This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCSVT.2016.2527318, IEEE Transactions on Circuits and Systems for Video Technology

Adaptive Multi-View Video Delivery using Hybrid Networking

Erhan Ekmekcioglu, C. Goktug Gurler, Ahmet Kondoz, *Senior Member, IEEE*, A. Murat Tekalp, *Fellow, IEEE*

Abstract— **Multi-view entertainment is the next step in 3D immersive media networking owing to its improved depth perception and free-viewpoint viewing capability, where users can observe the scene from the desired viewpoint. This paper outlines a delivery system for multi-view plus depth video combining the broadcast and broadband networks. The Digital Video Broadcast (DVB) network is used along with adaptive Peer-to-Peer (P2P) distribution over the Internet to deliver high-volume multimedia to users. The DVB network has been used to deliver part of the 3D service owing to its robustness and wide availability, as a mechanism to guarantee the minimum 3D Quality of Experience (QoE). The developed system brings key contributions in the P2P transport for real-time multimedia delivery, including a user preference-aware adaptation mechanism, adaptive redundant chunk scheduling for robustness, incentives to decrease the load on the content server for improved system scalability, and resynchronisation capability with the DVB transmission. The introduced features are compared to those of some other well-known P2P solutions to highlight the quantitative gains. A subjective testing campaign has also been organised on the developed hybrid platform, which proves the effectiveness of user-aware adaptation over network-based adaptation on a Mean Opinion Score (MOS) scale.**

Index Terms— **Multi-View Video, P2P, 3D, Hybrid delivery**

I. Introduction

THE concept of 3D has been popular for a long time in the research community and in the entertainment industry. Cinema customers have widely embraced the 3D technology and the success of 3D in the cinema has led to renewed efforts to introduce 3D technology to the home. A number of on-demand and live stereoscopic 3DTV services have been deployed by various broadcasters in the past few years, such as Sky and Virgin in the UK, and ESPN in the US. Despite the availability of 3D content and displays, and the feasibility of delivering it over the existing systems, stereoscopic 3DTV services at home have not been widely popular. Factors such as the limited depth effect, lack of free-viewpoint interactivity and motion parallax, and partly the physical degradations such as the requirement to wear glasses, have resulted in less-than-expected immersive experience. These could be overcome by deploying multi-view video. Nevertheless, the existing broadcast networks and the IP transport techniques do not sufficiently scale up to carry the required amount of visual information. It will take a long time before the existing networking technologies are able to cope with multiple high quality views accompanied by side information, such as depth maps, demanded by some modern multi-view display prototypes. These displays are capable of synthesising as many virtual views as necessary in real-time with the help of the delivered high quality viewpoints [1]. Thus, it is necessary to consider new methods for delivering multi-view content to the home.

Multi-View plus Depth (MVD) stands out as the simple yet effective extension of the Multi-View Video (MVV) format with per camera dense depth information, which enables efficient Depth-Image-Based-Rendering (DIBR) [2] to produce virtual views from a limited number of source views. For stereoscopic 3D video, the frame-compatible representation has been a widely deployed choice despite the inherent reduction in the spatial resolution of both video components [3]. The MPEG-4/H.264 AVC [4] standard allows the signalling of such frame-packing information within the Supplemental Enhancement Information (SEI). Multi-View Coding (MVC) standard [5], [6] has been the successor with the implementation of Disparity Compensated Prediction (DCP) exploiting the inter-view redundancies to improve the overall rate-distortion performance. However, it became apparent that even MVC is not sufficient to keep the multi-view bit-rate within manageable limits as the number of source cameras increases. It is impractical to encode and transport the entire set of views required to be displayed on most multi-view displays. Hence, it was necessary to define a new framework of 3D Video (3DV) coding that is able to deliver a limited number of camera views with dense depth information to render a higher number of virtual views at high quality. Two different standardisation efforts have emerged, both exploiting cross-domain similarities in the compression cycle (i.e., multiple views and depth maps), but providing compliance with different base standards, namely MVC (such as in MVC + D [7]) and High Efficiency Video Coding (HEVC) [8], [9]. The primary advantages of the new 3DV paradigm are to

