# QoE Aware Multi-view Video Transmission over P2P Networks

by

Saadet Sedef Savaş

A Thesis Submitted to the
Graduate School of Engineering
in Partial Fulfillment of the Requirements for
the Degree of

Master of Science
in
Electrical and Computer Engineering

Koc University

September 2011

Koc University

Graduate School of Sciences and Engineering

This is to certify that I have examined this copy of a master's thesis by

Saadet Sedef Savaş

and have found that it is complete and satisfactory in all respects,

and that any and all revisions required by the final

examining committee have been made.

Committee Members:

______________________________

Prof. A. Murat Tekalp (Advisor)

______________________________

Assoc. Prof. Öznur Özkasap

______________________________

Assoc. Prof. Çağatay Başdoğan

Date: ____________________

***to my family***

# ABSTRACT

Multi-view Video (MVV) is the next step in 3D evolution and quality of experience (QoE) aware MVV transmission is a hot research topic. In MVV transmission, the required bit rate depends on the number of views transmitted therefore unlike the previous video standards, it is difficult to operate over channels with fixed bandwidth capacity; making the Internet Protocol (IP) the natural choice for 3D video transmission. IP provides varying channel capacity and allows users to receive content at different bit rates according to their 3D displays setup. We addressed two major problems of IP: i) scalability problem, ii) varying amount of available bit-rate problem.

In this thesis, we have adopted an adaptive MVV streaming mechanism over Peer-to-peer (P2P) networks to address both of these challenges. The P2P approach addresses the scalability problem by distributing the task for delivery 3D media among users; allowing a more scalable solution. The adaptive streaming approach addresses the unreliable channel problem by adjusting the content to the available channel properties. However, properly addressing the perceived quality of experience (QoE) of MVV is very crucial in order to achieve a successful adaptive P2P MVV service. Therefore, we have particularly focused on QoE of MVV that is subject to typical IP network failures such as packet loss and varying channel capacity. Our first goal is to reduce the effect of the errors generated due to the packet losses during transmission by utilizing forward error correction (FEC) algorithms and concealment methods. The optimal allocation of FEC packets among views is investigated in order to achieve a better MVV perception. The subjective test results indicate that asymmetric allocation of FEC packets results in better perceived quality. Next, we evaluated the perceived quality of MVV, when various adaptation methods are adopted to match MVV bit-rate to a given rate. For this purpose, we have conducted subjective tests to evaluate the visual distortions caused by scaling the content rate (adaptation methods). These tests showed that in order to obtain the best perceived video quality, intermediate views should be scaled asymmetrically as much as possible. Once asymmetric coding threshold (~32dB) has been reached then it is possible to drop all the views between the first and the last one and interpolate them using depth image-based rendering (DIBR) technique. Finally, we propose a mesh-based P2P streaming architecture that employs rate adaptation according to the findings to deliver the best QoE under diverse network conditions.

# ÖZET

3 boyutlu (3B) görüntü teknolojilerinin bir sonraki adımı Çoklu-Görüntülü Videodur (ÇGV). Görsel deneyim kalitesine duyarlı ÇGV iletimi günümüzün güncel bir araştırma konularında bir tanesidir. ÇGV iletiminde gereken bant genişliği gönderilen görüntü sayısına bağlıdır, bu yüzden daha önceki video standartlarından farklı olarak, sabit bant genişliğine sahip kanalları kullanmak zordur. Dolayısı ile 3B video iletimini değişken bağlantı kapasitesi sağlayabilen İnternet protokolü (IP) üzerinden yapılması makul bir tercihtir. IP sayesinde kullanıcılar bağlantı hızlarına göre farklı sayıda görüntüyü alabilirler. Bu çalışmada IP'nin iki temel problemine değindik: i) ölçeklenebilirlik problemi, ii) değişken miktarda mevcut bit hızı problemi.

Yukarıdaki problemlere çözüm olarak eşten eşe (P2P) ağlar üzerinden ÇGV iletimi uygulanabilir. P2P ağlarını kullanmak, 3B video gönderme görevini eşlere dağıttarak ölçeklenebilir bir iletim sisteminin oluşmasına imkan sağlar. Ayrıca, iletim sisteminin uyarlanabilir olması, içeriğin mevcut kanal özelliklerine göre şekillenmesini mümkün kılarak, değişken kanal kapasitesine uygun akıtım yapılmasına olanak verir. Gene de, başarılı ÇGV iletim hizmeti sağlanabilmesi için algılanan deneyim kalitesinin detaylıca incelenmesi gerekmektedir. Bu çalışmada paket kaybı ve değişken kanal kapasitesi gibi tipik IP ağ hatalarından etkilenen ÇGKların deneyim kalitesi üzerine odaklandık. İlk amacımız, transfer sırasındaki paket kayıplarından kaynaklanan hataların, ön hata düzeltme (FEC) algoritmalarıyla ve hata gizleme methotlarıyla etkisini azaltmaktı. Daha iyi bir ÇGV algısı için FEC paketlerinin görüntüler arasında optimum paylaştırılması incelendir. Görsel test sonuçları bu paketlerin asimetrik paylaştırılmasının daha kaliteli bir algı sağladığını ortaya çıkardı. Sonrasında, 3B görüntünün kanalın değişken veri hızına uyarlanması için farklı adaptasyon metotları önerdik. Ayrıca bu metotları izleyici tarafından algılanan kalite üzerinden değerlendirdik. Bu amaçla içerik boyutunu ölçeklemekten kaynaklanan görsel bozuklulukları görsel testlerle değerlendirdik. Bu testler, iyi bir video kalitesi elde etmek için ara görüntülerin mümkün olduğu kadar asimetrik olarak ölçeklendirilmesi gerektiğini gösterdi. Eğer yeterli olmaz ise, asimetrik kodlama eşiğine (~32dB) ulaşıldığında, tüm ara görüntülerin iletiminin durdurulmasının ve sadece ilk ve son görüntüye ek olarak pixel derinlik bilgisinin gönderilmesinin en yüksek kaliteyi sağladığı saptandı. Yukarıdaki sonuçlara ek olarak, bu çalışma ÇGV'nun iletimi için Torrent-tabanlı bir iletim sistemi önermektedir.

# ACKNOWLEDGEMENTS

I would like to thank Prof. Tekalp who has guided and inspired me with his profound knowledge and genuine support during my graduate study.

I am grateful to members of my thesis committee for their valuable comments.

I would like to express my intimate thanks to C. Göktuğ Gürler for always helping me to overcome the problems. He had important contributions on this thesis by the experience and knowledge he provided.

Also, I would like to thank TÜBİTAK for providing support for my research.

Above all, I would like to present my deepest thanks to my father and mother for always supporting, encouraging and watching over me through my all life.

## TABLE OF CONTENTS

## LIST OF TABLES

## LIST OF FIGURES

## NOMENCLATURE

| | |
|---|---|
| *HVS* | Human Visual System |
| *QoE* | Quality of Experience |
| *PSNR* | Peak Signal to Noise Ratio |
| *FEC* | Forward Error Correction |
| *JSVM* | Joint Scalable Video Model |
| *MVC* | Multi-View Coding |
| *MVV* | Multi-View Video |
| *NALU* | Network Abstraction Layer Unit |
| *QP* | Quantization Parameter |
| *DIBR* | Depth Image Based Rendering |
| *SVC* | Scalable Video Coding |
| *TCP* | Transmission Control Protocol |
| *UDP* | User Datagram Protocol |
| *MVD* | Multi-view video plus depth |
| *LDV* | Layered Depth Video |
| *JVT* | Joint Video Team |
| *IP* | Internet Protocol |
| *RD* | Rate-Distortion |
| *DTV* | Digital TV |
| *DVB* | Digital Video Broadcasting |
| *DSCQS* | The double stimulus continuous quality scale |

# Chapter 1

# INTRODUCTION

## 1.1 Overview

Stereoscopic video has already made an impact on the multimedia industry. The market share of stereoscopic 3D movies in Hollywood has increased significantly over the last few years. More recently, stereoscopic TV broadcasts over DVB have begun in the UK in April 2010 using a frame-compatible format, which combines right and left views of a stereoscopic video pair in a single HD video frame. 3D-compatible TV sets that can display stereoscopic video using various technologies, such as using polarization or time-shutter glasses, are available in the market and the demand for them is increasing exponentially as the content amount increases. Meanwhile, standards for transmitting 3D media signals over peripherals, such as HDMI 1.4a, have been defined.

We are now progressing to the next phase in 3D media services, which will be based on Multi-View-Video (MVV) formats. MVV technology first takes its place in theatres as in stereoscopic video then Multi-view video applications will be available to the mass consumer market as well. MVV enables viewing a scene in 3D from multiple angles (within a viewing cone), which may make it possible to see behind an object by tilting the head; whereas in stereoscopic video, the viewer can only see what has been captured from a single viewpoint. Each view position provides different stereo perception from adjacent two views.

Moreover, MVV displays are auto-stereoscopic which do not require wearing special 3D glasses.

Besides the perceived video quality and standardization problems, the major difficulty with MVV is that high bitrate requirement. There are many studies to reduce the required bitrate for MVV. An annex called multi-view video coding (MVC) has been appended to H.264/AVC standard for encoding multi-view video (MVV) with high efficiency [1][2]. Researches also aimed to decrease the overall transmission rate by exploiting the human visual system (HVS) that can tolerate lack o f high frequency components in one of the views. Even if the bandwidth problem could have been alleviated up to a manageable level, the issue of display dependent input video formats, varying amount of bitrate requirement, makes transmission over fixed bit-rate channels, such as the DVB, unfeasible.

Transmission of MVV over IP (Internet Protocol) seems the most flexible solution for 3D media delivery, which can provide different transmission rates to different users according to their available connection rate and display technology. In addition, it is also possible to receive feedback from the user over an IP channel, enabling effective rate adaptation in addition to personalized services such as user centric advertisements or interactive TV. The MVV over IP can be offered as a stand-alone service or as a supplement to broadcast of stereoscopic video over DVB.

## 1.2 Problem Definition

One of the biggest challenges for MVV content delivery is to support wide range of 3D displays that require different number of views (and possibly different bitrates); making it difficult to use fixed bit-rate channels, such as DVB. DVB can be suitable for transmitting stereoscopic 3D where the bitrate is fixed e.g., using frame based methods (such as horizontal multiplexing) [3].

Transmission of MVV over IP seems the most flexible solution for 3D media delivery because IP can provide different transmission capacities to different users.

It is well-known that video over IP has some limitations. Two major challenges must be addressed in order to establish successful 3D video services over IP.

First, IP operates in the best effort sense; hence, it is possible to experience packet losses and varying amount of delays between end nodes. The second challenge is the high bandwidth requirement in server-client based solutions; making it difficult to achieve a scalable distribution system against increasing number of recipients. This problem is intensified with high bandwidth requirements of MVV format. Fortunately, it is possible to alleviate the former problem by using adaptive streaming and scalable video coding techniques, and the latter problem by using a peer-to-peer streaming architecture which distributes the burden of data dissemination over peers and utilizes their upload capacity.

This paragraph will present the first problem in detail. The recent Internet infrastructure is not adequate for transmitting all MVV views at highest quality due to insufficient bandwidth and its fluctuation property. Always sending views with associated per-pixel depth maps as in the ATTEST project is not the best option, because high quality outputs may not be generated due to the occlusions or high disparity in the scenes [5]. Instead, in an adaptive streaming solution, the video send-rate is dynamically adjusted according to the available user link capacity. How to degrade the bitrate a.k.a video quality is an open research area. Deciding the best degrading strategy benefiting from the subjective visual quality tests is one of our focuses in this thesis. This degrading policy forms one of the two important criteria that affect the perceived video quality of streamed MVV. The other criterion is packet losses. Although an adaptive streaming solution alleviate the packet loss problem, the packets can still be lost or delayed and if the delay is long enough such that the play out time for the video-segment have passed, both considered as lost data.

Forward error correction (FEC) methods can be employed as a remedy to packet losses as they can recover lost packets without requiring repeated requests from the client side at the cost of augmented bitrates requirement. Fortunately, it is possible to decrease the video bitrate by encoding at lower quality and create a gap to channel coding. For the remaining unrecoverable lost packets, error concealment methods should be employed. In this thesis, the effect of packet losses and different concealment techniques over the perception of both stereoscopic and multi-view video is evaluated by conducting subjective tests.

To overcome the second problem which is scalability, we propose an adaptive P2P solution. The MVV specific changes like piece and peer scheduling on a Bittorent-like P2P system will be presented in the contributions section.

## 1.3 Related Works

### 1.3.1 3D Video Formats

Simulcast (sequential) is the most common stereo or multi-view video representation where views are encoded independent of each other. Since it does not exploit the interview dependencies, this method has low complexity and it is also backward compatible with the 2D displays [4]. There are also frame compatible stereo video formats that make 3D video transmission over DVB possible. The main idea is to subsample views in order to halve the size of the 3D frames to fit it into a regular 2D frame size. After the subsampling, the subsampled frames are multiplexed into a frame. Different subsampling patterns exist such as side-by-side, top and bottom, line interleaved and checkerboard [3]. Video plus depth (V+D) format consists of a video signal and its per-pixel depth map. The other view is rendered by 3D warping at the decoder [4].

