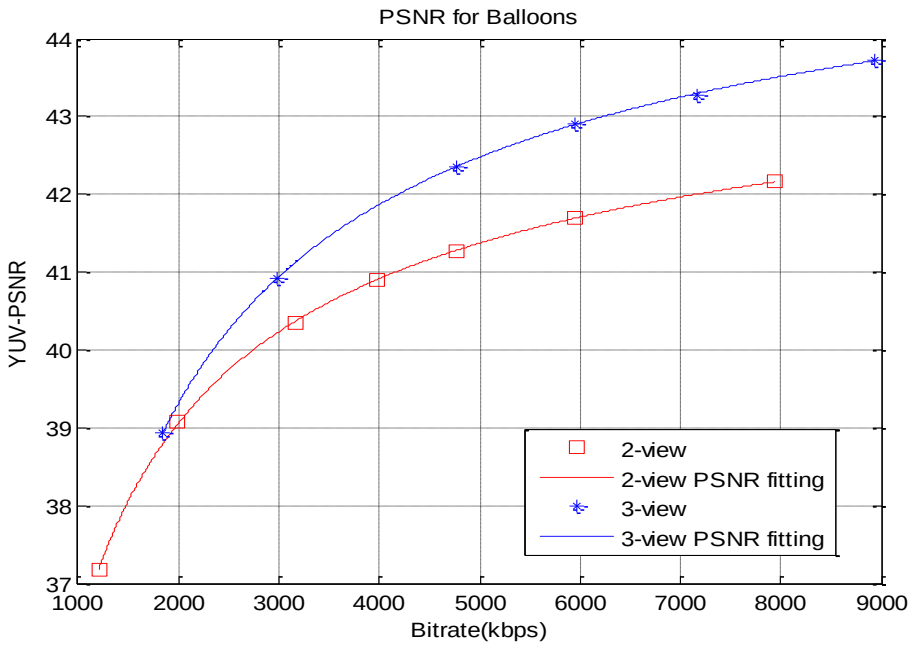


Figure 23 PSNR for Balloons

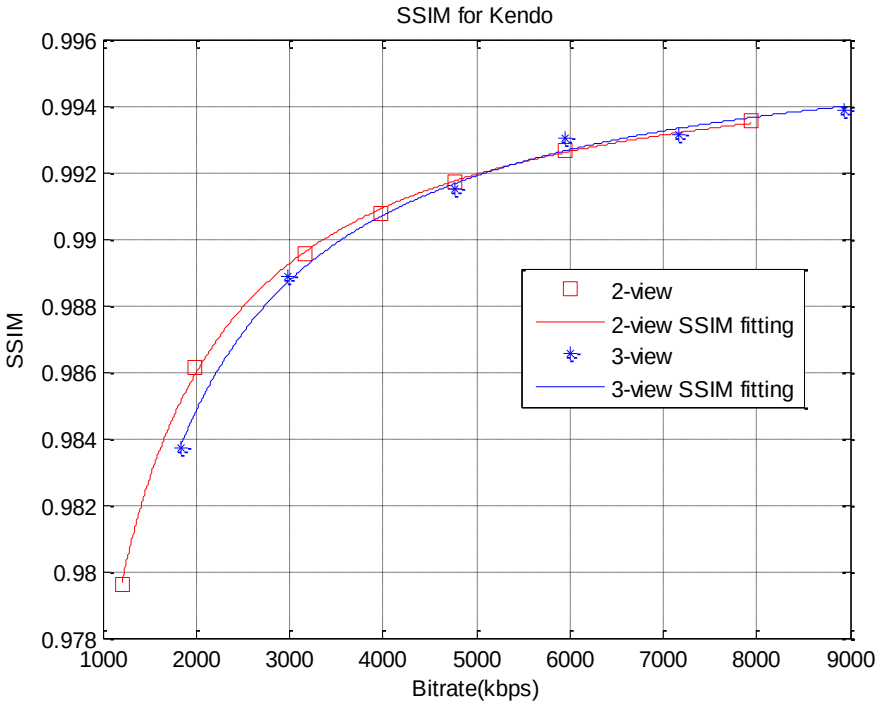


Figure 24 SSIM fitting for Kendo

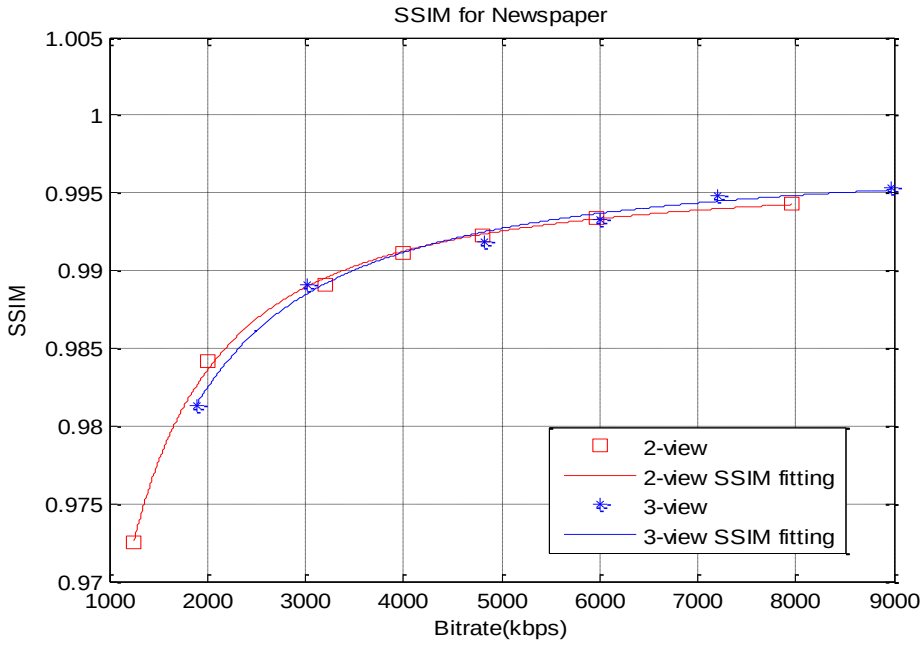


Figure 25 SSIM fitting for Newspaper

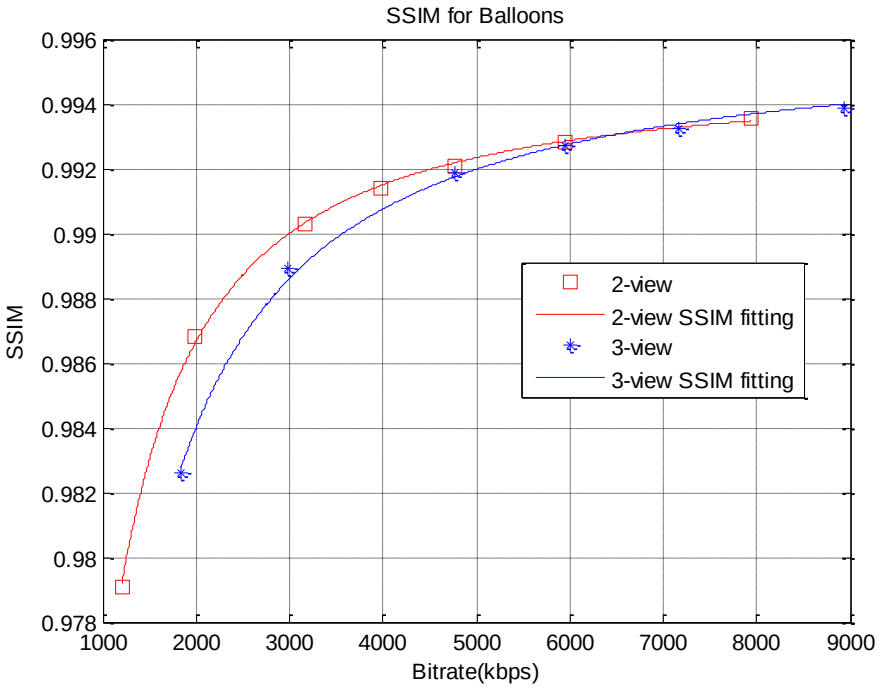


Figure 26 SSIM fitting for Balloons

Figures 21, 22 and 23 show the results of average PSNRs, while Figures 24, 25 and 26 show the result of the average SSIMs at different numbers of transmitted views (2V+D and 3V+D) for two test sequences.

When we use the PSNR as the QoS metric for the average numbers of virtual views (Figures 21, 22 and 23), given the same total video bitrate, the 3V+D format input always had a higher PSNR than the 2V+D. This means that, for a fixed video bitrate, based on the PSNR metric, 3V+D offers better quality. However, when we use the SSIM as a metric for evaluating the quality of virtual views of different total bit rates, we can see that, for lower total video bit rates, the 2V+D stream has a higher SSIM than the 3D+V stream. However, as total video bitrate increases beyond a certain point (which is around 5000kbps), 3V+D outperforms 2V+D, as can be seen in Figures 24 and 25. In Figure 26, that point is around 6000kbps. In other words, the effect of the transmitted number of views (i.e., the distance of the camera baseline) and the quality of

each view on the rendered virtual view depends on the total video bitrate. Using the SSIM metric, we can see that, in some bit rate range (from 0 to 5000kbps or from 0 to 6000), transmitting a lower number of views with optimized quality is better. However, when the total video bit rate is higher than 5000kbps or 6000kbps, it is better to first increase the number of views and then increase the quality of each view. Moreover, the variations between the rendered virtual views can be interpreted as global scene movements related to the amount of distance between two consecutive virtual viewpoints. Since the SSIM takes into account structural similarity and has a better correlation with human perception [13], it can predict the quality of rendered virtual views better than PSNR, revealing a threshold of network bandwidth that allows us to accommodate higher numbers of views.

Based on the objective test results, we can define a policy for selecting the most appropriate MVD video segment based on the number of views in terms of QoE, which can be used by the DASH client. In other words, increasing the number of views in reaction to the increasing network bandwidth does not mean a higher quality. Hence, we can select segments with lower numbers of views, but better quality, until the available bandwidth is greater than a pre-defined threshold. For instance, in these test sequences, when the