Manuscript received June 10, 2015; revised November 15, 2015; accepted January 14, 2016. This paper was recommended by Associate Editor B. Pesquet-Popescu. The work in this paper was supported in part by the ICT COST Action 3D-ConTourNet (IC1105) and in part by DIOMEDES Project (Grant No. 247996), both funded by the European Commission.
E. Ekmekcioglu is with Institute for Digital Technologies, Loughborough University London, E20 3BS, UK (email: E.Ekmekcioglu@ lboro.ac.uk)
C. G. Gurler is with Anvato, Mountain View, CA 94043, USA (e-mail: goktug_gurler@anvato.com)
A. Kondoz is with Institute for Digital Technologies, Loughborough University London, E20 3BS, UK (email: A.Kondoz@lboro.ac.uk)
A. M. Tekalp is with College of Engineering, Koc University, Istanbul 34450, Turkey (email: mtekalp@ku.edu.tr)

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCSVT.2016.2527318, IEEE Transactions on Circuits and Systems for Video Technology

enable depth adaptation for a wide range of different 3D display devices (e.g., increasing or decreasing the depth effect as necessary) and to support multi-view autostereoscopic displays with various numbers of output views [7].

What is equally important as the employed compression technique is the transportation technique. Delivery of multi-view video over the Internet is the most flexible solution that can provide different transmission rates to different users according to their context (i.e., available bandwidth and viewing conditions). However it comes with its problems. Network failures, such as packet losses due to various reasons, pose a problem. Similarly, the scalability of the network in terms of the number of served users is another challenge. Error resiliency and scalable coding techniques can cope with network induced problems by facilitating efficient bit-rate adaptation. Peer-to-Peer (P2P) overlay networks are effective to address the issue of delivering high-volume multi-view data to many users. Adaptive streaming has recently gained momentum especially within the context of HTTP based video applications that work in a server-client model. MPEG's Dynamic Adaptive Streaming over HTTP (DASH) is such an example [10], [11] where various encoded representations of videos are stored in the web servers each at different bit-rates. Other popular examples include Adobe's HTTP Dynamic Streaming [12], Apple's Live Streaming [13], and Microsoft's Smooth Streaming [14]. Considering multi-view video delivery to very large groups of consumers simultaneously, the client-server model becomes inefficient putting a burden on the media servers. Nevertheless, adaptive streaming of 3D media over P2P overlay networks has not been as widely studied as other client-server based models.

Terrestrial broadcast, such as DVB, is the most widely deployed network for delivering multimedia content to homes. Its availability and reliability is higher compared to broadband networks. However, it has not enough capacity to carry compressed multi-view content that has significantly higher bit-rates than the stereoscopic content. It is possible to transmit the multi-view video jointly through the DVB and the broadband networks, where part of the content is carried over the DVB network and the rest is carried over the P2P overlay network. The peers become the multi-view media consumers. The wider access to DVB ensures that the users (used interchangeably with peers in the rest of this paper) have a minimum level of guaranteed QoE. However, the independence of these networks imposes the need for their synchronisation. The delay differences between the two networks need to be compensated for in order to ensure correct multi-view rendering.

There are rather a limited number of contributions in the literature that make combined use of the broadcast and the IP networks for high-volume video distribution. A system is constructed in the proposed work, where the DVB network and a P2P overlay network transport multi-view media together. The proposed novel P2P overlay distribution network combines several features, including resynchronisation capability with the DVB; user-preference-aware multi-view video adaptation; incentives to minimise the server usage; maximisation of peer-to-peer link utilisation for improved system scalability; and deployment of redundant streaming based on the buffer occupancy to prevent packet losses. The performance analysis has been done using scalable coded multi-view test videos. A combination of peers installed in various cities across Europe and a world-wide P2P platform has been used, which consists of globally distributed nodes connected with each other without firewall restrictions.

The remainder of the paper is as follows: Section II describes the related work. In Section III, the proposed delivery architecture is outlined in detail, paying particular attention to the features of P2P multi-view video delivery. In Section IV, the objective and the subjective test results are presented. Finally, Section V gives the concluding remarks.

II. Related Work

This section explains the recent works in the literature that are related to the content of this paper.