MVD (Multi-view video plus depth) and LDV (layered depth video) format were introduced to support multi-view auto-stereoscopic displays.

***Multi-View-Video-Plus-Depth Format***

The view-plus-depth format was first proposed for the case of stereo video in the European project Advanced Three-Dimensional Television System Technologies (ATTEST) [5]. In this representation, views are enhanced with single channel depth information which can be used to render artificial intermediate views. MPEG has specified a container format for view-plus-depth stereo video in “ISO/IEC 23002-3 Representation of Auxiliary Video and Supplemental Information” also called MPEG-C Part 3 [6][7]. This format has later been extended to multi-view-video-plus-depth (MVD) [8], which consists of N-views plus N-depth maps to render intermediate views at the receiver side.

***Layered depth video***

Layered depth video (LDV) [9] is based on MVD and it is alternative to MVD. This format reduces bit-rate by sending only the central view the background information and its depth of the other views. The background information which has the disoccluded information is derived from the other views by comparing pixel-wise to determine which parts belong to the background. This format models occlusions better than MVD.

***Depth Enhanced Stereo***

Depth enhanced stereo is extending high quality stereo video with additional depth and possibly occlusion layers. It provides backward compability with stereo systems, easily usable for stereoscopic glasses-based consumer displays, and also support for multi-view auto-stereoscopic displays. This format has a flexible structure such as more than two views can be added or depth maps can be discarded [10] [11].

### 1.3.2 3D Video Perception and Subjective Evaluation Methods

Although 2D objective assessment models that are highly correlated are defined, an objective metric that reflects the Human Visual System for 3D systems is not developed [12]. The deficiency of a reliable objective quality metric for 3D video makes subjective video quality evaluation the most trustworthy way to assess 3D video quality. This fact proves the necessity of considering QoE concept in every stage of the 3D video transmission systems.

***3D artifacts***

Many artifacts may cause reduction of user experience like signal processing errors, packet losses, occlusion problems, errors in display [19]. Content creation, representation format, coding, transmission, post processing and visualization processes are the main source of artifacts [15]. Stereoscopic artifacts that can be seen in Table 1.1 are classified in four groups in a previous work by Boev, et. al; namely structure, color, motion and binocular [15][16]. Structure artifacts demonstrate the problems that are related to the perception of the structure (contours and texture) of the images. Color and motion artifacts are related to the vision of color and motion. Binocular artifacts contain the artifacts that are special to the stereoscopic perception [13].

| | Capture | Representation | Coding | Transmission/ Error Resilience | Visualization |
|---|---|---|---|---|---|
| Structure | - defocusing blur<br>- barrel distortions<br>- interlacing<br>- aliasing | - temporal and spatial aliasing<br>- line replication | - blocking artifacts<br>- mosaic pattern<br>- ringing | - data loss<br>- data distortion<br>- jitter | - flickering<br>- aspect ratio and display geometry distortions |
| Color / Motion | - Chromatic aberration<br>- motion blur<br>- temporal mismatch | - temporal and spatial aliasing | - color artifacts<br>- color bleeding<br>- motion compensation artifacts | - color bleeding<br>- loss/distortion in motion<br>- jitter | - smearing<br>- viewing angle dependant colour representation<br>- rainbow artifact |
| Binocular | - depth plane curvature<br>- keystone-distortion<br>- cardboard effect | - ghosting by disocclusion<br>- Perspective-binocular rivalry | - cross distortions<br>- cardboard effect<br>- depth "bleeding"/depth "ringing" | -data loss, one channel<br>- data loss, propagating | - ghosting by crosstalk<br>- angle dependant binocular aliasing<br>- puppet theater effect |

Table 1.1: Classification of stereoscopic artifacts

Disparity problems caused by calibration and rectification are the main artifact of content creation process. Disocclusion artifacts are caused by the representation of the video as video plus depth format since the occluded parts in the views are interpolating during the view rendering process [16]. Blocking artifact of DCT based coding techniques are causing the following artifacts that degrade 3D video quality; block-edge discontinuities, color bleeding, blur and staircase artifacts [17]. Artifacts due to transmission errors are strictly related with the usage of error resilient coding, concealment methods and the channel coding. Both in DVB and the Internet, burst errors and packet losses are common [14]. If inter-view prediction is employed, the errors in one view are propagated to the predicted ones as well.

### *3D Video Perception*

So far, the results about the perception of stereoscopic 3D video have revealed that human visual system (HVS) can tolerate lack of high frequency components in one of the views to a certain level. This result suggests asymmetric quality distribution among views [20][21] and can be exploited for stereoscopic video streaming purposes [22] and researches aimed to decrease the overall transmission rate by exploiting the human visual system (HVS)'s this feature. Hence, one of the views may be presented at a lower quality without degrading the 3D video perception. This approach is known as asymmetric coding and very recent results can be found in [21] [23]. In the literature, it is reported that HVS can compensate the low quality pair but whether it can compensate the distortion due to packet losses is not known. Similar studies for multi-view video quality assessment are still in its early stages and so far there is no convincing QoE optimized communications system thus it remains an open research area. A few initial studies have been presented on multi-view video streaming over P2P networks [24].

### *Qoe Measurements*

QoE in 3D refers to the quality of the received video and includes the depth sense and visual comfort factors in addition to the metrics like fidelity in monoscopic video. Unlike fidelity, it is very difficult to quantify sense of depth or comfort factors therefore the most reliable method to assess the QoE in 3D is to perform subjective assessment tests. Advancement of 3D video visual quality is affected adversely from the lack of reliable objective metrics [25]. Human observers evaluate the video quality in subjective tests, which requires more time and effort.

Different subjective assessment methods are described in the ITU-R BT.1438 recommendation [26][27]. There are mainly: DSIS-Double Stimulus Impairment Scale, DSCQS-Double Stimulus Continuous Quality Scale, SCM-Stimulus Comparison Method, SSMSingle Stimulus Method, SSCQM-Single Stimulus Continuous Quality Evaluation, requires many observers that evaluate image quality [28]. DSCQS method is especially suitable for evaluation of 3D video pairs. In this method, the subjects are shown one test sequence and reference video consecutively twice. Observers are expected to rate the overall perceived quality of the sequences on a continuous rating scale. The assessment results show the score difference between the reference and test images. For attaining better standards, ITU-R WP6C is working towards the identification of requirements for the broadcasting and subjective testing of 3DTV [29]. ITU-T Study Group 9 announced that its scope will contain 3D video quality [30].

### 1.3.3 Multi-View Video Coding

There are two main options using open codec standards for adaptive streaming: i) *Simulcast encoding* using scalable extension of H.264/AVC, called SVC, which allows SNR, temporal or spatial scalability. ii) *Dependent encoding* using the multi-view extension of H.264/AVC, called MVC, which can exploit inter-view redundancies and provide higher encoding efficiency compared to simulcast coding. Note that the depth maps associated with each view may or may not be encoded as auxiliary data in both options. More detailed information on 3D video coding can be found in [3].

In simulcast coding, each view is coded independently without exploiting similarities between views. This approach allows independent transmission and decoding of streams, eliminating possible complexities especially in P2P solutions. One option to achieve scalable coding is to utilize SVC extension of H.264/AVC standard that provides spatial, temporal and quality scalability. When compared against the H.264/AVC standard, the SVC extension provides better QoE in the case of limited resources such as link capacity, processing power and display size. It has backward compatible syntax that has been standardized by the Joint Video Team (JVT) of the ITU-T VCEG and the ISO/IEC MPEG [31]. For backward compatibility, SVC has a layered structure with base layer that is compliant with the H.264/AVC syntax and discardable enhancement layer that increases the quality in either one the scalability dimensions.

MVC extension of H.264/AVC offers the best compression efficiency for MVV by dependent coding among views to exploit similarities between them [32]. It features flexible prediction structures, allowing frames to be predicted from neighbor frames in both time and view dimension (full prediction) [33][34]. In the other extreme, it is possible to remove all inter-view dependencies, resulting in simulcast coding. One commonly used prediction structure is simplified prediction scheme that restricts inter-view prediction to certain time instants. It provides similar rate-distortion (RD) performances with full prediction [35] with far fewer inter-view dependencies. The encoding efficiency of MVC is highly affected by lightning conditions, camera orientation (disparity) and noise.

In multi-view-plus-depth encoding, selected views and associated depth maps can be can either simulcast or dependently encoded using non-scalable or scalable codecs. It is also possible to exploit correlations between the texture video and associated depth maps. For example, in [36] SVC is employed to compress texture videos and associated depth maps jointly, where up to 0.97 dB gain is achieved for the coded depth maps, compared with the simulcast scheme. Joint coding approaches most commonly target sharing some entities between the color (view) component and the depth component, such as the motion vectors. Nevertheless, there are some handicaps in making use of shared motion vector information between the two components. One of them is that the motion vectors computed during rate-distortion optimization process are selected to minimize the energy of the texture residual, which does not show 100% correlation for two components. In addition, the motion in the third dimension also affects the luminance of depth pixels and this is why there is the need for compensating the motion in Z-direction for depth maps. A previous work by Kamolrat, et. al, has investigated the utilization of motion search in the third dimension (in addition to 2D motion estimation) to further increase the block based depth map coding performance [37]. A previous research work had also utilized SVC in the scope of coding backwards-compatible view-plus-depth map content, such that the view was put in the base layer, and the depth map was put in the enhancement layer without inter-layer prediction [38].

### 1.3.4 3D Video Delivery

Transmission of 3D media, to end users with varying 3D display terminals and bandwidths is one of the biggest challenges to bring 3D media to the home and mobile devices. There are two main platforms for 3D video delivery: digital television (DTV) platforms e.g., DVB in Europe and the Internet Protocol (IP) platform in the form of IPTV or WebTV, as depicted in Figure 1.1. For the stereoscopic case both approaches have established services. For the multi-view 3D case however, content delivery has been performed by various research groups using only IP network. The major drawback of DTV is the lack of variable bitrate allocation, which is a requirement for multi-view video streaming for varying display types with different number of views.

#### *3D Stereoscopic Video Delivery*

The service of stereoscopic 3D video delivery in IP and DTV is quite new. Digital Video Broadcasting (DVB) is a suite of open standards for DTV, which has been used to broadcast stereo video using frame-compatible formats. Sky TV in England has adopted the standard and started broadcasting Premier League matches in 3D at HD resolution in 2010. The service has drawn significant attention, thousands of fans wearing 3D glasses in the pubs, showing the great interest in public for 3D entertainment.

Similarly, in 2009 July, YouTube engineers have announced that users may watch 3D contents. The user interface allowed various 3D representation formats, such as passive glasses (Red/Cyan, Green/Magenta) and also special 3D display types such as lenticular sheet or parallax barrier in Interleave mode and side by side format. It even has a feature that enables watching 3D content without any special hardware by just crossing your eyes as shown in Figure 1.2.

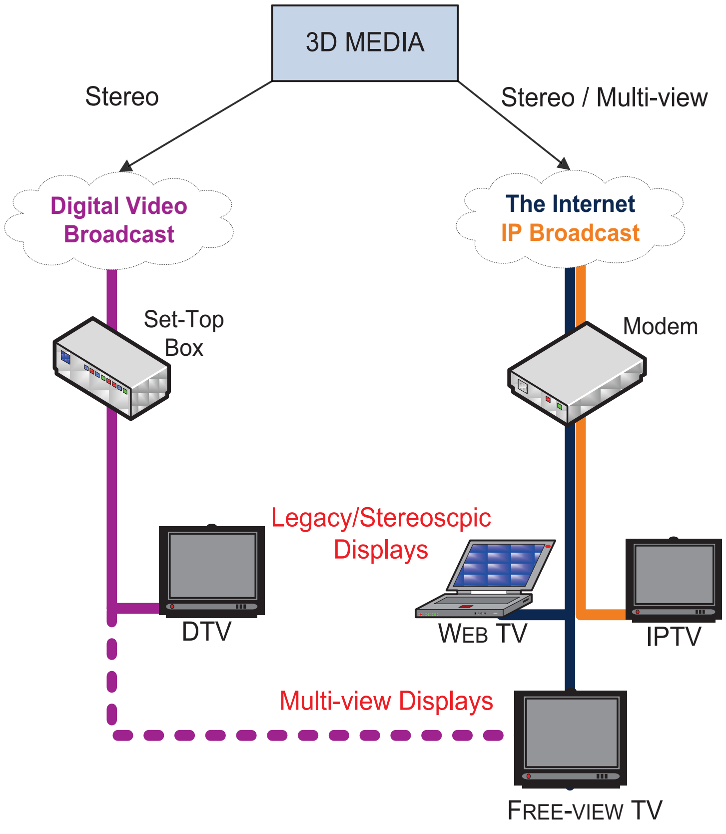



Figure 1.1: Platforms for 3D media transport.