available bandwidth is lower than 5Mbps or 6Mbps, we select the segment with a total MVD bit rate that is lower than the available bandwidth. Otherwise, when the available bandwidth is higher than 5Mbps, increasing the number of views has a higher priority than increasing the quality of each view. That is to say, once the available bandwidth reaches 6 Mbps, we select MVD segments with higher numbers of views, moving from 2V+D to 3V+D. If the available bandwidth kept increasing, we would select the segments that had, not only higher number of views (3V+D), but also higher qualities for each view.

4.4 Subjective test

4.4.1 Subjective test setup

Our subjective test aims to test the perceived quality of the generated views. Due to the lack of matching monitors available in the market for 3D multi-view displays, we chose one specific generated view number (3.5 views) on a 2D display to test the rendered results. We used a 46” HD Parnos Monitor (provided by our industrial partner, Magor Corp) [53] to run the tests.

This subjective test seeks to compare the subjective quality of different numbers of input views (i.e., two views or three views) at seven different bitrates, which we chose from the RD model. We design the test according to ITU-R recommendation BT.500-13 [17].

In the setup of the subjective test, fifteen people from our lab with no 3D video test experience participated. All of the test sessions began after a training phase. During the training phase, the participants were told what they were going to see in the test, what they had to evaluate and how to express their opinions [17].

4.4.2 Subjective Test Scenario

We used two HD screens to show the video sequences at the same time. The participants scored what they saw. The test had two parts to gauge subjects’ opinions of the selected synthesized views for both two-view and three-view situations.

In this case, the first part of the designed test was based on the Simultaneous Double Stimulus for Continuous Evaluation (SDSCE) [17]. The subjects watched two sequences simultaneously. The reference test sequence and the test sequence were

displayed side by side on the same monitor. For the first part, we let the subjects watch the reference video, which was the nearest original raw video (view 4) from the "newspaper" test sequence, together with the generated views for the three input views under seven different bitrates (from 0.6 Mbps to 4.2 Mbps). Then, we let the subjects evaluate the synthesized view video based on the reference raw video (view 4) by marking both videos between 0 and 5. As a rule of thumb, 1 was bad, 2 as poor, 3 was fair, 4 was good, and 5 was excellent. We also asked the participants whether the tested video was acceptable as a video to watch.