A. *Studies related to adaptive stereoscopic video streaming*

Adaptive streaming in the context of 3D stereoscopic video has been widely studied. In this context, asymmetric coding has been a popular technique. The authors in [15] compared the performances of symmetric and spatially asymmetric stereoscopic video coding. It was concluded that the performance of spatially asymmetric coding, where one of the views is encoded at half resolution in both dimensions, is similar to that of the symmetric and asymmetric quality coding at full resolution. However it has been found out that lower subsampling rates reduce the perceptual performance. A temporal-scaling based asymmetric coding technique was presented in [16] suggesting to reduce the temporal resolution of one of the views by half. SNR-scaling based approaches are more common, leading to a smoother 3D video viewing experience than the other approaches. Scalable Video Coding (SVC) extension [17] of the AVC [4] standard has been widely used in this scope. In a comparison between symmetric and asymmetric 3D video adaptation using SVC, it was observed that the performance gain of asymmetric coding changes at different operating points [18]. Especially at higher bit-rates, the asymmetric quality adaptation has proved to be superior in terms of maintaining the 3D perceptual quality. In [19], the impact of utilising visual attention models in asymmetric stereoscopic video coding was studied. It was suggested that the unequal treatment of regions with different saliency can further improve the performance of conventional quality-asymmetric stereoscopic video coding.

B. *Studies related to adaptive multi-view video streaming*

Since the multi-view displays are not widely available, the delivery of a complete multi-view video ensemble to users is not common. This is primarily for maximising the bandwidth efficiency. Besides, users are able to watch the scene from one viewing angle at a time, which is related to a sub-set of views. Selective (or interactive) multi-view streaming involves the retrieval of a sub-set of the views depending on the viewing context [20]. The delivered set of views can dynamically change depending on the users' viewing preferences (i.e., viewpoint selection), triggering adaptation as in the case of network variations. Previously, an IP-based multi-view streaming architecture was proposed that selectively transmits the packets of encoded viewpoints depending on the tracked head position [21]. Frequent changes in the viewing preferences can hinder the compression performance by leading to more frequent use of switching- or I-frames in the multi-view stream. This problem was addressed by the use of redundant frame structures, consisting of I- and P- frames at various qualities, appropriately trading off between transmission and storage [22]. In a more recent work, the authors have proposed to use a content replication strategy [23] to reduce the view switching cost in the domain of interactive free-viewpoint video streaming. In that work, the users are served by the viewpoint that is nearest to their selection if the requested view is unavailable in a nearby server (for fast streaming) while the remote server only transmits a pre-encoded view differential. The aforementioned works are effective in the Content Delivery Network (CDN) type applications, where a number of local content servers are maintained to support a larger base of users. Another recent work addressed the interactive delivery of the stereoscopic scene of interest (out of a multi-view video) over wireless transmission networks in an error-resilient way by exploiting Multiple Description Coding (MDC) [24]. The proposed approach generates a couple of descriptions from the texture-plus-depth videos of the left and the right views of the associated views of interest, which are then transmitted through separate wireless channels to exploit path diversity. Affected descriptions are recovered using a combination of the temporal and the inter-view predictions. Free-viewpoint video has also attracted attention in the domain of video conferencing applications. In [25], the authors have proposed an error-resilient streaming scheme for free-viewpoint video conferencing. The presented approach favours the more reliably reconstructed view in the decoder-side view synthesis process and employs an encoder-side optimised reference picture selection technique for real-time operation.