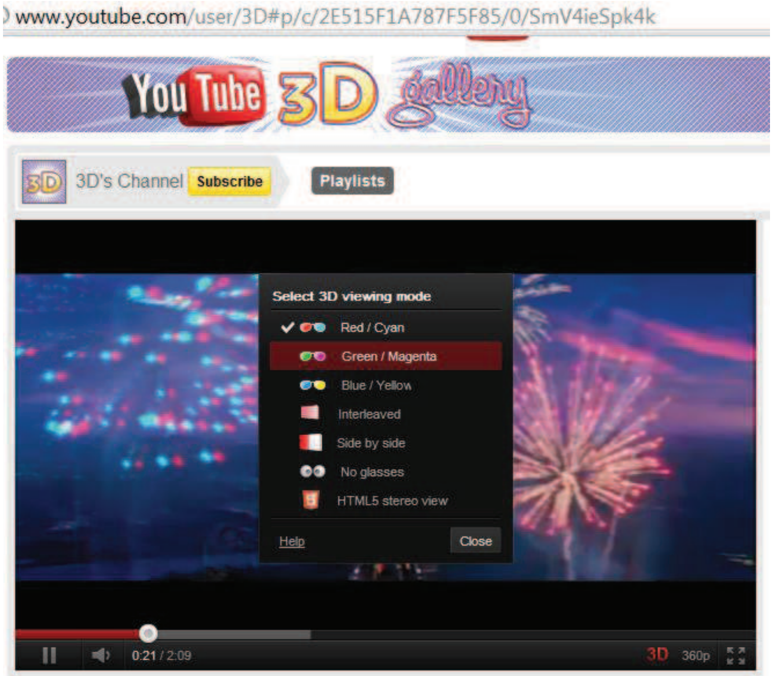



Figure 1.2: 3D Video over YouTube

***Multi-view Video Delivery***

Although there have been several research groups that target delivery of Multi-view video, so far there is no established service for two major reasons. First, there is no standard for multi-view video representation for 3D displays. It is very difficult to achieve such standard with so many different techniques to display multi-view video correctly. In the stereoscopic case, the view representation formats over HDMI connection has been standardized, allowing use of various stereoscopic displays over special hardware [39]. The second challenge is the high bit rate requirement, hindering the deployment of multi-view video transmission services from becoming feasible.

Nevertheless, several research groups have established their prototype streaming applications. In European project 3DTV [40], Kurutepe et al. has built an architecture for streaming multi-view video using server-client architecture and open standard protocols such as RTP, RTSP and SDP [41]. The authors have proposed minor modifications to these standards to enable streaming multi-view video that has been encoded using MVC. Their major contribution is the identification of inter-view dependencies, enabling intelligent stream selection for different display types.

Following server-client based approaches, Kurutepe et al. has studied the P2P solutions in order to distribute the burden of high bitrate requirements over the users [41]. In this study, the authors have proposed serving each view over a separate tree based P2P network. Thus, a user with display that needs n views can subscribe to n trees to receive the content. However, this work was only at preliminary stage. Neither of these works has considered the quality of experience which is a significant issue in 3D video streaming.

Finally, we have proposed torrent based streaming. We have employed Scalable video coding (SVC) instead of MVC, although it provides better compression efficiency, to make the system adaptive in terms of the required views of different displays. Another reason not to employ MVC is the only way to deliver adequate number of views is to use depth rendering. In this case the views are separated, thus MVC does not provide much gain and it does not provide scalability as well. By exploiting SVC, discardable chunks are formed to degrade the video quality when necessary. MVV specific piece scheduling algorithm was proposed.

### 1.3.5 3D Display Systems

There are different types of displays that can be used to produce the 3D effect. 3D display technologies are presented in Figure 1.3. The display systems that require observers to wear special glasses are called stereoscopic displays and the displays that do not require the use of glasses are called auto-stereoscopic displays. We will only represent a part of stereoscopic and auto-stereoscopic displays which is widely adopted by 3D systems today [43]. Polarized displays are the most common used type in stereoscopic displays and used in 3D cinema today. Two images are projected onto the same screen through polarizing filters. Viewer's glasses have these filters as well and they only pass the similarly polarized light. By this way each eye only sees one of the images. Two type of polarizing filters are employed: plane or circular. In plane-polarized displays, ghosting and crosstalk artifacts are possible if the observers are not in the correct position [44].Circular polarized displays solve the problems caused by the viewer position. Intensity of the views also decreases due to the polarizers. RealD cinema systems are using circular-polarization. Another method for stereoscopic displays is time-sequential displays which use shutter glasses. These glasses ensure that only one eye can see an image each frame [45]. A shutter blocks light from the appropriate eye when the image for the other eye is projected on the screen. It keeps opening and closing the shutters in synchronization with the display. Meanwhile, the display alternately displays left and right images by adopting alternate-frame sequencing technique. Xpand 3D cinema system uses shutter glasses [46]. Today's commercial auto-stereoscopic displays adopt parallax barriers or lenticular sheets placed on top of LCD screens [47]. Parallax barriers are placed on top of the LCD displays. Parallax displays block light from certain pixels to certain direction. Viewers should stay at a particular place to get the views correctly. The disadvantages of this display are to reduce both brightness and sharpness of the image. Lenticular sheet is another 3D display technology that depends on a linear line of lenses which divert the light and project a subset of display pixels to each eye [48]. The comparison of lenticular sheet and parallax barrier is depicted in Figure 1.4.

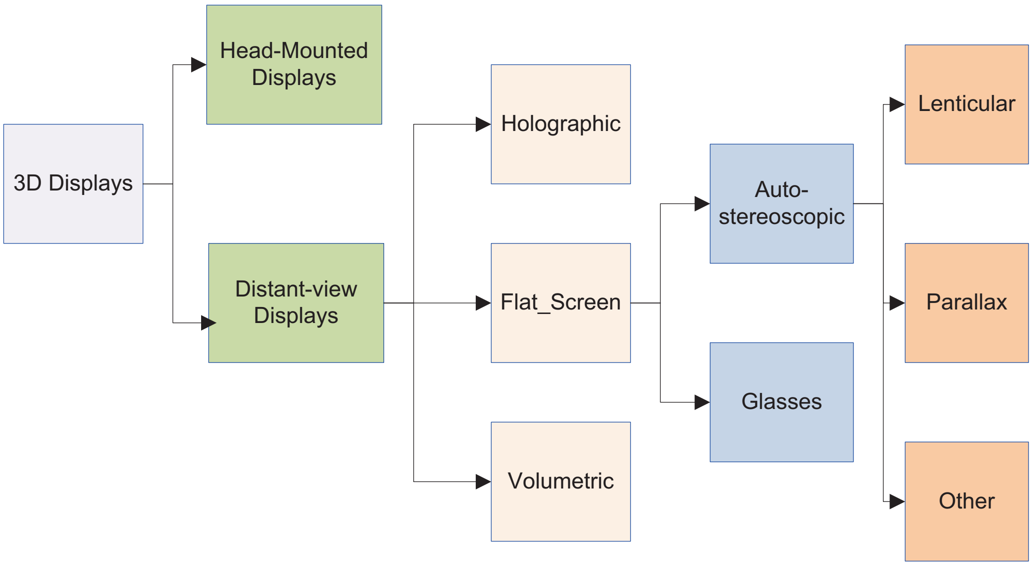



Figure 1.3: 3D display technologies

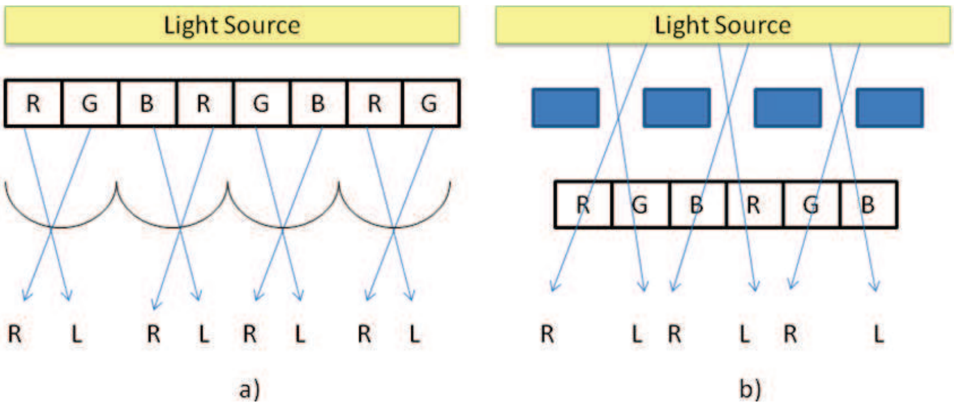



Figure 1.4: Redirecting the light of autostereoscopic display: a) lenticular sheet, b) parallax barrier

## 1.4 Contributions

This thesis proposes a framework for adaptive streaming of MVV using a server-assisted P2P overlay over IP. In the literature, most studies on P2P video neglect issues related to video coding, which has critical impact on the efficiency of the solution. We believe that a successful P2P video streaming system should employ a cross-layer solution, in which the network layer is video coding aware. Especially in 3D video, different components of the MVV format, such as different views and their depth maps, and possibly different video layers depending on their depth and geometry, affect the overall perception of 3D video experience differently. Hence, it is natural to consider unequal encoding and/or adaptation priority for them. Therefore, our adaptive P2P streaming solution carefully considers 3D perception issues specific to MVV, such as the effect of view and depth adaptation and the effect of packet losses in designing the proposed adaptation strategies. Our pull-based P2P MVV streaming solution that responds to fluctuations in the network conditions according to the findings of subjective tests.

The contributions of this thesis can be classified under three categories.

First one is the examining the affect of packet losses over MVV perception:

- We have used Raptor coding as the FEC algorithm due to its high efficiency to protect stereo video streams.
- Determining the distribution of FEC packets among views and observing their performance under varying channel packet loss probability rates.
- Determining whether HVS can compensate the distortion due to the packet losses like it compensate the low quality pair as reported on previous researches.
- Figuring out which concealment method gives the best perceived quality. We provide results in both objective (PSNR) and subjective metrics.

Second one is deciding a quality degradation strategy to dynamically matching content bit rate to channel capacity:

- Exploiting new scalability options of MVV format that are not possible in 2D or 3D stereo video such as changing the number of views.
- By conducting subjective tests for comparing many scaling options for MVV, trying to find best strategy about how to decrease bit rate of the video.

Third one is implementing an adaptive, MVV specific P2P streaming application:

- Our adaptive P2P system responds any fluctuation of network conditions by changing the adaptation strategy defined in the previous contribution.
- Our system employs scalable video coding format to increase scalability which allows high adaptation capabilities since it is possible to discard a portion of the encoded bit stream at the cost of graceful quality degradation.
- A centralized server will assist P2P service start-up and in case of failures, such as ungraceful peer exit.
- New piece scheduling algorithm to prevent requesting uncritical pieces before the critical ones.

## 1.5 Organization

The rest of the thesis is as follows. Next chapter presents determining features of Multi-view video perception that contains both effect of packet losses and view adaptation techniques. Chapter 3 presents application of adaptation strategies for P2P streaming. Other aspects of proposed system such as chunk size will be described in this chapter. Finally, Chapter 4 includes the conclusion and gives future directions.

# Chapter 2

# MULTI-VIEW VIDEO PERCEPTION

## 2.1 Introduction

In this chapter, we will analyze the two main things that affect the perception of 3D video. First one is the impact of packet losses and second one is the impact of scaling methods to adapt the video bitrate to channel bandwidth. In section 2.2, we will address the first factor that is the effect of network packet losses/delay on visual perception of 3D video. For achieving better multi-view video compression, researchers are focusing on three main issues. Firstly, exploiting inter-view dependencies provide significant gain in terms of bitrate [49]. Second one is sending only a portion of color views and rendering intermediate views with the help of low-cost (low bitrate) depth maps [50]. Third one is asymmetric rate allocation by exploiting human visual system which compensates the high frequency components in one view [51][53]. According to previous researches, performing scaling methods (spatial, temporal or SNR) to one view while the other view is hold at a high quality, rate-distortion performance stays at acceptable levels [21]. In coherence with the previous findings, we will address the impact of asymmetric allocation of channel coding over video perception. Moreover, different concealment methods have examined to alleviate packet loss effect by conducting the subjective quality evaluation tests in that section. In section 2.3, we will address the second factor. First a number of scaling methods will be defined for determining the best degrading strategy in compliance with the available bandwidth, and then these methods will be assessed by subjective tests. The following section gives information about the technique employed in the subjective tests.

### 2.1.1 Objective Quality Metrics

Peak Signal to Noise Ratio (PSNR) is the most widely adopted and accepted quantitative metric for assessing the quality of impaired monoscopic video [54]. It is reported that the existing 2D objective quality measures of 3D video sequences fail to capture perceptual errors and causes misleading results [55]. Although some researchers employ PSNR for assessing 3D video, the limitations of PSNR for assessing 2D video have been revealed in [56]. PSNR causes same kind of limitations to the 3D video quality evaluation as well [54]. Moreover, PSNR metric does not reveal the comfort of depth information. A number of approaches for objective 3D video quality assessment are listed in [57]. However, no such objective measure has been widely accepted or adopted in the community. Therefore, subjective tests yield the most accurate results for the evaluation of 3D video.

### 2.1.2 Subjective Evaluation Methodology

The double stimulus continuous quality scale (DSCQS) methodology was used throughout the subjective tests. The DSCQS method is a standardized subjective quality assessment technique [59][60] and is considered to be appropriate for evaluating 3D video quality. Alternately, an original (reference) sequence and a test sequence are shown. The identical sequences are presented a second time in the same order. Subjects are not informed which one is the reference and which one is the test sequence. After the second presentation, observers are asked to evaluate the perceived video quality for each sequence separately. As can be seen in Figure 2.1, a standard continuous scale from 0 to 100 is used for grading. Red box represent the first pair shown and the green box represents the next one. This scale is labeled by the word expressions as 100-80 corresponds to excellent, 80-60 corresponds to good, 60-40 corresponds to fair, 20-40 corresponds to poor and 0-20 corresponds to bad to guide the observer [62].