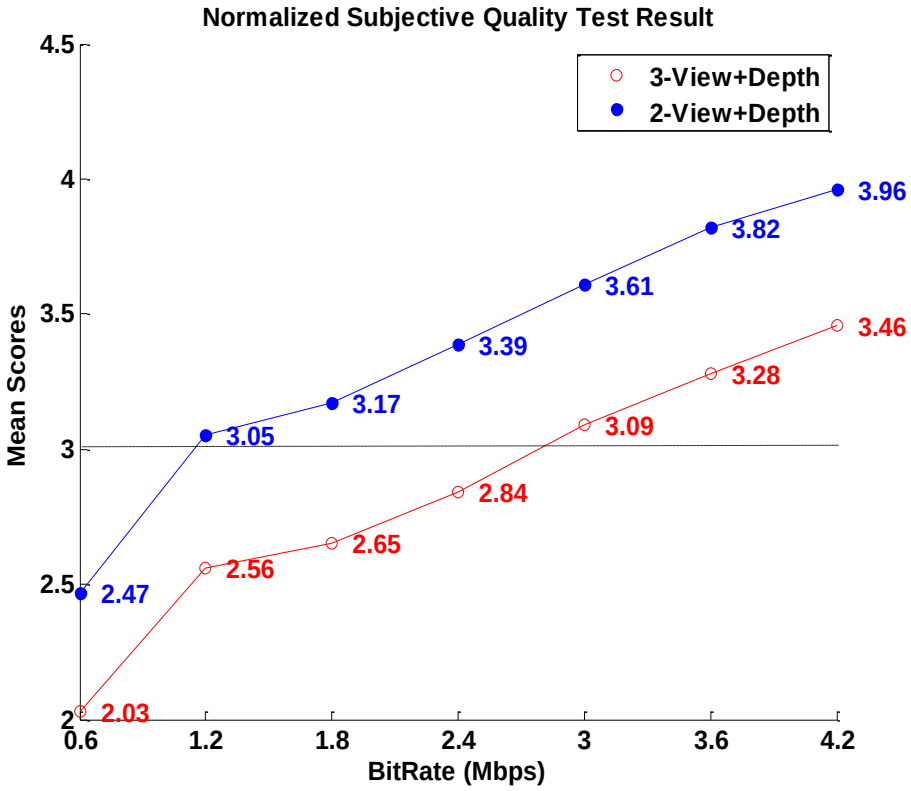


Figure 27 Subjective test results for the Newspaper sequence

The second part of the test was designed based on the Double-Stimulus Continuous Quality-Scale (DSCQS) method introduced by ITU-R recommendation BT.500-13 [17]. Between two 10-second test sequences, one 4-second grey interval is given. We let the subjects compare the generated views from both the two-view and the three-view

scenarios and evaluate the scores for each scenario. The participants also wrote down which side was better.

To analyze the subjective test results, we set the marks for the 3-view input from 0 to 5 (3 and below was poor, and 4 and above was good) as the first test group and collected the relative scores for the 2-view input from the second part of tests. The Mean Opinion Scores (MOS) were determined using the data of all 15 subjects. The results are shown in Figure 27. As the total bitrate increased, the average opinion score of the subjects rose, which is intuitive. Figure 28 also shows that 2V+D surpasses 3V+D at each bitrate. However, as stated above, videos with scores under 3 are considered poor; thus, it can be seen that, for 3V+D, the lower acceptable bound for the total bitrate is about 3 Mbps, while, for 2V+D, the minimum acceptable bitrate is about 1.2 Mbps. We can conclude that the minimum quality required for two views should be no less than 1.2 mbps. However, we decided that the switching point between 3V+D and 2V+D is 3Mbps, which also determines the minimum bitrate for 3V+D, based on the assumption that 3V+D has higher priority for streaming, since it can provide more views for the clients.

4.4.3 The limitation of the subjective test

Although the subjective test is used in this thesis, there are still some limitations. Our subjective test was conducted under the assumption that the quality of each view had an effect on the QoE for the final 3D multi-view display. However, due to the lack of available hardware devices for 3D multi-view displays in the market, our subjective test was conducted using high definition 2D displays. We selected one of the rendered views from all generated views and then let the users score what they saw. The limitation is that:

- Only one view from all rendered views was selected.

- The subjective test takes place on a two-dimensional high definition display. 2D quality cannot prove sufficiency regarding the quality of 3D views.

We decide to rely on the objective test measurement, which is the average of the Structural Similarity scores, to predict the quality of experience for the 3D multi-view video content.

4.5 System Behavior

In order to show our transmission policy based on the average SSIMs for rendered views at different available bandwidths (as introduced in Chapter 4) and the corresponding objective test results (presented in 5.2), the goal of this subsection is to show the policy chosen based on the algorithm introduced in 5.2. This is presented from the perspective of the scalability of different numbers of views at different available bandwidths in order to maximum the final QoE for multi-view 3D video streaming.

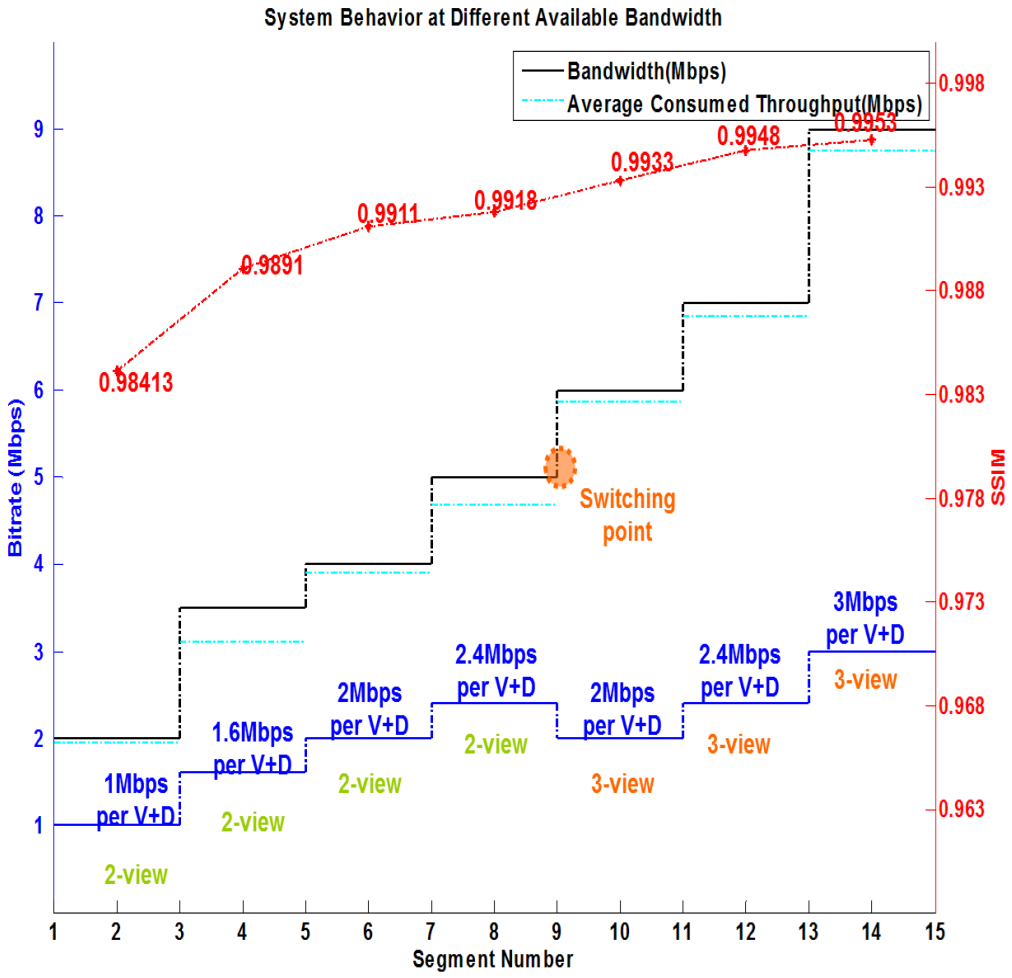


Figure 28 System behaviors at different available bandwidths (Newspaper)