C. *Studies related to multi-view video streaming over P2P*

Dissemination of multi-view video packets over P2P networks augments the utilisation of the peers' networking resources, thus reducing the burden on the content servers and increasing the service scalability. Various approaches exist in the literature to form P2P overlays, i.e. tree-based, mesh-based, or a combination of both. It had previously been observed that commercial P2P deployments that rely purely on tree-based solutions do not exist [26], mainly due to the difficulty of maintaining the optimum tree structure in the cases of frequent peer churns and bandwidth fluctuations. A comprehensive comparison of the P2P solutions for video streaming had shown that mesh-based methods consistently exhibit superior performance over tree-based methods [27]. BitTorrent [28] is a receiver-driven, mesh-based P2P overlay topology that has originally been designed to share files. Tribler [29] is one of the commonly used video sharing platforms compatible with BitTorrent. It utilises a windowing mechanism to enable chunk scheduling in a time-aware manner, and features a give-to-get type incentive mechanism. Based on its specifications Tribler has no application layer concept, since the chunks are filled with multimedia data in a chop-and-ship manner. In chop-and-ship approach that maps video bit-stream into Torrent chunks, Network Abstraction Layer (NAL) unit boundaries are disregarded and a NAL unit can be split into two chunks. Since a partially delivered NAL unit is useless, in the events of chunk losses, it will not be possible to decode the following chunk. P2PNext [30] is another scheme that is based on Tribler. It targets scalable video distribution over P2P networks and aims at augmenting the received video quality by performing rate adaptation. In that system, padding is used to align the Group of Pictures (GOP) boundaries to the fixed sized chunks. Constant Bit-Rate coding (CBR) technique is adopted in order to generate bit-streams at a certain rate [31]. CBR is useful to correctly estimate the buffer duration by the peer, which influential on the efficiency of network adaptation. However, it is less bandwidth efficient than Variable Bit-Rate (VBR) coding. Mesh-based topologies are more robust to transient peers, despite an increase in the delay for peer search. The start-up delay characteristics of purely mesh-based topologies can be improved by grouping peers and assigning to them powerful seeder peers, who would retrieve the multi-view content directly from the content provider in a push-based fashion. In [32], peers were allowed to retrieve simulcast encoded multiple views' streams from different peers (depending on packet availability) to hold them in disjoint buffers in a mesh-topology, which led to improved flexibility. In [33], it was proposed to improve the robustness of low delay P2P multi-view streaming system by grouping peers based on their geographical locations and routing the multiple description multi-view streams through separate trees. For each peer, multiple fall back trees were also calculated dynamically to compensate for the negative impact of peer churns. This approach however brought demand for extra computing resources and signalling overhead. In another work, views within an MVC bit-stream were also forwarded through multiple independent multicast trees [34]. Authors in [35] proposed to construct two types of overlays simultaneously: one type comprising peers only watching the same view (intra-view) and another type comprising peers watching different views (cross-view). A reduced view switching delay was achieved, since every peer has also several neighbour peers in the cross-view overlay. This prevents the peers to leave an overlay completely and try to join another overlay from scratch. In [36], the authors optimised the allocation of the common anchor view streams to groups of peers depending on the view synthesis distortion and network reconfiguration cost. The access cost of common anchor views are collectively shared by peers at the expense of higher distortion.

III. Proposed Multi-View Video Delivery System

This section outlines the novel P2P architecture, generation of the scalable multi-view content, used transport protocols and the streaming related features (e.g., chunk selection mechanism, adaptation, redundant streaming for improved

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCSVT.2016.2527318, IEEE Transactions on Circuits and Systems for Video Technology

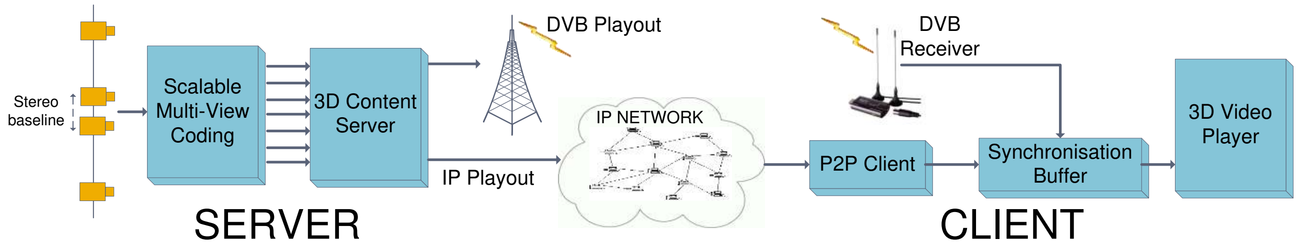

Fig. 1. Overview of the hybrid multi-view video content delivery architecture (note: only one client has been depicted for clarity)

robustness) in detail.

A. Overview of the proposed hybrid content delivery system

An overview of the hybrid content delivery architecture is shown in Fig. 1. The scalable encoded multi-view video bit-stream is packetised in the 3D content server to be sent out to the end-users via the DVB and the IP networks. The IP delivery block is built on top of the BitTorrent protocol that was introduced in Section II.C.

The work presented in this paper extends BitTorrent by introducing: i) variable length chunks as the transmission unit, ii) buffer driven chunk scheduling policy, iii) score based centralized incentive mechanism facilitating adaptive streaming. Variable length chunk format minimises the effect of data losses by aligning the GOP boundaries with the chunks for proper application layer framing. The buffer driven chunk scheduler ensures rate adaptive video delivery and minimises the bandwidth consumption at the server. The scheduling policy is also combined with a centralized incentive mechanism to restrict the video service to a quality that is proportionate to the peers' contribution in content distribution. The incentive mechanism utilises the sharing history of a peer to adjust its chunk reception rate from its neighbours when it has just joined a session, which prevents long delays as a consequence of the tit-for-tat mechanism. Before describing the P2P delivery system in further detail, the following sub-section explains the content creation process for distribution.