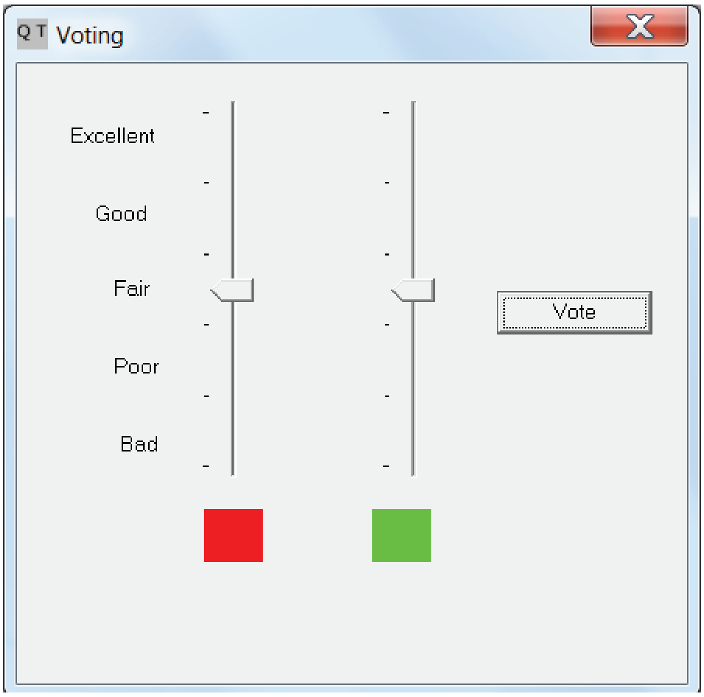



Figure 2.1: Quality Grading Scale

### 2.2 Perception of 3D under Packet Losses

Packet losses are inevitable in IP networks. In UDP a packet can simply be dropped due to best effort nature of the Internet and in TCP, data can still be considered as lost if it is delayed until play-out elapses and it becomes completely useless. So far, researches have mostly focused on perception of 3D video without packet losses, investigating different encoding options such as asymmetric rate allocation. As it is mentioned before, most of the studies reveal that human visual system (HVS) tends to neglect loss of high frequency components in one of the views. However, the packet losses during video transmission are a major QoE issue and should be investigated deeply. The effect of network packet losses/delay on visual perception of 3D video and whether the human visual system can compensate for the artifacts generated by packet losses has not been studied before. In this section, the methods that are aiming to achieve best QoE in case of packet loss scenario by utilizing forward error correction (FEC) and concealment algorithms will be addressed.

As a preliminary test, we have simulated the packet losses on encoded bit streams over channels with different loss rates to observe the impact of packet losses over stereoscopic video. The method used for channel loss simulation and test content preparation phase will be explained later in this section. Two options have been analyzed to see the effect of the losses. In the first case, errors due to packet losses are occurred on similar portions of the both views. In the second case, different portions of the views are affected from the packet losses. In Figure 2.2, both scenarios can be seen. We have investigated if HVS favors one of these cases.

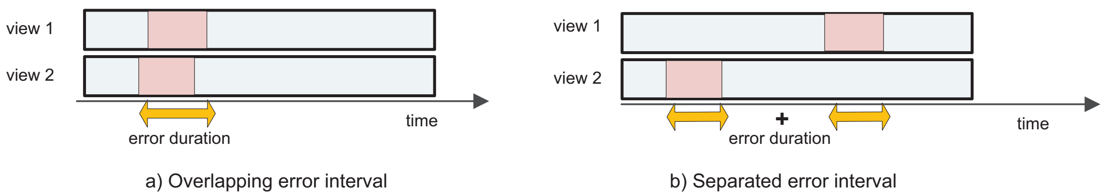



Figure 2.2: Loss intervals in 3D video

The subjective test results reveal that even though the error interval has been extended, people significantly prefer the second option when the different portions of the views are erroneous. In this experiment, it is understood that HVS tends to neglect the errors due to packet losses in one view if the other one is at a good shape.

The next step of this experiment is to figure out how to minimize errors due to packet losses on at least one view by utilizing from forward error correction methods. And this leads to consider the asymmetrical allocation of FEC packets to protect one view better. In section 2.2.1, we will address our fist aim which is to find the optimum FEC allocation among the views to achieve the best perceived quality for a given channel condition by considering the previous test finding.

### 2.2.1 Symmetric vs. Asymmetric Allocation of FEC Packets

As we have discovered from the previous tests, HVS is capable of compensate the artifacts of one view generated by losses, up to a level. Therefore, we focused to get rid of the packet losses for at least one view and tried to reveal how well HVS can compensate artifacts due to less protection of the other view. We compared the symmetrical allocation of FEC packets and applying channel coding asymmetrically to right and left views by expecting to get similar results with the previous researches did it for unequal compression of left and right views [58].

In this study, we have used Raptor coding as the FEC algorithm due to its high efficiency to protect stereo video streams. We have distributed FEC packets symmetrically and asymmetrically among views, and then by conducting visual tests we have decided the best strategy for a better QoE. Now, we will explain our test setup in detail.

***Channel Simulation Tests***

*Generation of Test Streams*

In this study, we have conducted our experiments with stereoscopic 3D multi-media. The reason for that is our aim is to observe the reaction of HVS to differently protected right and left views without affecting artifacts generated by display technology which is a well-known fact that although stereoscopic displays give an acceptable experience, MVV display-dependent artifacts.

MVC aims to offer high compression efficiency by exploiting interview redundancies. MVC is a backward compatible standard, meaning that one of the views must be in compliance with the H.264 syntax and cannot have interview dependency. Consequently, that stream is named as independent view (reference view) and has higher importance for the decoding purposes.

We have encoded the contents presented in Table 2.1 using JMVC (ver. 8.3.1) with quantization parameters (QP) 22, 23 and 24. Slice mode is enabled with mode 2 and slices are restricted to be smaller than 1460 bytes to leave a gap for the UDP and IP headers in a typical LAN with path maximum transmission unit 1500 byte. Encoding results with QP 22 forms the baseline for each sequence. The bitrate gain by using higher QP (23, 24) is used for channel coding (FEC) with three different methods. These methods are based on the idea that independent view may be more important than the other view. In first distribution mode, FEC bits are distributed equally (symmetric), in the second case only independent view is protected, and finally the FEC bit budget is split as 2/3, 1/3 for independent and dependent view respectively.

| **Name** | **Resolution** | **Information** |
|---|---|---|
| Adile | 640x480 | Computer |
| Flower | 704x448 | Moving camera |
| Train | 704x576 | Fixed camera |

Table 2.1: Sequences in simulation

*Simulation of Packet Losses*

We have performed packet loss simulation for %3, %5 and %10 rates. Although, it is possible to perform independent loss chances for each packet, that kind of loss scheme is not realistic for the streaming applications over the Internet because packets are commonly lost in a bursty manner when an intermediate router becomes congested. In order to simulate the packet losses in a realistic way, we have used the trace files provided in [63]. Naturally losing different slices is different cause different artifacts. Therefore, we have 100 runs for each test condition and averaged PSNR results to obtain a reliable value.

***Subjective Tests***

The best error correction result for 10% packet loss scenario is obvious and highlighted in Tables 2.4-2.5. For the remaining two cases (3% and 5% packet loss rate), two subjective tests are performed to evaluate the perception of 3D video for pairs with close PSNR values as highlighted in the same tables. The testing methodology is based on the Double-Stimulus Continuous Quality-Scale (DSCQS) method [60]. We have tested with 7 male and 3 female assessors and 7 of those were experts in 3D video coding.

*Test Setup*

The setup consists of a pair of Sharp MB-70X projectors, two polarized filter glasses that polarize the light emitted in opposite directions, a silver dielectric screen that maintains the polarization after reflection and a PC to drive the projectors. Each projector reflects one of the views on the screen. The assessors wearing polarized filter glasses sit approximately 3.5 meters away from the screen.

*Test Methodology*

The assessors are expected to grade 12 test sequences (four sequences for three different contents) for each test. Each test sequence and its original are displayed twice, in random order so that the assessor cannot know which one is the original. Scoring is on a continuous quality scale in accordance with the DSCQS standard as explained in section 2.1.2. Analysis is based on the difference in rating for each pair rather than the absolute values. Once all participants grade the test sequences, an overall score is calculated for each sequence for analyzing the overall test result. This score is normalized between 0 and 100. The evaluation procedure defines a confidence interval, which is calculated based on the standard deviation and serves as a safety margin for the validity check of the users. A tester is ignored if scored out of safety margins frequently. We note that larger scores indicate poorer perceived video quality.

***Results***

*Encoding and Error Correction Results*

The rate distortion values are presented in Table 2.2. The gains in bitrates are used to add FEC codes in three different approaches. In the first case, FECs are distributed evenly between the dependent and independent stream. Secondly, only the independent stream is protected. And finally FEC packets are distributed at 1/3 (dependent view) and 2/3 (independent view) ratio. Table 2.3 provides the number of FEC packets introduced per 100 video packets.

| | Bitrate (Kbps) | | | PSNR (dB) | | |
|---|---|---|---|---|---|---|
| **QP** | **Adile** | **Flower** | **Train** | **Adile** | **Flower** | **Train** |
| **22** | 1549 | 3699 | 5199 | 41.82 | 38.45 | 38.37 |
| **23** | 1390 | 3195 | 4508 | 41.75 | 37.86 | 37.83 |
| **24** | 1263 | 2801 | 3989 | 40.26 | 37.26 | 37.29 |

Table 2.2: Rate distortion values without packet losses

| | | Symmetric | | Asymmetric 1 | | Asymmetric 2 | |
|---|---|---|---|---|---|---|---|
| | **QP** | **Left** | **Right** | **Left** | **Right** | **Left** | **Right** |
| **Adile** | **23** | 10 | 10 | 14 | 0 | 12 | 4 |
| | **24** | 18 | 18 | 25 | 0 | 22 | 7 |
| **Train** | **23** | 13 | 14 | 25 | 0 | 19 | 7 |
| | **24** | 26 | 26 | 49 | 0 | 37 | 13 |
| **Flower** | **23** | 14 | 14 | 25 | 0 | 19 | 7 |
| | **24** | 28 | 28 | 51 | 0 | 40 | 13 |

Table 2.3: FEC packets amount for all QP and contents

**Error Concealment Using Modified FFMEG**

Since lost data significantly affects the perception, error concealment at the receiver side is an important issue for transmitting MVV over IP. In the situation like the FEC amount is not adequate for the channel loss rate, some of the loss packets remain unrepaired and causes errors. To alleviate this problem and minimize the affect of packet losses over the perception of 3D video, different 3D error concealment algorithms can be adopted in the decoding stage. We have compared two main concealment algorithms; first one is slice repetition and second one is frame repetition. In order to provide a reliable concealment mechanism, we have modified FFMPEG library in compliance with the MVC syntax and implemented slice level error concealment algorithm based on slice repetition. When a slice of a frame is lost the corresponding region's residue is replaced from another frame that is closest in picture order count metric. This approach performs well for slices that has limited or no motion, but introduces distortions when there is a significant motion within the lost slice's region.

To overcome the problem of high motion part of the video, frame repetition seems a trivial solution for not to deteriorate a part of the frame by repeating only the lost slice. We have applied the both concealment methods to the encoded and FEC added video sequences above. According to the conducted subjective test results, people have preferred slice based error concealment method even in sequences with high motion (such sequences generate more disturbing mismatches in slice based error concealment methods compared to frame repetition). Therefore, we have used slice based error concealment method in this test.