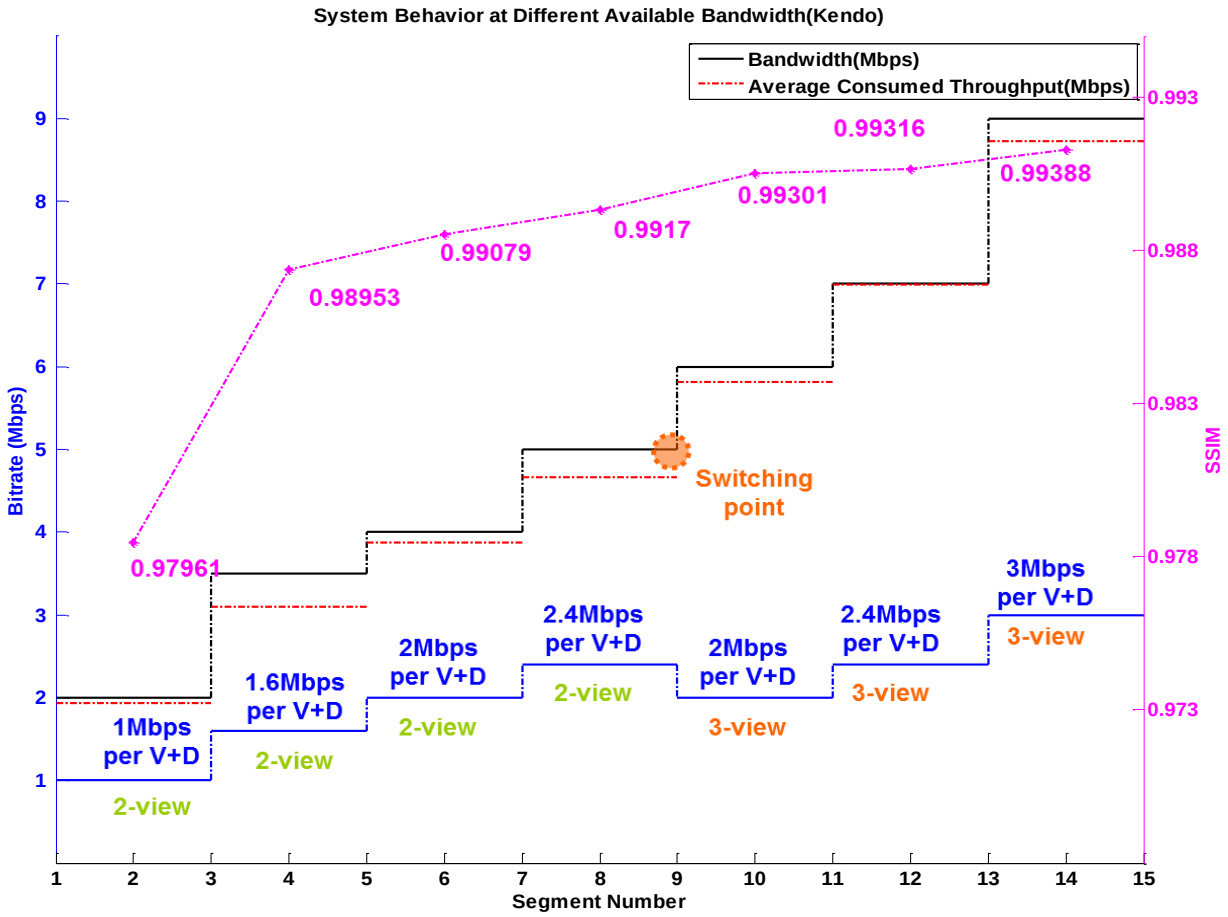


Figure 29 System behaviors at different available bandwidths (Kendo)

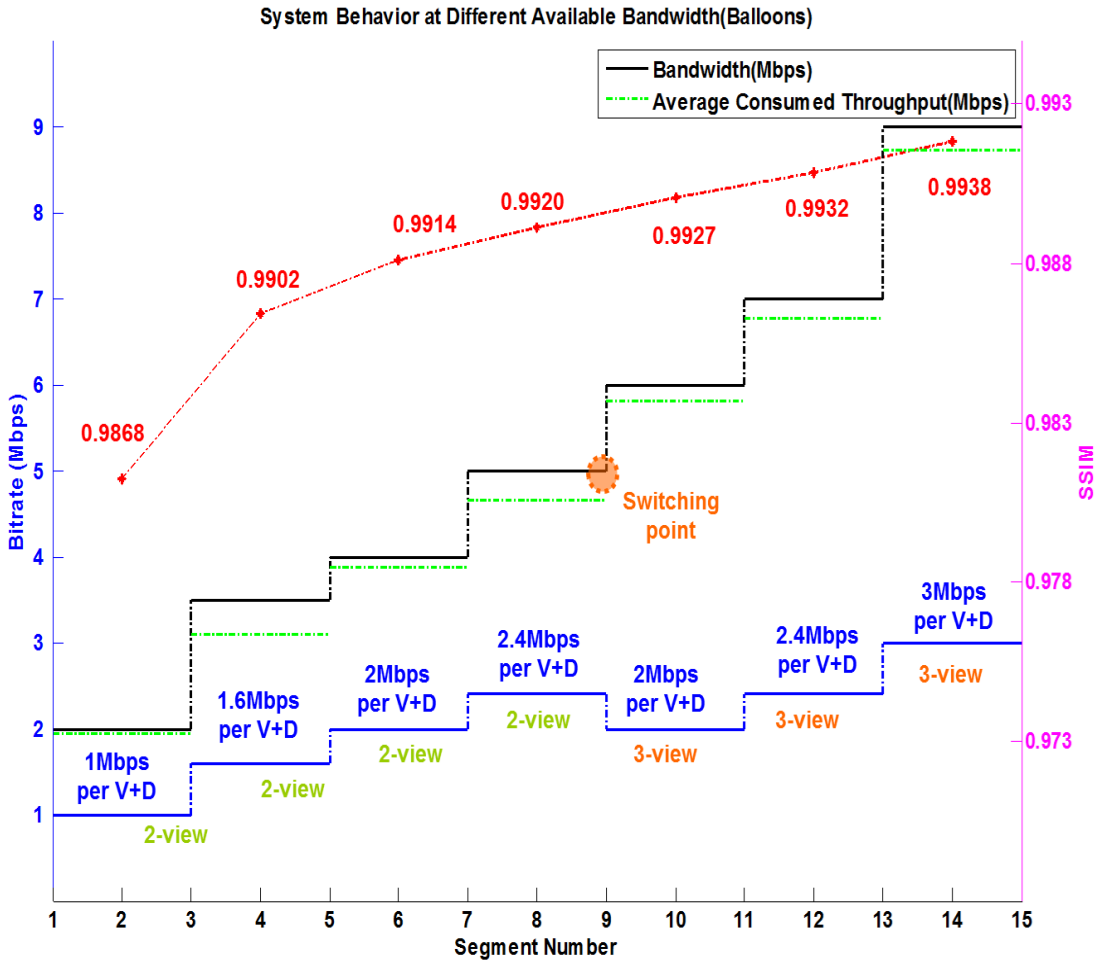


Figure 30 System behaviors at different available bandwidths (Balloons)