B. Content creation for distribution

The SVC standard [17] has been used to individually encode each view and the depth maps. In this work, quality (SNR) scalability and viewpoint scalability are used. Each view is encoded using two SNR layers (one base and one enhancement layer), where each layer has approximately the same bit-rate. This way, a sensible bit-rate adaptation range is reached, where the target bit-rate is adjusted by truncating a layer. Depth maps are encoded using a single layer. VBR is used for all views and depth maps, which has the advantage of achieving a better compression ratio than CBR coding.

After the scalable video bit-streams are generated, the quality layers and the depth maps of each view are split to separate streams and then encapsulated into MPEG-2 Transport Streams (TS). MPEG-2 Transport Streams consist of fixed length transport packets (188 bytes each). Each video Access Unit (i.e., frame) that consists of one base layer NAL unit and one quality enhancement layer NAL unit is split to two blocks, one containing the base layer NAL unit and the other containing the enhancement layer NAL unit. Each block is wrapped into individual Packetised Elementary Streams (PES) packets. Then, all PES packets are further fragmented into smaller fixed-length chunks to form the transport stream (TS). TS encapsulation is a mandate for multimedia transmission over the DVB system. In order to ease the synchronisation of the streams delivered over the DVB and the P2P overlay networks, the same encapsulation format is deployed for P2P. The same Program Clock Reference (PCR) is used to inject the Presentation and Decoding Time Stamps (PTS and DTS) into the transport streams of both delivery networks. MPEG-2 TS encapsulation brings approximately 4% overhead in contrast to pure RTP encapsulation that is a more popular encapsulation method for IP streaming and that brings approximately 2% overhead. Nevertheless, the difference is not significant that would impact the transport performance. Fig. 2 depicts the encapsulation procedure for one of the views.

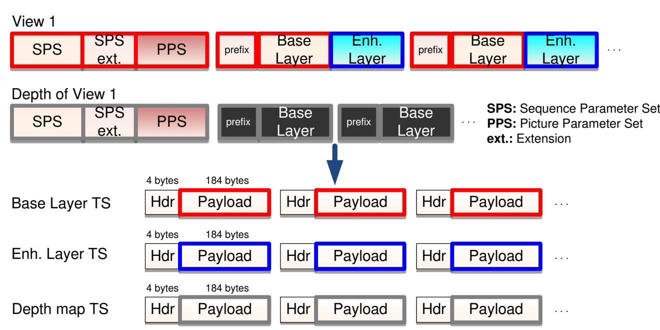

Fig. 2. Splitting SVC stream to multiple MPEG-2 Transport Streams

Note that the videos that are sent over the DVB network are stuffed with NULL[1] transport stream packets to produce constant bit-rate streams as implied by the DVB standard. This is not the case for the streams delivered through the P2P overlay network. The chunks for the P2P transport are generated by grouping together the fixed-length TS packets that make up a whole GOP (fixed number video frames). Separate chunks are generated from the base layer, enhancement layer, and the depth map transport streams. And each transport stream has a unique identifier, referred to as Programme ID (PID).

[1] NULL packets are used for stuffing the MPEG Transport Stream to achieve constant bit-rate streaming and carry no multimedia-related or other information.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCSVT.2016.2527318, IEEE Transactions on Circuits and Systems for Video Technology

As previously mentioned in the introduction section, part of the multi-view content is broadcast to all users over the DVB network to guarantee a minimum service. This corresponds to the AVC-compatible base layer transport streams of the central stereoscopic camera pair. This pair has a stereoscopic baseline (as shown in Fig. 1) and can directly be displayed on a stereoscopic 3D display after decoding. All other transport streams, including the enhancement layers and the depth maps of the central stereoscopic camera pair and the transport streams associated with the side-most cameras are delivered over the P2P overlay network in synchrony with the DVB stream. Fig. 3 depicts the association of the multi-view streams with the delivery networks. When a peer receives all four views and the depth maps, other intermediate virtual views can be synthesised inside a multi-view rendering after decoding. As a result, the peer can enjoy the multi-view content either on a compatible multi-view auto-stereoscopic display or on a conventional 3D stereoscopic display with free-viewpoint navigation capability.