*Channel Simulation Results*

First, the FEC encoded video packets are simulated as to be sent over a channel with 3%, 5% and 10% data loss rate. All received video packets and the ones that are recovered by FEC decoder are forwarded to MVC decoder. The PSNR results of the obtained test sequences are represented in Tables 2.4-2.5. The results show that for the network simulation with 10% loss rate, FEC was unable to compensate the lost packets for all but the symmetric scheme at 24 QP. For the two remaining simulations, the asymmetric 1 scheme performed significantly worse than other options, and thus eliminated. We have conducted further subjective tests for 3% and 5% channel loss rate simulations to evaluate the performance of the remaining highlighted options.

| | Adile | | | | | | |
|---|---|---|---|---|---|---|---|
| | | Symmetric | | Asymmetric 1 | | Asymmetric 2 | |
| | QP | Left | Right | Left | Right | Left | Right |
| 3% Loss | 22 | 31.1 | 30.6 | 31.1 | 31.0 | 30.9 | 30.9 |
| | 23 | 39.9 | 40.5 | 40.7 | 31.2 | 40.6 | 33.3 |
| | 24 | 40.0 | 40.0 | 40.0 | 31.5 | 40.0 | 37.5 |
| 5% Loss | 22 | 28.9 | 28.9 | 29.6 | 29.1 | 29.1 | 29.0 |
| | 23 | 36.7 | 36.7 | 39.1 | 29.5 | 37.6 | 31.1 |
| | 24 | 39.0 | 40.0 | 40.0 | 29.4 | 40.0 | 33.8 |
| 10% Loss | 22 | 26.7 | 26.6 | 26.7 | 26.7 | 26.7 | 26.6 |
| | 23 | 27.9 | 27.5 | 29.6 | 27.0 | 28.8 | 26.9 |
| | 24 | 34.3 | 34.5 | 37.2 | 27.1 | 36.8 | 27.5 |

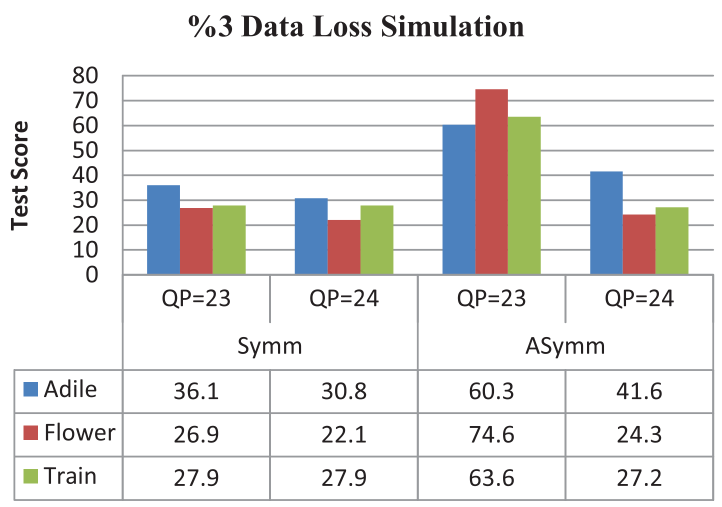
Table 2.4: PSNR values of Adile content after FECs

| | Train | | | | | | |
|---|---|---|---|---|---|---|---|
| | | Symmetric | | Asymmetric 1 | | Asymmetric 2 | |
| | QP | Left | Right | Left | Right | Left | Right |
| 3% Loss | 22 | 29.6 | 28.0 | 29.6 | 28.1 | 29.7 | 28.1 |
| | 23 | 37.2 | 37.4 | 37.3 | 29.8 | 37.3 | 35.4 |
| | 24 | 36.9 | 37.0 | 36.9 | 29.5 | 36.9 | 36.9 |
| 5% Loss | 22 | 26.9 | 25.5 | 26.8 | 25.5 | 27.0 | 25.5 |
| | 23 | 34.9 | 34.2 | 37.3 | 26.6 | 36.7 | 30.9 |
| | 24 | 36.9 | 37.0 | 36.9 | 26.8 | 36.9 | 34.5 |
| 10% Loss | 22 | 23.2 | 22.0 | 23.2 | 21.9 | 23.3 | 21.8 |
| | 23 | 26.4 | 26.0 | 32.7 | 22.8 | 31.7 | 23.1 |
| | 24 | 33.1 | 33.6 | 36.8 | 22.6 | 34.6 | 26.8 |

Table 2.5: PSNR values of Train content after FECs

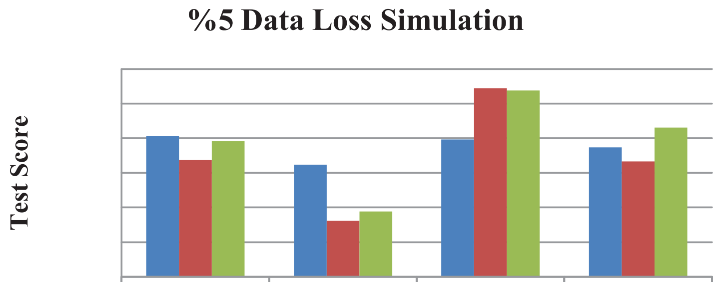
*Subjective Test Results*

The experimental results are summarized for the highlighted sequences in Tables 2.4-2.5 is shown in Figure 2.3-2.4. In these figures the users score for each QP value and channel coding scheme. The subjective test scores clearly indicate that symmetric using higher channel coding rates outperforms the quality gain from source coding, indicating that the artifacts from packet losses are much more noticeable.




| | Symm QP=23 | Symm QP=24 | ASymm QP=23 | ASymm QP=24 |
|---|---|---|---|---|
| Adile | 36.1 | 30.8 | 60.3 | 41.6 |
| Flower | 26.9 | 22.1 | 74.6 | 24.3 |
| Train | 27.9 | 27.9 | 63.6 | 27.2 |

Figure 2.3: Subjective test results




| | Symm QP=23 | Symm QP=24 | ASymm QP=23 | ASymm QP=24 |
|---|---|---|---|---|
| Adile | 40.7 | 32.4 | 39.6 | 37.4 |
| Flower | 33.7 | 16.2 | 54.4 | 33.3 |
| Train | 39.1 | 18.9 | 53.8 | 43.1 |

Figure 2.4: Subjective test results 2

The objective and subjective evaluation results are in coherence and based on their results we can make the following conclusions:

- It is clear that the artifacts due to packet losses/delays cannot be compensated by HVS and so should be avoided at all cost. Therefore, during the transmission of the MVV over the Internet, it is best to decrease the quality of the video to provide gap for error correction algorithms. For the transmission over TCP, this gap can be used for delays in packet retransmission.
- Even though MVC generates streams at unequal importance, in the face of network errors both stream generates noticeable artifacts.

## 2.3 SCALING METHODS

Second important factor affecting perceived quality of MVV is the effect of scaling methods. Our goal is to identify the features of human visual system and understand the best method of graceful quality degradation for adaptive video streaming purposes. Since it is difficult to quantify the QoE (especially for 3D content) by objective metrics, we have conducted subjective visual tests to quantify the effect of different bit-rate degradation strategies on QoE. In each test, MVV sequences are scaled to match a certain bitrate using various methods and assessed according to perceived QoE. The subjective tests results form the basis of adaptation strategy in the proposed MVV streaming solution.

### 2.3.1 Scaling Methods for Multi-view Video

The goal of each scaling method is to decrease overall bitrate of transmitted MVV to match the available network rate. This can be done either by decreasing quality (quality scaling) in one or more scalability dimensions (SNR, spatial or temporal) or by discarding view(s) at the cost of transmitting depth-maps (view scaling).

**Quality scaling** can be applied equally on each view (*symmetric scaling*) or unequally among different views (*asymmetric scaling*). In *symmetric scaling*, all views are affected thus the decrease in perceived video quality is inevitable unless the video is coded at high quality unnecessarily. *Asymmetric scaling* is a more advanced approach. The studies on stereoscopic 3D video have revealed that unequal scaling may yield higher perceived quality compared to symmetric coding at the same bitrate [21]. In the case of multi-view video, asymmetric scaling corresponds to scaling adjacent views at different quality levels. (e.g., High-low-high…)

**View scaling** is an alternative method in which a subgroup of views is transmitted with associated depth maps. Missing views are interpolated at the client side using depth-image-based-rendering (DIBR) [61]. Depth maps are single channel images with less high frequency components and can be coded with higher efficiency compared to color images. However, the quality of the associated colored image and the depth map must be high; otherwise artifacts may occur in the interpolated views.

***Subjective Tests***

Subjective tests are to evaluate the performance of scaling methods in terms of delivered QoE under different network conditions. In each test, a different target bitrate is set to simulate a certain network condition and test video sequences are encoded to match the target bitrate using one of the scaling methods listed in Table 2.6. Then assessors have evaluated each scaling method by comparing the scaled sequences against original sequence based on the Double-Stimulus Continuous Quality-Scale (DSCQS) standard [60]. 12 male and 4 female assessors have attended the tests and 7 of them were experts in 3D video coding area.

The tests are performed using 5-view 3D display at 1920x1200 screen resolution that is equipped with lenticular sheet technology.

| No | Description | Detail |
|---|---|---|
| 1 | Symmetric | SNR |
| 2 | Symmetric | Spatial |
| 3 | Asymmetric | SNR |
| 4 | Asymmetric | Spatial |
| 5 | DIBR | 3 ref. view |
| 6 | DIBR | 2 ref. view |

Table 2.6: Scaling Methods

| Provider | Sequence | Image Property | # camera |
|---|---|---|---|
| Nagoya University Tanimoto Lab | Dog | 1280x960 | 80 cameras |
| | Pantomime | 1280x960 | 80 cameras |
| ETRI and MPEG-Korea | Lovebird1 | 1024x768 | 12 cameras |

Table 2.7: Recorded test sequences

*Test Content Preparation*

In each subjective test, test sequences in Table 2.7 are encoded using scaling methods explained below to match the target bitrate. In some tests, some methods cannot be applied if they fail to match target bitrate at an acceptable quality level or fail the asymmetric coding criterion. (In asymmetric coding approach, it is stated that the low quality pair should be above ~32dB [21].) Table 2.8 presents the bit-rate limitations and Figure 2.5 depicts their RD performances. The scaling methods used in the subjective tests are as follows:

**Method 1:** In symmetric SNR scaling, all views are encoded at a QP value higher than that of the reference view.

**Method 2:** Symmetric spatial scaling is performed by decimating each view to one fourth of its original resolution and then coding at target bitrate. We note that content at high definition resolution are perceivably less affected by decimation followed by interpolation compared to content at standard definition resolution.

**Method 3:** In asymmetric SNR scaling, the intermediate views (view 2 and 4) are encoded at a low quality threshold whereas remaining are first coded at high quality such that the overall bitrate is matched. The drawback of this method is that, it yields marginal bitrate reduction thus it could be applied only in a Test 1.

**Method 4:** In asymmetric spatial scaling, the high quality views (view 1, 3 and 5) are first downscaled and then coded.

**Method 5 and 6:** In these methods, some of the views are not transmitted at all and estimated by using (DIBR) techniques [61]. In method 5, only two views are discarded whereas in method 6 only two views are kept. Therefore, the former one has three reference views at lower quality, whereas method 6 has only two reference views but at a higher quality. As a sample we provide Table 2.9 that summarizes the conditions of subjective tests for Pantomime sequence. (Adopted scaling methods, achieved PSNR values for each view, bit rate and the PSNR value of the reference views and bit budget for each test.)

| Test No | Pantomime | Dog | Lovebird |
|---|---|---|---|
| 1 | ~ 4600 | ~ 3800 | ~ 2500 |
| 2 | ~ 3500 | ~ 2500 | ~ 1900 |
| 3 | ~ 2900 | ~ 2100 | ~ 1500 |
| 4 | ~ 2300 | ~ 1700 | ~ 1300 |

Table2.8: Bit Budget of Tests for Each Sequence (kbps)

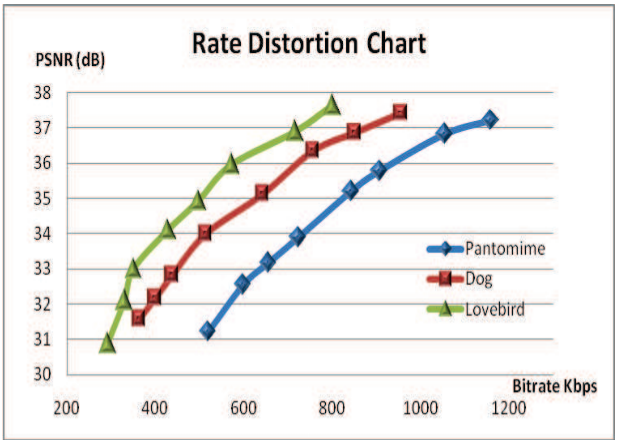



Figure 2.5: RD Performance of Test Sequences

| | View Number 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| **Reference:** ~5700kbps | | | | | |
| | 37.6 | 37.6 | 37.6 | 37.6 | 37.6 |
| Method | **Test 1:** ~4600kbps | | | | |
| 1 | 35.7 | 35.7 | 35.7 | 35.7 | 35.7 |
| 3 | 37.6 | 32.6 | 37.6 | 32.6 | 37.6 |
| 4 | 37.6 | 35.6 | 37.6 | 35.6 | 37.6 |
| 5 | 37.6 | | 37.6 | | 37.6 |
| Method | **Test 2:** ~3500kbps | | | | |
| 1 | 33.2 | 33.2 | 33.2 | 33.2 | 33.2 |
| 2 | 35.6 | 35.6 | 35.6 | 35.6 | 35.6 |
| 4 | 35.6 | 34.9 | 35.6 | 34.9 | 35.6 |
| 5 | 35.7 | | 35.7 | | 35.7 |
| Method | **Test 3:** ~2900kbps | | | | |
| 1 | 32.6 | 32.6 | 32.6 | 32.6 | 32.6 |
| 2 | 34.9 | 34.9 | 34.9 | 34.9 | 34.9 |
| 4 | 35.6 | 32.6 | 35.6 | 32.6 | 35.6 |
| 5 | 33.9 | | 33.9 | | 33.9 |
| 6 | 37.6 | | | | 37.6 |
| Method | **Test 4:** ~2300kbps | | | | |
| 2 | 34.2 | 34.2 | 34.2 | 34.2 | 34.2 |
| 5 | 31.2 | | 31.2 | | 31.2 |
| 6 | 34.8 | | | | 34.8 |

SNR scaled view

Spatial scaled view

view+depth

discarded view

Table 2.9: PSNR (dB) values of test sequences for scaling methods (Pantomime Sequence)

*Subjective Test Results*

Figure 2.6 presents the subjective test results for all test sequences with corresponding adaptation method number. Our observations for each test are as follows:

In the first test, the bit-rate budget is high enough to code at high quality by all scaling methods. Therefore almost all methods scores close to the reference view. Only symmetric SNR scaling introduces barely noticeable artifacts (observed by trained assessors). Method 3 has lower PSNR average than Method 1 but HVS can compensate the difference in quality in stereoscopic 3D [21] [64]. So we have experience a similar result for the MVV case.