In Figures 29, 30 and 31, we demonstrate the transmission behaviors of our proposed system in relation to the adaptive client. The results can be seen in the two test sequences: Kendo and Newspaper. As mentioned early in this section, we initially limit the bandwidth to 2Mbps. After downloading two segments, the available bandwidth is increased by 1Mbps. Thus, the following available bandwidths used are 3.5Mbps, 4Mbps, 5Mbps, 6Mbps, 7Mbps, and 9Mbps, at seconds 20, 40, 60, 80, 100, and 120 respectively. It can be seen from both Figures 29 and 30 that, when the available bandwidth is below 5Mbps (in Figure 31, this point is 6Mbps), the priority is to adjust the bit rate and depth for each view to meet the perceived quality, rather than to increase

the number of views. On the other hand, when the available bandwidth is above 5Mbps, the priority is to increase the number of views, rather than the quality of each view. Our simulation results show that when the available bandwidth is above 5Mbps, the performance of 3V+D is better than that of 2V+D in terms of the average perceptual quality of views, as estimated by the SSIM metric. Thus, we select those segments with more transmitted views, rather than increasing the quality of each view. As can be seen from Figures 29 and 30, when the available bandwidth at segments 9 and 10 (80 seconds) is increased from 5Mbps to 6Mbps, MVD segments (3V+D) were chosen. However, the bitrate for each view stays at the same level without any increase. In this way, we can select different bit streams from the server according to variations in the available bandwidth to provide the maximum perceptual quality of virtual views to the user.

4.6.4 Emulation

In the emulation test section, the real network emulation scenario is proposed, and the results are shown from the perspective of the network transmission. Our proposed SSIM-based exponential moving average adaptation algorithm is tested and analyzed in this subsection.

4.6.4.1 Emulation test scenario

Figure 32 introduces the network scenarios used in the emulation. These scenarios emulate the long-term variations in the network fluctuation of different available bandwidths. The test takes 240 seconds and includes 24 segments (we assume that each segments takes 10 seconds). The available bandwidth increases every 40 seconds, moving from 2 Mbps to 9 Mbps. This scenario evaluates the behavior of the client of the proposed system with regard to adapting to the changeable bandwidth.

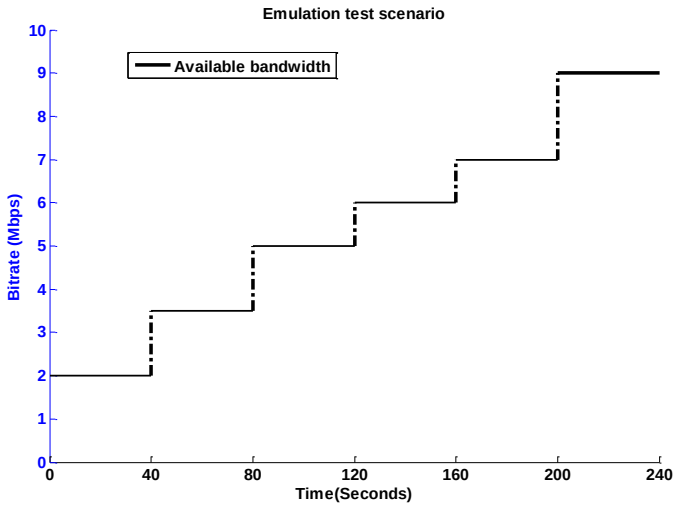


Figure 31 Emulation test scenario

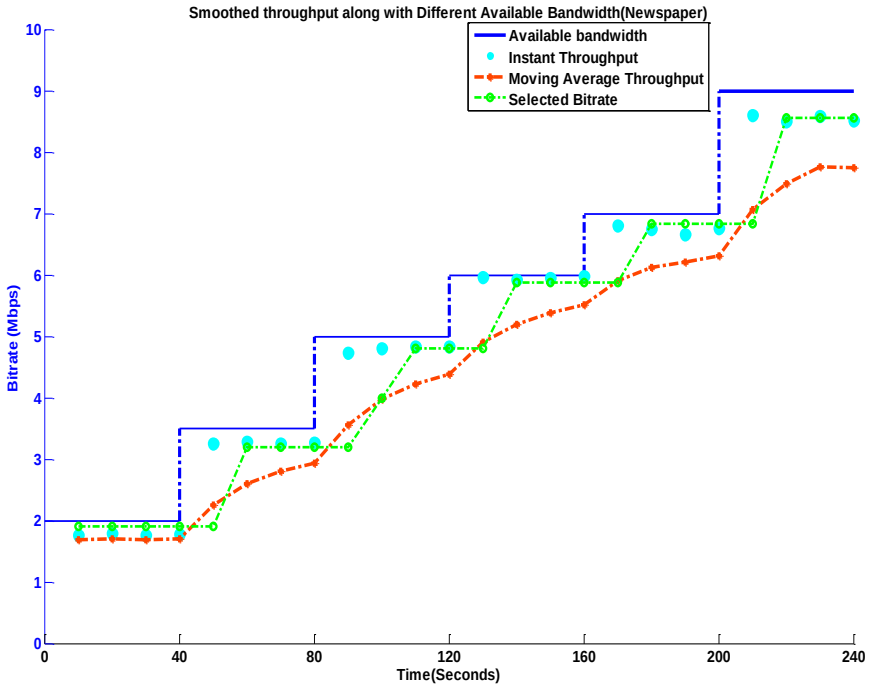


Figure 32 Smoothed throughput, along with different available bandwidths (Newspaper)

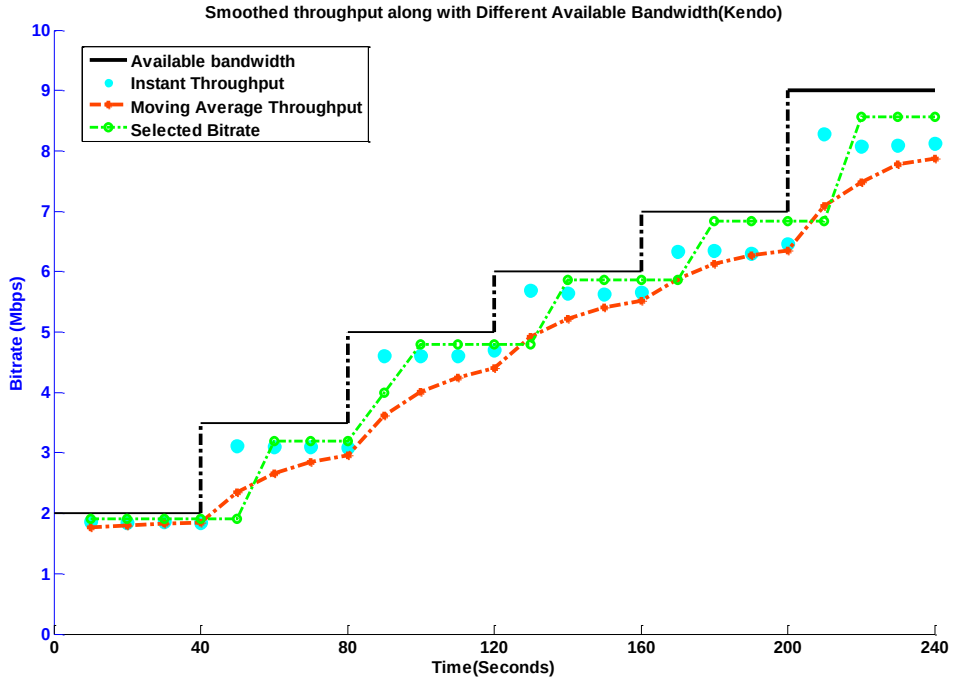


Figure 33 Smoothed throughput, along with different available bandwidths (Kendo)