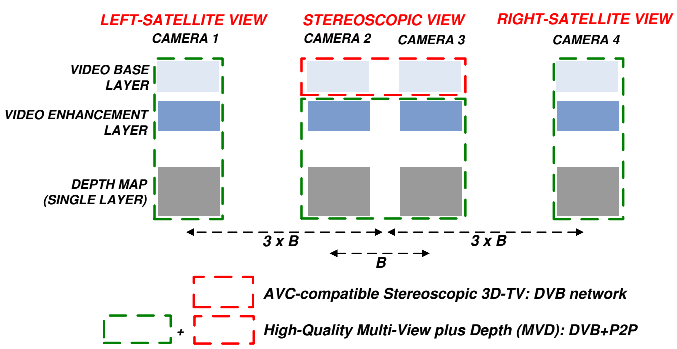

Fig. 3. Association of the multi-view streams with the delivery networks

A content metadata file that is in XML format is generated during the encoding and encapsulation processes. The content metadata file is used in the content server. The metadata is inserted into a special chunk that is exchanged with peers to let them correctly decode and display the content. Each processed MVD content is registered in the content server with a title and a Globally Unique Identifier (GUID). Apart from the title, GUID and tags, multi-view geometry related information is also involved in the content metadata file. This metadata is used by peers' media players to correctly render the 3D scene. The number of cameras, extrinsic and intrinsic camera parameters, scene setup (cameras' positions with respect to each other and with respect to the centre of the scene), the PIDs of all involved transport streams and their types (e.g. base, enhancement, depth map) are the other components of the content metadata file.

C. P2P multi-view content delivery

The generated chunks in the multi-view content server are requested and delivered using the HTTP protocol. HTTP protocol has been a powerful option for adaptive video streaming solutions. Using HTTP, it is possible to create pull-based streaming mechanism that is controlled by the peer. The content server, which is also called Main Seed Server (MSS) in the rest of the paper, additionally operates as the tracker server and helps new peers find other peers in the swarm. The tracker server informs peers about the current state of the swarm. In BitTorrent, the tracker server randomly forwards a subgroup of peers upon request from a newcomer. Instead of selecting peers randomly, authors in [37] claimed that buffer maps can be utilised to monitor the network characteristics of a peer and proposed to match peers based on the states of their buffer maps. Based on this idea, in this work the tracker server clusters the peers according to the requested multi-view streams. Since rate adaptation is enabled thanks to SVC, peers that request the same set of streams can be considered to have similar bandwidth capacity. When a peer connects for the first time, the tracker forwards a random subgroup of peers, since the views that are requested by this peer are initially unknown. In the following iterations, the peer forwards the PIDs of the transport streams that it is interested in. The tracker server updates its entry for that particular peer and then returns a sub-list of peers that are interested in the same set of streams.

Modern P2P video streaming solutions use a windowing mechanism for timely delivery. This is done by restricting chunk scheduling inside a window and randomising the chunks among peers [38]. Randomisation is crucial, because if each peer downloads the same chunk, they cannot exchange distinct chunks among themselves. In the proposed solution, separate windows are used for each transport stream. The windows can be synchronised or slightly shifted according to the state of the downloading process. Fig. 4 depicts these windows. Only the chunks inside a window can be scheduled for download (i.e., chunks in dashed line frame with yellow background). There can be some chunks that are downloaded but not played yet (light green), or currently being downloaded (yellow chunks in straight line frame).