In Test 2, Method 1 has a lower score; indicating that the artifacts become more visible (See Figure 2.6). The reason clearly can be seen in for Pantomime sequence, all views are coded at around 33 dB and this amount is at a visually perceivable low quality.

For the remaining last two tests, transmitting color images (whether they are coded symmetric or asymmetric) for each views yield perceivably lower quality. Again, this finding is coherent with the previous studies in which authors state that beyond a certain PSNR threshold value, asymmetry performs poorly [21]. When we compare frame based methods (asymmetric or symmetric) against DIBR based methods, we observe that DIBR methods are more favorable at low bit rates if implemented correctly. The results indicate that using less number of reference frames at higher quality is better that using more reference frames at lower quality for DIBR techniques. Then, DIBR based methods either provide the best visual quality or at least they match the score of frame based methods. This observation is evident from the test scores of Method 5 and 6.

Yet another observation is that, in spite of the high PSNR value, blurriness has been observed in interpolated low-resolution videos in the subjective tests. So it is preferable to pick view plus depth solution if the budget is very limited.

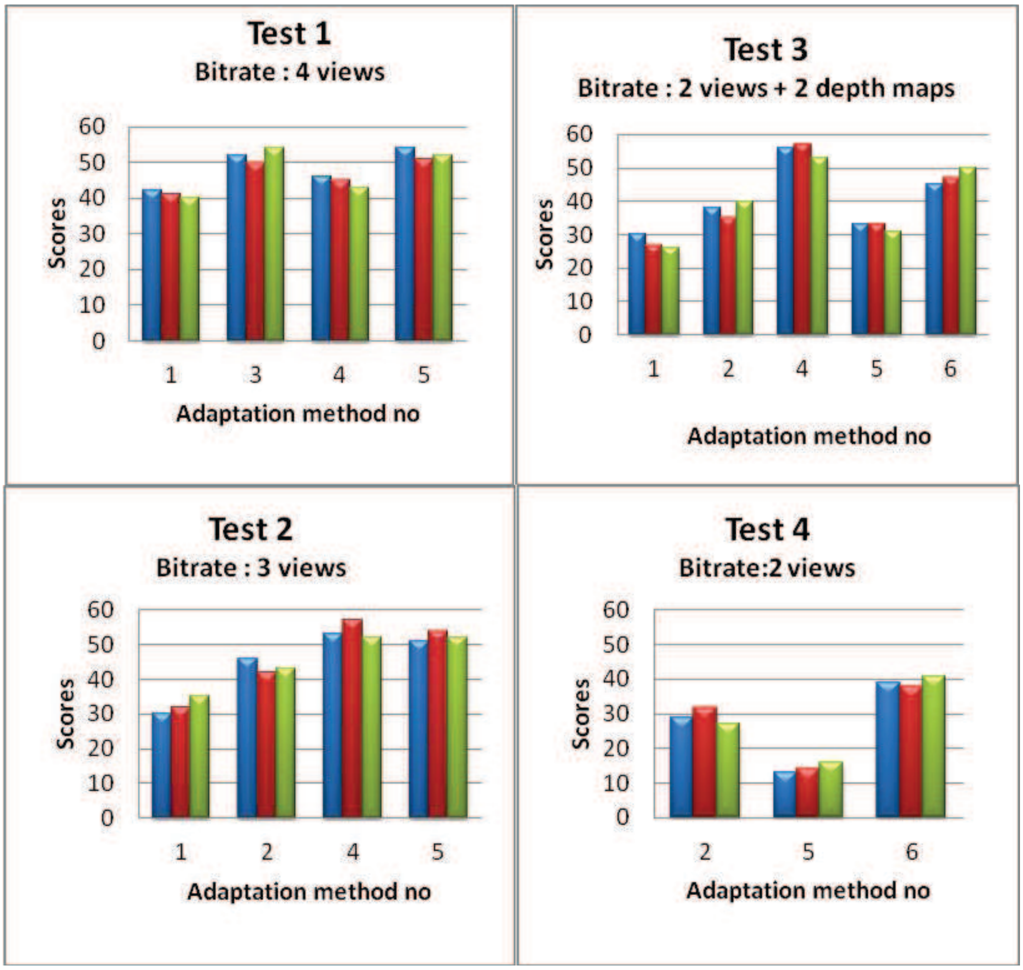



Figure 2.6: Subjective test results

*Adaptation Decision Chart*

Based on the results of the subjective tests we have the following conclusions. In order to obtain the best perceived video quality, intermediate views should be scaled asymmetrically as much as possible. Once asymmetric coding threshold (~32dB) has been reached then it is possible to drop all the views between the first and the last one and interpolate them using DIBR technique. Table 2.10 summarizes the adaptation decisions.

| **Step No** | **Method** |
|---|---|
| 1 | All views at highest quality. |
| 2 | Asymmetric scaling for intermediate views |
| 3 | Keep only the first and last view and use DIBR |

Table 2.10: Adaptation Decision Chart

# Chapter 3

# A FRAMEWORK FOR QoE AWARE MVV STREAMING OVER P2P NETWORKS

## 3.1 Introduction

In this study, we aim to determine the necessary modifications to the currently available P2P technologies to establish a successful 3DTV service that have two key features: i) to be able to operate in wide range of network conditions and perform rate adaptation based on the QoE tests results ii) to require minimum network resources at the content originator. To this end, we have designed a hybrid content distribution system in which data originates from content server(s) and disseminates to swarm using mesh based P2P network. The rest of this section describes the novel approaches in order to achieve the above mentioned requirements.

## 3.2 Chunk Generation

*Video Coding Scheme*

For video streaming applications over IP, rule of thumb is to keep the overall bitrate of the content as low as possible because each additional data increases the possibility of experiencing IP failures. Since the adaptation decision chart (See Table 2.10) indicates that only the intermediate views should be scalable, only those views are encoded using SVC extension of H.264/AVC which causes 10% extra overhead. Remaining views and the depth-map-video of the views at the edges are encoded using standard H.264/AVC which is currently the best codec in terms of compression efficiency. Once the video coding is over every stream is mapped into P2P chunks independently.

*Chunk Mapping*

Packetization of multimedia content considering the underlying network infrastructure has a critical impact over the performance of any streaming solution. For a Torrent based P2P system, packetization corresponds to formation of video chunks. In Bittorent protocol, all the chunks have a fixed size which is based on the total size of the shared content. This approach is not very suitable for video coding applications because the rate may fluctuate over time. Moreover, chopping the bit stream at fixed locations generates chunks that are not independently usable since the actual frame data and the required header information may become separated. Therefore, we adopt variable sized chunks using group of picture (GOP) boundary as possible separation points. It is possible to use multiple GOPs to create a single chunk, if the payload size is too small.
An encoded video starts with non-video coding layer (non-VCL) NAL units, such as sequence parameter set (SPS). They provide vital information to initiate video decoding process. (e.g., picture resolution in macroblocks (MB) and decoded picture buffer size). In order to ensure video decoding, Non-VCL NAL units are provided inside the metadata file to protect against packet loss.

*Chunk Generation using SVC bit stream*

The SVC stream starts with non-VLC NAL units similar to H.264/AVC. When slice mode is disabled, there is one base layer NAL unit and one enhancement layer NAL unit for each frame. The base layer bit stream can generate a video frame (at a lower quality) even if the enhancement layer bit stream is not received; indicating that enhancement layer NAL units are discardable. Therefore, when generating chunks using SVC encoded video, we propose to split the base and enhancement NAL units into separate chunks (See Figure 3.1) Please note that, due to hierarchical coding in SVC, the first GOP has two I-frames (represented with *) creating a relatively larger chunk.

### 3.3 Peer Selection

The tracker is responsible for tracking peer-list and content servers and informs peers about the current state of the swarm. In BitTorrent, tracker server randomly forwards a subgroup of peers. Definitely it is possible to follow an intelligent approach for peer selection. Instead of selecting partners randomly, Hei et al. claim that buffer maps can be utilized to monitor network behavior of a peer and suggested matching peers based on the state of their buffer maps (a data structure that indicates the chunks that are available in a peer)[65]. The authors state that it is important to choose a peer with similar network resources.

In our solution, tracker server clusters peers according to the requested views. Tracker always forwards a random subgroup of peers among the ones with same or close number of view requirement. When a peer connects for the first time, peer selects a subgroup of these forwarded peers by tracker. In the following connections, peers with similar buffer map pattern are preferred among the received peers.

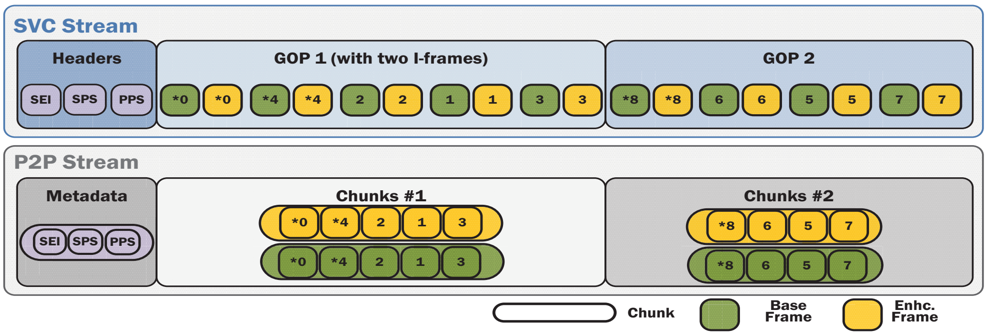



Figure 3.1: Chunk generation using one GOP per chunk

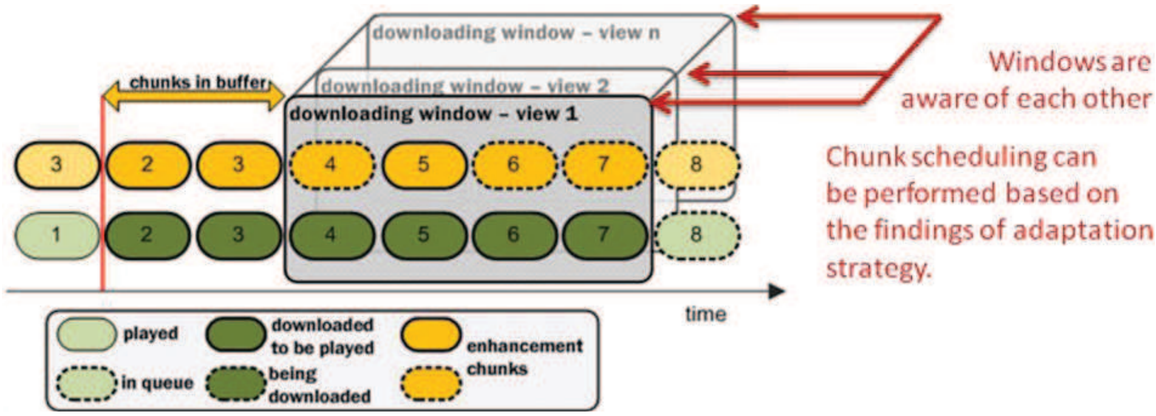



Figure 3.2: Downloading window with two layers and n views

(The red line indicates current location of the player.)

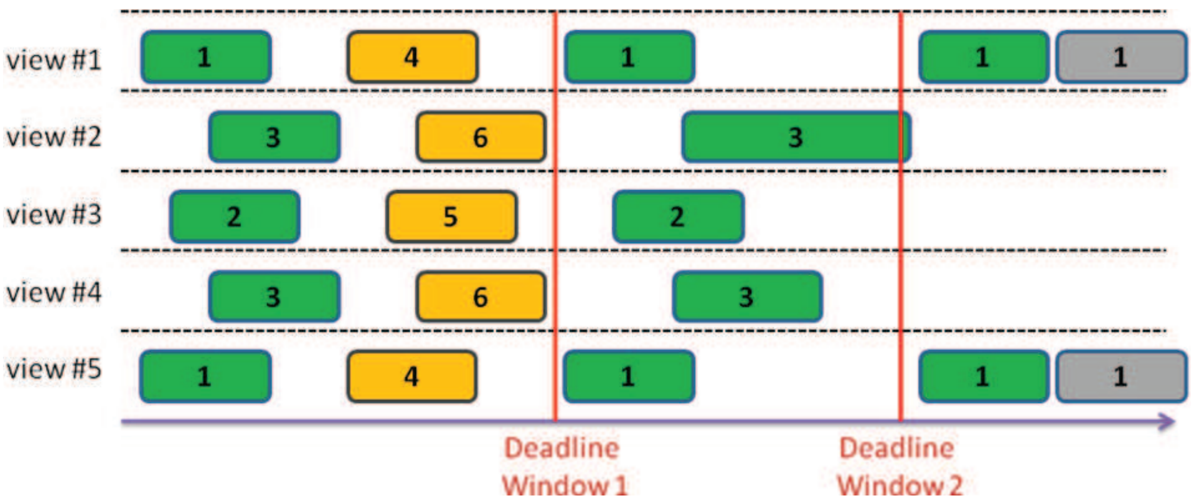



Figure 3.3: QoE Aware Scheduling

## 3.4 Piece Selection

Piece selection algorithm is responsible for defining the rules of data exchange among peers. Commonly, peers exchange buffer map to indicate the availability of chunks to the neighboring peers. We adopt a similar architecture in which peers notify the available chunk, along with its view and layer identifier. A similar strategy is adopted for selecting the peers to connect. When selecting peer to make chunk request, a recipient peer compares its buffer maps with the candidate peers and choose the one that have downloaded similar number of enhancement chunks.