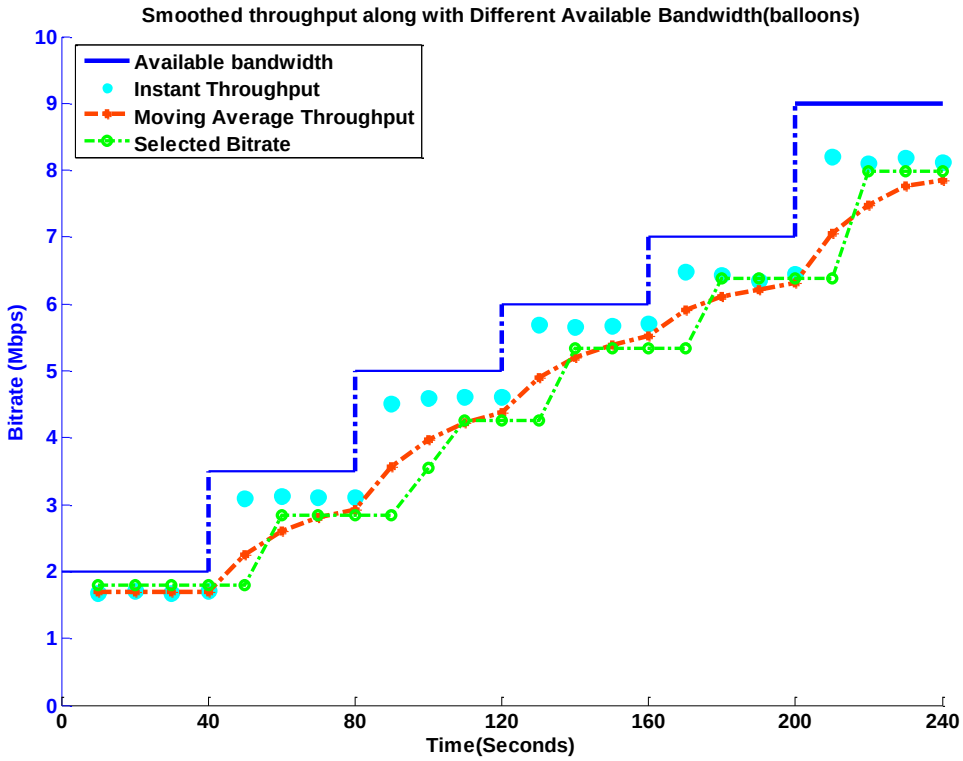


Figure 34 Smoothed throughput, along with different available bandwidths

(balloons)

In Figures 33, 34 and 35, the instant throughput, the moving average throughput and the selected video bitrate are shown, respectively. As introduced in Chapter 3,The client began the download with the lowest available video bitrate (2Mbps in this test scenario). There are eight switching operations in this emulation test scenario.

The exponential moving average of the throughput for the last four segments is used as a predictor for the appearance of the next throughput. As can be seen in each figure, the instant throughput was quickest in terms of increasing the available bandwidth. However, our algorithm selected the moving average of the last four segments as the predictor for the next segment to download, for the reason that the moving average of the last four segments can avoid the situation of a short variation in available bandwidth. This is good because it allows the video bitrate to stay stable when short fluctuations occur. Therefore, the policy can be based on the value of the moving average. When the moving average value is greater than the previous selected bitrate, a higher bitrate level can be selected for the next download segment (and vice versa). When the moving average of the last five download segment is higher than the previous download segment, the increase of video bitrate switch up (and vice versa) for the next download segement. This policy can explain why switching occurs at 50 seconds, 90 seconds, 130 seconds, 170 seconds, and 210 seconds, instead of at 40, 80, 120, 180, and 200 seconds (i.e., when the available bandwidth changes).

Another thing shown in the figures is that, at 100 seconds, the client selected 3.5 Mbps instead of 6 Mbps, after the 3 Mbps video bitrate was selected previously. This is a smoothed method of adaptation for the DASH concept. It is implemented because users prefer smoothed adaptations of video bitrate change over fluctuating adaptations.

Chapter 5: Conclusions and Future Work

In this chapter, the conclusions and findings of the thesis are presented. In Section 5.1. the conclusion of the research work is presented. The importance of this research work and the result that was achieved is presented. Section 5.2 enumerates several possible improvements for this thesis.

5.1 Conclusions

In this thesis, the architecture of our DASH-based 3D multi-view video streaming system is proposed. Two state-of-the-art techniques (HEVC and DASH) were used in our system. The importance of this work is to attempt to propose a system based on the new techniques (HEVC and DASH) for Multi-view plus Depth for 3D content. There is not much this work in this field of research. The proposed system is implemented. And it is evaluated based on the simulation and emulation test.

This thesis introduces the preparation of the scalable cloud server using the HEVC 3D extension encoder at the scalable server side. We also proposed a new type of scalability in terms of changing the number of views to adaptively stream multi-view video for the auto-stereoscopic 3D display.

Furthermore, based on objective test results, we were able to devise a policy to adaptively select different versions of bit streams, which were compressed using the MVD format at different available bandwidths in order to present the best quality for every view to the user. We also proposed a smoothed throughput-based average SSIM maximum bit stream selection algorithm for the DASH client to adaptively download

video content in order to improve the objective quality of the delivered multi-view video content.

The client side is designed to adaptively download the segments provided by the server based on different available bandwidths. The emulation results are shown in this thesis.

5.2 Future Work

In future work, we expect to improve our system in the following ways:

5.2.1 Create a higher-resolution test sequence

The trend of 3D technology is to focus on high detail and high fidelity. Right now, all of the test sequences for 3D multi-view video research is in high definition (1080p or 720p) formats; however, in 3D multi-view coding, especially in the context of view synthesis technology, the high details of transmitted views and depth maps can provide better performance for rendered views, which is good for the quality of the experience of the 3D multi-view video.

5. 2.2 Speed up the encoder

One of the limitations of the real-time 3D multi-view video is the speed of the encoder. The speed of the reference software is so slow that real-time streaming is impossible. This thesis represents the first stage of research for HEVC-based 3D multi-view video streaming. However, for the next stage of research, a faster encoder speed is needed. There are two solutions for achieving faster encoder speeds:

- Hardware solution: Speed up the encoder via the Field-Programmable Gate

Array (FPGA). This is suitable for the codec cause most of codecs are implemented through hardware. Despite the ability to achieve good performance, the expanse is high compared to the following solution.