*QoE Chunk Scheduling for Request*

The windowing mechanism in our framework is an extended version of the currently available in two additional dimensions as depicted in Figure 3.2. The first dimension is the number of views meaning that there are separate windows for each view and depth-map. The second dimension is the quality layer which represents the *discardibility* of chunks. (Available only if scalable version of a view is to be requested.) For scalable views, all base layer chunks must be acquired prior to requesting enhancement layer chunks. By this way, the scheduling mechanism tries to ensure smooth, uninterrupted video playback as much as possible and provides high quality in surplus of network resources.

Once a streaming session is initiated, peers start to schedule chunks within the first window randomly. Once all the chunks in window is downloaded (at maximum quality), the window slides and schedules following chunks. At this moment, player may start to consume downloaded multimedia. The time required to download the first window is considered as buffering duration.

Once buffering is over, the state of the buffer determines the adaptation decisions. If the duration of buffer is low, the first action is to discard the enhancement chunks. If problem persists and the peer fails to receive base layer chunks, then views are discarded based on their priority, which is defined by result of subjective tests. In the other case, if the buffer size gets large, then higher quality is layer chunks are requested.

The number of concurrent chunk requests (downloads) is a system parameter. In recommendations, it has been limited to 5; stating that the download capacity of a peer is saturated by finding good neighbors, not by increasing the number of connections [66]. In our case, when a chunk is scheduled, it is requested from server only if it is not available in neighboring peers. However, we limit the chunk requests from the server to one at a time to decrease the workload on the server side.

Yet another important issue is the prioritization difference among windows. We provide a sample scheduling time-chart in Figure 3.3 that considers QoE results found in Chapter 2 (Table 2.10) for 5 view display system. This figure is divided by the window download deadline boundaries. The rounded rectangles correspond to downloading window of each view. The numbers indicate the order of request for each window.

According to the results in Table 2.10, a peer should first try to download the views at the edges. Therefore, in Figure 3.3, the peer first schedules the windows of view 1 and 5. Next, the intermediate chunks are requested. Once the base layer has been downloaded, the peer tries to download the enhancement layer in the same order. According to Figure 3.3, the first window could have been downloaded before its deadline. A similar downloading strategy has been adopted for the next turn. However this time, the window of view 2 could not have been finished on time. Since even the base layer chunks could not be received, in the next iteration, the peers go to the last option in the adaptation decision, which is to download only the edge maps along with the corresponding depth maps.

*Scheduling Responses to Chunk Requests*

Peers occasionally receive chunk requests from other peers. Each request is first placed in a queue, similar to the approach in Layered P2P [67]. However, we allow a number of chunk requests to be processed at a time.

When the system can handle a new request, requests are sorted according to two criteria. First, the requested view id is considered. If it is a critical stream, such as the depth-map-video then it has been served first because the recipient may be in a critical situation. Next, the layer has been considered. The base layer chunks are prioritized and the upload scheduling process.

## 3.5 Networking Tests

### 3.5.1 Test Environment

We have implemented a P2P video streaming architecture with the above mentioned features. In order to test the software we have used the PlanetLab environment in which multiple nodes all over the globe form a network with no firewall or NAT complexities. In our test setup one of the nodes has been assigned as content server. That particular server also provides the tracker service which enables peers to find each other. When a peer joins the network, it first connects the tracker server and receives a sub-list of peers that are recently active. The tracker server also receives periodic feedback from peers about their state which includes the buffer map and number of connected peers. Throughout the tests, we have monitored the traffic over the content server and the state of all peers.

When the P2P test starts, randomly selected peers are signaled via remote procedure call (RPC) which makes each peer wait a random amount of time before connecting to the tracker server (between 0 to 3 seconds). Once this period is over, the peer connects to the tracker server and receives the sub-list of peers. Then, it sends a handshake message to each peer in the list (it may get connected to maximum 50 peers). In this step, the peers exchange their current buffer map and learn about the chunks that are available currently.

When a peer schedules a chunk to be downloaded from its download window (DW), first the peers are checked to see if the chunk is available or not. If a peer's buffer map indicates that the chunk is available then the chunk is requested from that peer. Otherwise, the chunk has been requested from the content server.

Throughout the test, when a peer successfully downloads a chunk, it sends update messages to indicate the availability of the chunks. Note that, if the chunks are too small, then there are too many message exchanges which increases the overhead. On the other hand, when there are many distinct small but more chunks (considering that the total window duration is equal) are present; peers are more likely to find to exchange chunks among each other. Therefore, the traffic over the service may change according to the scheduling of chunks by the peers. If the window size is set larger without decreasing the size of each chunk, then the average pre-buffering time of the peers increases.

This study aims the deployment of a MVV service; however the current infrastructure is not adequate to stream 5 views even in PlanetLab due to the high bitrate requirement of content server at the beginning of the session. Therefore, we have tested to stream single view to be able to observe server burden. This stream has been coded using SVC at 1.2 Mbps (620 Kbps base layer with PSNR value 33.5, 580 Kbps enhancement layer with PSNR value 36.5). The window size is 10 chunks (~5 seconds).

### 3.5.2 Test Results

We have the following comments about the results presented in Table 3.1 and Figure 3.4:

- The increase in number of peers has diminishing effect over the bandwidth requirement of the server; indicating that the proposed architecture is scalable.
- The peak load at the server side occurs at the initial stage, in which peers unable to exchange data among each other. This is probably due to the fact that, at the initialization stage it is more difficult to find chunks in the P2P network.
- Nevertheless, the server load is not linear with the number of peers. Although the content has been at 1.2 Mbps, the maximum server-load is around 10Mbps at the initial stage. This has been achieved by the different startup delay (random of 3 seconds) by the clients which makes finding the initial requested chunks among peers.
- Pre-buffering delay increases with increasing number of clients. However, it is still shorter than the window duration which indicates that the available server bitrate is higher than the content bitrate. Only in the test with 15 peers, the buffering duration is higher than the window duration. This suggests that the server may be unable to serve all peers at the initial stage.

We run each test a large number of times until the results saturate. On the average, only a few peers have received the content at the lowest quality. Almost all of the peers could acquire all base layer chunks. The average ratio of receiving enhancement layer chunks is about 50% percent.

| # Peers | Buffering Duration (sec.) | Server Load (Mbps) |
|---|---|---|
| 5 | 2.53 | 3.71 |
| 10 | 3.97 | 5.36 |
| 15 | 5.26 | 6.80 |

Table 3.1: Average Buffering Duration (peers) and Server Load

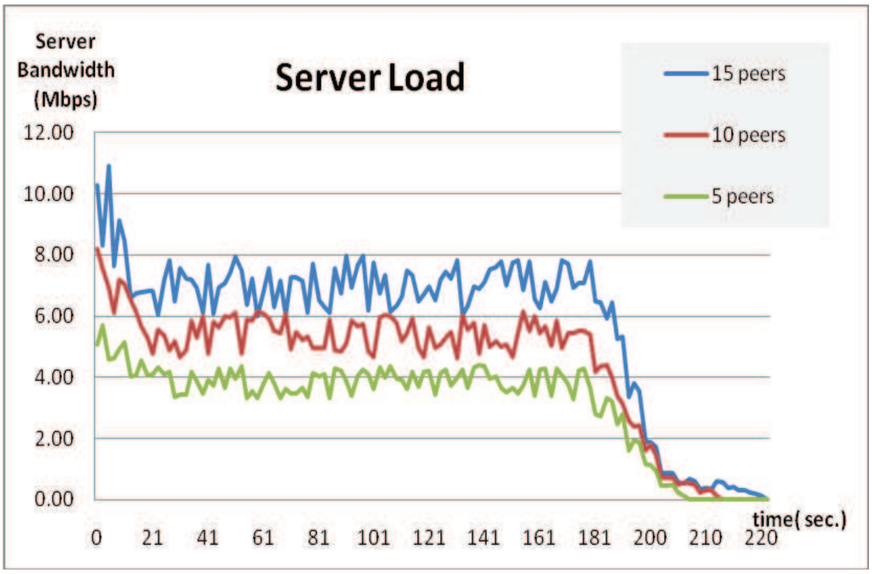



Figure 3.4: Server Load over Time

# Chapter 4

# Conclusions

## 4.1 Summary

In this study, we propose a framework for multi-view video streaming considering perceived quality of experience. We have performed series of subjective tests to understand the characteristics of the important factors that affect the perception of MVV and focused on finding ways to improve perceived video quality. The results from these tests allow us to develop a MVV streaming system with a better QoE. In the rest of this section, we will present the findings from the tests and mention our P2P system for MVV video delivery as the summary of the work done.

In order to establish a successful 3DTV (free-TV) application, one should consider the challenges of using IP as the network protocol. In this thesis, we propose an adaptive MVV streaming mechanism over Peer-to-peer (P2P) networks for two of the major challenges in IP streaming which are scalability problem and varying link capacity between nodes.

We have utilized adaptive streaming for the varying available bit-rate problem by dynamically matching content bit rate to channel capacity. The adaptation methods described in this thesis provides means to adapt diverse network conditions. Moreover, these methods regard both the video coding efficiency and perceived video quality to maximize the visual experience at a given channel condition and also minimize the risk of IP failures. These adaptation methods require quality degradation with different amounts to match the available bit-rate. To investigate them in terms of QoE and identify the best rate adaptation strategy, we have performed subjective quality tests. According to findings of these subjective tests, in order to obtain the best perceived video quality, intermediate views should be scaled asymmetrically as much as possible. Once asymmetric coding threshold (~32dB) has been reached then it is possible to drop all the views between the first and the last one and interpolate them using DIBR technique.

Next, we have addressed the scalability of media delivery solution in the case of increasing number of clients. Since MVV has large bandwidth problem, experiencing bottleneck is almost inevitable in server client based architectures. As a remedy, we have introduced a novel adaptive P2P video streaming approach that is build on top of legacy Bit-torrent. Using our solution, it is possible to alleviate centralized high bandwidth requirement at the server side and distribute the burden over all peers. However, a practical adaptive P2P MVV service cannot be achieved without properly addressing the perceived quality of experience (QoE) of MVV. Therefore, we have performed rate adaptation based on the QoE tests results to operate in wide range of network conditions. Due to the Internet inadequacy, the implementation of the proposed protocols has been tested for one view on PlanetLab testbed. Additionally, the effect of network packet losses during transmission over perception of MVV is an important factor of a successful QoE. We aimed to reduce the effect of the errors generated due to the packet losses as much as possible by utilizing forward error correction (FEC) algorithms and concealment methods. The optimal allocation of FEC packets among views was investigated in order to achieve a better MVV perception. The subjective test results indicate that asymmetric allocation of FEC packets results in better perceived quality.

## 4.2 Future Work

The actual P2P streaming tests of this study have been performed in PlanetLab with only one view. This was due to the limited bandwidth capacity of the serving node, which is responsible for forwarding a chunk if it is not present in the P2P swarm. We believe that a streaming test with less than 15 peers cannot be considered as a realistic P2P environment. Unfortunately, we could reach this number with only one view in our test environment. We conclude that the current infrastructure of the IP networks is not enough to transfer multi-view video with an acceptable quality and so we are not able to evaluate the QoE over the streamed MVV.

As future work, we are planning to lower the bitrates of all views to match the available capacity of current Internet infrastructure and then evaluate the success of the proposed adaptation framework. And also we will continue our QoE evaluation and observe the effect of developments in the infrastructure to send MVV.

**BIBLIOGRAPHY**

[1] A. Vetro, P. Pandit, H. Kimata, A. Smolic and Y-K. Wang "Joint draft 8.0 on multiview video coding," Joint Video Team, Doc. JVT-AB204, July 2008.

[2] Y. Chen, M. Hannuksela, L. Zhu, A. Hallapuro, M. Gabbouj, and H. Li, "Coding techniques in multiview video coding and joint multiview video model," in Proc. of the Picture Coding Symp. (PCS), May 6-8, 2009, Chicago, Illinois, USA.

[3] C. Gurler, B. Gorkemli, G. Saygili, and A. Tekalp, “Flexible transport of 3-D video over networks,” Proceedings of the IEEE,vol. 99, no. 4, pp. 626–642, Apr. 201.

[4] A. Vetro, S. Yea, A. Smolic, Towards a 3D video format for autostereoscopic displays, In: Proceedings of the SPIE: Applications of Digital Image Processing XXXI, San Diego, CA, USA, 2008.