- Software acceleration solution: Because the hybrid quad-tree structure of the HEVC codec is designed for the parallelization computation, the nature of the HEVC offers a better approach to the acceleration of GPU-based codec acceleration.

5.2.3 Subjective test of the 3D multi-view display

The limitations of subjectively testing 3D multi-view videos on 2D high definition are explained in Chapter 4. To tackle these limitations, a 3D multi-view display is proposed for the subjective test. This scenario is the same as the subjective test scenario introduced in Chapter 4. However, users can directly view the multi-view 3D content on the auto-stereoscopic display and score what they are seeing.

5.2.4 Multi-client emulation

The emulation in this thesis is based on the assumption that only one client is downloading the 3D multi-view content over the network. In future works, the emulation should consider multi-client situation, in which each client competes with one other to stream multi-view videos. Moreover, a policy and an algorithm should be modified to accommodate multiple users.

Reference

[1] P. Benzie, J. Watson, S. Member, P. Surman, I. Rakkolainen, K. Hopf, H. Urey, V. Sainov, and C. Von Kopylow, "A Survey of 3DTV Displays : Techniques and Technologies," IEEE Transaction Circuit Syst. Video Technol., vol. 17, no. 11, pp. 1647–1658, 2007

[2] A. Buchowicz, "Video coding and transmission standards for 3D television — a survey," Opto-Electronics Rev., vol. 21, no. 1, pp. 39–51, Dec. 2012.

[3] M. Tanimoto, "Free-Viewpoint Television," in Image and Geometry Processing for 3-D Cinematography, vol. 5, R. Ronfard and G. Taubin, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2010, pp. 53–76.

[4] P. Kauff, N. Atzpadin, C. Fehn, M. Müller, O. Schreer, a. Smolic, and R. Tanger, "Depth map creation and image-based rendering for advanced 3DTV services providing interoperability and scalability," Signal Process. Image Commun., vol. 22, no. 2, pp. 217–234, Feb. 2007.

[5] K. Müller, H. Schwarz, D. Marpe, C. Bartnik, S. Bosse, H. Brust, T. Hinz, H. Lakshman, P. Merkle, F. H. Rhee, G. Tech, M. Winken, and T. Wiegand, "3D High-Efficiency Video Coding for Multi-View Video and Depth Data," IEEE Trans. IMAGE Process., vol. 22, no. 9, pp. 3366–3378, 2013.

[6] C. Ozcinar, E. Ekmekcioglu, and A. Kondoz, "Dynamic adaptive 3D multi-view video streaming over the internet," in Proceedings of the 2013 ACM international workshop on Immersive media experiences - ImmersiveMe '13, 2013, pp. 51–56.

[7] G. J. Sullivan, J. Ohm, W. Han, and T. Wiegand, "Overview of the High Efficiency Video Coding," IEEE Trans. Circuits Syst. Video Technol., vol. 22, no. 12, pp. 1649–1668, 2012

[8] K. Müller, H. Schwarz, D. Marpe, C. Bartnik, S. Bosse, H. Brust, T. Hinz, H.

Lakshman, P. Merkle, F. H. Rhee, G. Tech, M. Winken, and T. Wiegand, “3D High-Efficiency Video Coding for Multi-View Video and Depth Data,” IEEE Trans. IMAGE Process., vol. 22, no. 9, pp. 3366–3378, 2013

[9] J. Ohm, G. J. Sullivan, H. Schwarz, T. K. Tan, S. Member, and T. Wiegand, “Comparison of the Coding Efficiency of Video Coding Standards — Including High Efficiency Video Coding (HEVC),” IEEE Trans. circuits Syst. video Technol., vol. 22, no. 12, pp. 1669–1684, 2012.

[10] H.-H.-I. Fruanhofer, “HEVC 3D extension Test Model(3DV HTM) version 11.0,”2013.[Online].Available:https://hevc.hhi.fraunhofer.de/svn/svn_3DVCSoftware/tags/HTM-11.0/.

[11] (MPEG) IJSW. Dynamic adaptive streaming over http. w11578, CD 23001-6, w11578, CD 23001-6. ISO/IEC JTC 1/SC 29/WG 11 (MPEG), Guangzhou, China, 2010.

[12] T. Stockhammer, “Dynamic Adaptive Streaming over HTTP – Standards and Design Principles,” in Proceedings of the Second Annual ACM Conference on Multimedia Systems (MMSYS 2011), 2011, no. i, pp. 133–143.

[13] Z. Wang, A. C. Bovik, H. R. Sheikh, S. Member, E. P. Simoncelli, and S. Member, “Image Quality Assessment : From Error Visibility to Structural Similarity,” IEEE Trans. circuits Syst. video Technol., vol. 13, no. 4, pp. 600–612, 2004.

[14] T. Wiegand, H. Schwarz, A. Joch, F. Kossentini, S. Member, and G. J. Sullivan, “Rate-Constrained Coder Control and Comparison of Video Coding Standards,” IEEE Trans. circuits Syst. video Technol., vol. 13, no. 7, pp. 688–703, 2003.

[15] Begen, A., Akgul, T. and Baugher, M. 2011. Watching Video over the Web: Part 1: Streaming Protocols. J. IEEE Internet Comput. 15, 2 (Mar. 2011), 54–63.

[16] Kuschnig, R., Kofler, I. and Hellwagner, H. 2011. Evaluation of HTTP-based

Request-Response Streams for Internet Video Streaming. In Proceedings of the second annual ACM conference on Multimedia systems. (San Jose, California, USA, February 23-25, 2011) MMSys `11. ACM, New York, NY, 245–256.

[17] Pantos, R. and May, W. 2010 HTTP Live Streaming. Internet Draft IETF Draft. IETF Tools. http://tools.ietf.org/html/draft-pantos-http-live-streaming-04

[18] Zambelli, A. 2009. IIS smooth streaming technical overview. Microsoft Corporation

[19]Hassoun, D.2010.Dynamic streaming in flash media server 3.5.Adobe. http://www.adobe.com/devnet/adobe-media-server/articles/dynstream_advanced_pt1

[20]Akamai HD Network Demo. http://wwwns.akamai.com/hdnetwork/demo/flash/zeri/

[21] Lohmar, T.; Einarsson, T.; Frojdh, P.; Gabin, F. 2011. Kampmann, M.; Dynamic adaptive HTTP streaming of live content. In Proceedings of the 12th IEEE International Symposium on a World of Wireless, Mobile and Multimedia Networks (Lucca, Italy, 20-24 June, 2011) WoWMoM ‘11. 1-8.

[22] A. Javadtalab, M. Semsarzadeh, A. Khanchi, S. Shirmohammandi, and A. Yassine, “Continuous One-Way Detection of Available Bandwidth Changes for Video Streaming Over Best-Effort Networks,” IEEE Trans. Instrum. Meas., vol. 64, no. 1, pp. 190–203, 2015.

[23] K. T. Ba and A. M. Tekalp, “ADAPTIVE STEREOSCOPIC 3D VIDEO STREAMING,” in Image Processing (ICIP), 2010 17th IEEE International Conference on, 2010, pp. 2409–2412.

[24] A. Hamza and M. Hefeeda, “A DASH-based Free Viewpoint Video Streaming System,” in Proceedings of Network and Operating System Support on Digital Audio and Video Workshop, 2013, p. 55.

[25] K. Calagari, “Anahita : A System for 3D Video Streaming with Depth

Customization Categories and Subject Descriptors," in Proceedings of the ACM International Conference on Multimedia, 2014, pp. 337–346.

[26] K. E. Psannis, M. Hadjinicolaou and A. Krikelis, "MPEG-2 Streaming Of Full Interactive Content", IEEE Transactions on Circuits and Systems for Video Technology, vol.16. no 2, pp. 280-285, 2006.

[27] T. Schierl, M. M. Hannuksela, Y-K. Wang, and S. Wenger, " System Layer Integration of High Efficiency Video Coding (HEVC)," IEEE Transactions on Circuits and Systems for Video Technology, vol. 22, issue 12, pp. 1871-1884, 2012,

[28] S. Wenger, "H.264/AVC over IP," IEEE Transactions on Circuits and Systems, vol. 13, no. 7, July 2003.

[29]H. Roodaki, M.R. Hashemi, and S. Shirmohammadi, "A New Methodology to Derive Objective Quality Assessment Metrics for Scalable Multi-view 3D Video Coding", ACM Transactions on Multimedia Computing, Communications, and Applications, Vol. 8, No. 3S, Article 44, September 2012, 25 pages.DOI: 10.1145/2348816.2348823

[30] H. Roodaki, M.R. Hashemi, and S. Shirmohammadi, "Rate-Distortion Optimization for Scalable Multi-View Video Coding", Proc. IEEE International Conference on Multimedia and Expo, Chengdu, China, July 14-18 2014, 6 pages.DOI: 10.1109/ICME.2014.6890275

[31] B. Oztas, M. T. Pourazad, P. Nasiopoulos, I. Sodagar, and V. C. M. Leung, "A Rate Adaptation Approach for Streaming Multiview Plus Depth Content," in Computing, Networking and Communications (ICNC), 2014 International Conference on, 2013, no. Mvd, pp. 1006–1010.

[32] E. Bosc, F. Racapé, V. Jantet, P. Riou, M. Pressigout, and L. Morin, "A study of depth/texture bit-rate allocation in multi-view video plus depth compression," Ann. Telecommun. - Ann. Des Télécommunications, vol. 68, no. 11–12, pp. 615–625, Apr.

2013.

[33] B. Li, H. Li, L. Li, and Z. Jinlei, “Rate control by R-lambda model for HEVC,” Jt. Collab. Team Video Coding(JCT-VC)of ITU-T SG 16 WP 3 ISO/IEC JTC 1/SC 29/WG 11, pp. 1–11, 2012.

[34]H. Roodaki, M.R. Hashemi, and S. Shirmohammadi, “New Scalable Modalities in Multi-view 3D Video”, Proc. ACM Workshop on Mobile Video, Oslo, Norway, February 27 2013, pp. 25-30.DOI: 10.1145/2457413.2457420

[35] P. Ndjiki-nya, M. Köppel, D. Doshkov, H. Lakshman, P. Merkle, K. Muller, and T. Wiegand, “Depth Image-Based Rendering With Advanced Texture Synthesis for 3-D Video,” IEEE Trans. Multimed., vol. 13, no. 3, pp. 453–465, 2011.

[36]Dimenco, “Non-glass 3D displayer,” 2014. [Online]. Available: http://www.dimenco.eu/3d-displays/displays/65-inch-4k/.

[37]Alioscopy, “Alioscopy 3D HD 55" LV data sheet,” 2010. [Online]. Available: http://www.alioscopy.com/en/datasheet.php?model=Alioscopy 3D HD 47%22 LV.

[38] T. L. at N. University, “Kendo Test sequences.” [Online]. Available: http://www.tanimoto.nuee.nagoya-u.ac.jp/.

[39] Y.-S. Ho, E.-K. Lee, and L. Cheon, “Newspaper, Multiview Video Test Sequence and Camera Parameters,” in INTERNATIONAL ORGANISATION FOR STANDARDISATION ORGANISATION I ISO / IEC JTC1 / SC29 / WG11 CODING OF MOVING PICTURES AND AUDIO, 2008, pp. 1–6.

[40] “Call for Proposals on 3D Video Coding Technology.” ISO/IEC JTC1/SC29/WG11 MPEG2011/N12036, Geneva, Switzerland, 2011.

[41] Rizzo, L. 1997. Dummynet : a simple approach to the evaluation of network protocols. ACM SIGCOMM Computer Communication Review. 27, (1997), 31–41

[42] Internet Information Service, 2014. [Online]. Available: http://www.iis.net.

[43] T. C. Thang, H. T. Le, S. Member, A. T. Pham, S. Member, and Y. M. Ro, “An Evaluation of Bitrate Adaptation Methods for HTTP Live Streaming,” IEEE J. Sel. ATREAS Commun., vol. 32, no. 4, pp. 693–705, 2014

[44] R.Szeliski. "Epipolar Geometry" in the book of Computer Vision: algorithms and applications, chapter11, pp.471-473.

[45]W. J. Tam, F. Speranza, S. Yano, K. Shimono, and H. Ono, “Stereoscopic 3D-TV: Visual comfort,” IEEE Trans. Broadcast., vol. 57, no.2, 2011.

[46]A.Fernando, S.T.Worral, E.Ekmekcioglu, “3D-TV: Processing and transmission of 3D video signals,” book, wiley.

[47]ImagefromWiki:http://en.wikipedia.org/wiki/File:Parallax_barrier_vs_lenticular_screen.svg.

[48]Jason Geng, “Three-dimensional display technologies”,Advances in Optics and Photonics, Vol. 5, Issue 4, pp. 456-535 (2013),http://dx.doi.org/10.1364/AOP.5.00045

[49] R. Hartley and A. Zisserman, Multiple View Geometry in Computer Vision. Cambridge,U.K.: Cambridge Univ. Press, 2000.

[50] I.-R. R. BT.500.13, “Methodology for the subjective assessment of the quality of television pictures,” Int. Telecommun. Union, vol. 13, 2012.

[51]M. Carbone and L. Rizzo. Dummynet revisited. SIGCOMM CCR, vol. 40, n. 2. 2010, April.Available from http://info.iet.unipi.it/~luigi/papers/20100304-ccr.pdf.

[52]M. Carbone and L. Rizzo. An emulation tool for PlanetLab. 2010, February. Available from http://info.iet.unipi.it/~luigi/papers/20100316-cc-preprint.pdf.

[53]Magor,“Aerus,”www.magorcorp.com,2014.[Online].Available:www.magorcorp.com.

[54] Download from http://www.fujii.nuee.nagoya-u.ac.jp/multiview-data/