[5] ATTEST [Online]. Available: http://www.hitech-projects.com/euprojects/attest/

[6] Text of ISO/IEC FDIS 23002-3 Representation of Auxiliary Video and Supplemental Information, ISO/IEC JTC1/SC29/WG11, Doc. N8768, Marrakesh, Morocco, Jan. 2007.

[7] Text of ISO/IEC 13818-1:2003/FDAM2 Carriage of Auxiliary Data, ISO/IEC JTC1/SC29/WG11, Doc. N8799, Marrakech, Morocco, Jan. 2007.

[8] P. Merkle, A. Smolic, K. Müller, and T. Wiegand, ―Multi-view video plus depth representation and coding,‖ in IEEE Int. Conf. on Image Process. (ICIP), San Antonio, TX, Sep. 2007.

[9] K. Müller, A. Smolic, K. Dix, P. Merkle, P. Kauff, and T.Wiegand, “Reliability-based Generation and View Synthesis in Layered Depth Video“, Proc. MMSP 2008, IEEE International Workshop on Multimedia Signal Processing,Cairns, Australia, October 2008.

[10] A. Smolic, H. Kimata, and A.Vetro, “Development of MPEG standardsfor 3D and free viewpoint video,” presented at the SPIE Optics East, Three-DimensionalTV,Video, and Display IV, Boston, MA, Oct. 2005.

[11] A. Smolic, K. Mueller, P. Merkle, P. Kauff, T. Wiegand, “An overview of available and emerging 3D video formats and depth enhanced stereo as efficient generic solution”, Picture Coding Symposium (PCS’09), Chicago, USA, May 2009.

[12] M. Martini, C. Hewage, “Objective quality assessment for 3-D video delivery”, IEEE COMSOC MMTC E-Letter, vol.6, No.8, August 2011.

[13] M. Lambooij, W. IJsselsteijn, M. Fortuin, I. Heynderickx,” Visual discomfort and visual fatigue of stereoscopic displays: a review”, Journal of Imaging Science and Technology 2009; 53(3):030201-(14).

[14] Boev A., Gotchev A., and Egiazarian K., Stereoscopic artifacts on portable auto-stereoscopic displays: what matters?, In Proc. Int. Workshop on Video Processing and Quality Metrics for Consumer Electronics, VPQM 09, Jan. 15-16, 2009, Scottsdale, Arizona, U.S.A.

[15] A. Gotchev, S. Jumisko Pyykkö, A. Boev, D. Strohmeier, "Mobile 3DTV System: Quality and User Perspective", European Workshop of Media Delivery (EUMOB‟08), Oulu, Finland, July 2008.

[16] A. Boev, D. Hollosi, A. Gotchev, and K. Egiazarian,” Classification and simulation of stereoscopic artifacts in mobile 3DTV content”, Proc. SPIE, Vol. 7237, No. 1, (2009).

[17] M. Yuen, “Coding Artefacts and Visual Distortions”, in (H. Wu. K. Rao eds), Digital Video Image Quality and Perceptual Coding, ISBN 9780824727772 , CRC Press, 2005.

[18] Hewage CTER, Worrall ST, Dogan S, Villette S, Kondoz AM. Quality evaluation of color plus depth map-based stereoscopic video. IEEE Journal of Selected Topics in Signal Processing 2009; 3(2):304–31.

[19] K. Seshadrinathan , R. Soundararajan, A. Bovik, L. Cormack,” Study of subjective and objective quality assessment of video”, IEEE Transactions on Image Processing 2010; 19(6):1427–1441.

[20] W. J. Tam, “Image and depth quality of asymmetrically coded stereoscopic video for 3D-TV,” Joint Video Team document JVT-W094, Apr. 2007.

[21] G. Saygili, G. Gurler, A. M. Tekalp, “Quality assessment of asymmetric stereo video coding,” in Proc. IEEE Int. Conf. on Image Processing (ICIP), Hong Kong, Sep. 2010.

[22] G. Gurler, K. Bagci, A. M. Tekalp, “Adaptive Stereoscopic 3D Video Streaming,” in Proc. IEEE Int. Conf. on Image Processing (ICIP), Hong Kong, Sep. 2010.

[23] P. Aflaki, M. M. Hannuksela, J. Hakkinen, P. Lindroos, and M. Gabbouj, “Subjective study on compressed asymmetric stereoscopic video,” in Proc. IEEE Int. Conf. on Image Processing (ICIP), Hong Kong, Sep. 2010.

[24] E. Kurutepe, T. Sikora, “Multi-view video streaming over p2p networks with low start-up delay,” in Proc. IEEE Int. Conf. on Image Processing (ICIP), San Diego, CA, 2008.

[25] K. Seshadrinathan , R. Soundararajan, A. Bovik, L. Cormack,” Study of subjective and objective quality assessment of video”, IEEE Transactions on Image Processing 2010; 19(6):1427–1441.

[26] ITU-R BT.1438, Subjective assessment of stereoscopic television pictures, Recommendation ITU-R BT.1438, ITU Telecom. Sector of ITU, 2000.

[27] ITU-R BT.500-11, Methodology for the Subjective Assessment of the Quality of Television Pictures, Recommendation ITU-R BT.500-11, ITU Telecom. Sector of ITU, 2002

[28] Hewage CTER, Worrall ST, Dogan S, Villette S, Kondoz AM. Quality evaluation of color plus depth map-based stereoscopic video. IEEE Journal of Selected Topics in Signal Processing 2009; 3(2):304–31.

[29] ITU-R, “Digital three-dimensional (3D) TV broadcasting,”Question ITU-R 128/6, 2008.

[30] ITU-T, “Objective and subjective methods for evaluating perceptual audiovisual quality in multimedia services within the terms of Study Group 9,” Question 12/9, 2009.

[31] H. Schwarz, D. Marpe, T. Wiegand, “Overview of the scalable video coding extension of the H.264/AVC standard,” IEEE Trans. on Circuits and Systems for Video Tech., vol. 17, no. 9, pp. 1103–1120, Sep. 2007.

[32] A. Vetro, T. Wiegand, and G. Sullivan, “Overview of the stereo and multiview video coding extensions of the H.264/MPEG-4 AVC standard,” in this issue.

[33] H. Schwarz, D. Marpe, and T. Wiegand, “Analysis of hierarchical B-Pictures and MCTF,” in Proc. IEEE Int. Conf. on Multimedia and Expo (ICME), pp. 1929–1932, Toronto, Ontario, Canada, Jul. 2006.

[34] Y. Chen, Y.-K.Wang, K. Ugur, M. Hannuksela, J. Lainema and M. Gabbouj, “The emerging MVC standard for 3D video services,” EURASIP Jour. on Advances in Signal Processing, vol. 2009, 2009.

[35] P. Merkle, A. Smolic, K. Muller, and T. Wiegand, "Efficient prediction structures for multiview video coding," IEEE Trans. on Circuits and Systems for Video Tech., vol. 17, no. 11, pp. 1461-1473, Nov. 2007

[36] S. Tao, Y. Chen, M. M. Hannuksela, Y.-K. Wang, M. Gabbouj and H. Li, ―Joint texture and depth map video coding based on the scalable extension of H.264/AVC,‖ in Proc. IEEE Int. Symp. on Cir. and Sys. (ISCAS), pp. 2253–2256, 2009.

[37] Buncha Kamolrat, Anil. Fernando, Marta Mrak, Ahmet Kondoz, "3D Motion Estimation for Depth Image Coding in 3D Video Coding," IEEE Trans. on Consumer Electronics, Volume 55, Issue 2, May 2009.

[38] C.T.E.R. Hewage, S. Worrall, S. Dogan, H. Kodikara Arachchi and A.M. Kondoz, "Stereoscopic TV over IP," Proc. of the 4th IET European Conference on Visual Media Production (CVMP'2007), London, UK, November 2007.

[39] HDMI Licensing, LLC., “HDMI Specification 1.4,” 2009.

[40] http://www.3dtv-research.org/

[41] E. Kurutepe, A. Aksay, C. Bilen, C. G. Gurler, T. Sikora,G. Bozdagi Akar, A. M. Tekalp, "A Standards-Based,Flexible, End-to-End Multi-View Video Streaming Architecture", PV'07, Lausanne, Switzerland, Nov. 2007

[42] E. Kurutepe, M. R. Civanlar, A. M. Tekalp, Interactive Transport of Multi-view Videos for 3DTV Applications, Journal of Zhejiang University Science, vol.7, no.5, 2006, pp.830-836.

[43] A. Vetro, W. Matusik, H. P. Pfister, and J. Xin, “Coding approaches for end-to-end 3-D TV systems,” in Proc. Picture Coding Symp. (PCS), San Francisco, CA, Dec. 2004.

[44] I. Sexton and P. Surman, “Stereoscopic and autostereoscopic display systems,” IEEE Signal Processing Mag., vol. 16, pp. 85–99, May 1999.

[45] Maneesh Agrawala , Andrew C. Beers , Ian McDowall , Bernd Fröhlich , Mark Bolas , Pat Hanrahan, “The two-user Responsive Workbench: support for collaboration through individual views of a shared space, Proceedings of the 24th annual conference on Computer graphics and interactive techniques”, p.327-332, August 1997.

[46] http://en.wikipedia.org/wiki/3-D_film

[47] Wojciech Matusik, Hanspeter Pfister. 3D TV: A Scalable System for Real-Time Acquisition, Transmission, and Autostereoscopic Display of Dynamic Scenes. ACM Transactions on Graphics. 24(3) 2004.

[48] Y. Zhu and T. Zhen, "3D Multi-view autostereoscopic display and its key technology", Asia-Pacific Conference on Information Processing, pp. 31-35, 2009.

[49] K. Mueller, P. Merkle, H. Schwarz, T. Hinz, A. Smolic, T. Oelbaum, "Multi-view video coding on H.264/AVC using hierarchical B-frames," *in Proc. of the Picture Coding Symposium*, Beijing, China, Apr. 2006.

[50] Krzysztof Klimaszewski, Krzysztof Wegner, Marek Domanski, "Distortions of synthesized views caused by compression of views and depth maps," *3DTV Conference: The True Vision - Capture, Transmission and Display of 3D Video, 2009 pages: 1 – 4*

[51] A. Aksay, S. Pehlivan, E. Kurutepe, C. Bilen, T. Ozcelebi, G Bozdagi Akar, M. R Civanlar,and A. M. Tekalp, "End-to-end stereoscopic video streaming with content-adaptive rate and format control," Signal Proc.: Image Communication, vol 22, pp 157-168, Feb 2007.

[52] N. Ozbek and A. M. Tekalp, "Unequal inter-view rate allocation using scalable stereo video coding and an objective stereo video quality measure," in *Proc. IEEE ICME*, Germany, June 2008. 54

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

[54] Quality Evaluation of Color Plus Depth Map-Based Stereoscopic Vide0

[55] Quality Analysis for 3D Video Using 2D Video Quality Models

[56] Video Quality Experts Group (VQEG), "Final Report From the Video Quality Experts Group on the Validation of Objective Models of Video Quality Assessment, Phase II," 2003 [Online]. Available: http://www. vqeg.org.

[57] Quan Huynh-Thu; Le Callet, P.; Barkowsky, M.; , "Video quality assessment: From 2D to 3D — Challenges and future trends," Image Processing (ICIP), 2010 17th IEEE International Conference on , vol., no., pp.4025-4028, 26-29 Sept. 2010

[58] Nukhet Ozbek, A. Murat Tekalp and Turhan Tunalı, “Rate allocation between views in Scalable Stereo Video Coding using an objective stereo video quality measure,” Acoustics, Speech and Signal Processing, 2007. ICASSP 2007. IEEE International Conference.

[59] “Subjective Assessment of Stereoscopic Television Pictures,” ITU, Recommendation BT.1438, 2000.

[60] “Methodology for the Subjective Assessment of the Quality of Television Pictures,” ITU, Recommendation BT.500–10, 2000.

[61] M. Tanimoto, T. Fujii, K. Suzuki, N. Fukushima, and Y. Mori, “Reference Softwares for Depth Estimation and View Synthesis”, ISO/IEC JTC1/SC29/WG11, M15377, April 2008.

[62] L. Meesters, W. A. Ijsselsteijn, and P. J. H. Seuntiens, “A survey ofperceptual evaluations and requirements of 3-D TV,” IEEE Trans. Circuits Syst. Video Technol., vol. 14, no. 3, pp. 381–391, Mar. 2004.

[63] S. Wenger, “Error patterns for Internet experiments,” ITU Telecommunications Standardization Sector, Doc. Q15-I-16r1, Oct. 1999.

[64] W. J. Tam, “Image and depth quality of asymmetrically coded stereoscopic video for 3D-TV,” Joint Video Team document JVT-W094, Apr. 2007.

[65] X. Hei, Y. Liu, and K. W. Ross, "Inferring Network-Wide Quality in P2P Live Streaming Systems," JSAC, vol.25, no. 10, Dec. 2007.

[66] R. Penno, S. Raghunath, R. Woundy, V. Gurbani, J. Touch “LEDBAT Practices and Recommendations for Managing Multiple Concurrent TCP Connections”, IETF draft, Feb. 26, 2000.

[67] Z. Liu, Y. Shen, K. W. Ross, S. S. Panwar, and Y. Wang, "Layerp2p: using layered video chunks in p2p live streaming," IEEE Transaction on Multimedia.,vol. 11, no. 7, pp. 1340-1352, 2009.