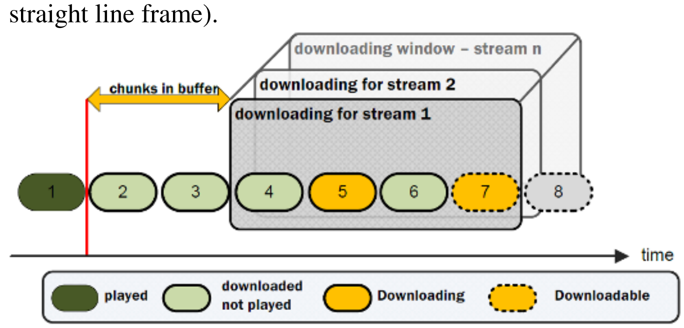

Fig. 4. Utilised windowing mechanism for downloading the chunks

C.1 Adaptive streaming

The adaptive streaming is implemented in two steps. The first step is the stream selection performed each time a download window slides. The second step is adaptive chunk selection performed each time a chunk is going to be scheduled within a download window. Once streaming starts, peers request the chunks of all quality layers to receive at the highest possible quality. Once the window is downloaded, a peer advances its window to schedule new chunks. Stream adaptation is based on the streams' relative prioritisation order, which is updated according to the peer's viewing preferences. This process is explained in Section III.C.2. The number of streams to be scheduled is updated each time the window slides. If the buffer duration is below a certain threshold, the number of downloaded streams is reduced by discarding the streams at the bottom of the prioritisation order. The buffer duration is determined by the duration of received sequential chunks that can be decoded in all streams. In the opposite case, where the duration of the buffer is once again over the threshold, then the next stream with the highest priority is added to the list of actively downloaded streams to match the capacity. Note that as the bandwidth permits, the lowest priority streams continue to be downloaded to compensate for the delays in frequent viewpoint changes. Once the number of streams is determined, chunks that are within the new window can be scheduled for download.

Chunk scheduling is performed in further two steps. Fig. 5 depicts the chunk selection process. First, a candidate stream is selected based on the prioritisation weights of the active streams. The probability of selecting the i^{th} stream (p_i) is the ratio of its weight (w_i) to the sum of the weights of all streams as shown in Equation 1.

$$p_i = \frac{w_i}{\sum_{i=0}^{n} w_i} \tag{1}$$

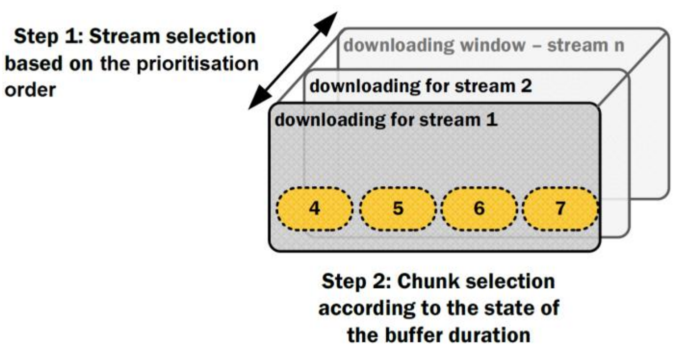

Fig. 5. Chunk selection process

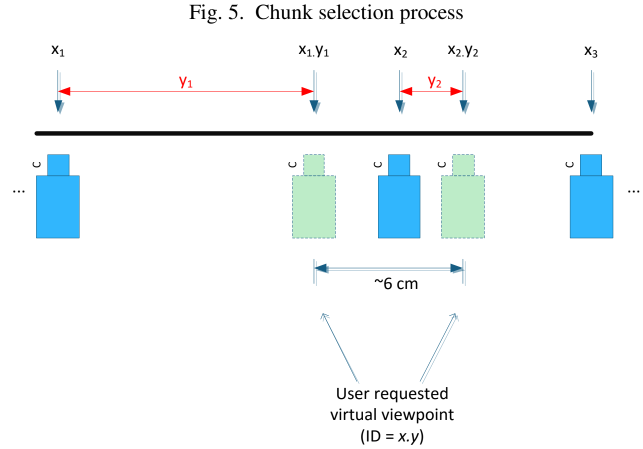

Fig. 6. Required views to render the requested virtual view pair

The second step is to determine the chunk request order for the selected stream. The criterion is the associated buffer length. If it is short, chunks are requested in a weighted random order that gives higher chance to the chunks close to the play-out deadline to be scheduled first. If the buffer is long, chunks that are available in the swarm are scheduled first. If no chunks are available in the swarm, then they are selected randomly within the window. Randomisation increases the diversity among the peers and augments the P2P communication, which decreases the overall bandwidth requirements on the content server.

C.2 User preferences based stream prioritisation

When the user changes his/her preferred viewpoint, higher streaming priorities are assigned to the streams associated with the closest three camera views that are required to synthesise the requested stereoscopic view pair, as shown in Fig. 6. Assuming that the requested viewpoint's position is represented as a floating number, the following steps are applied to perform the prioritisation action:

1. Calculate the virtual view pair $x_i.y_i = \{x_1.y_1, x_2.y_2\}$ to be rendered, based on the user input and the stereoscopic baseline (considered as 6 cm).
2. Determine the closest three camera views (x_i, x_{i+1}, x_{i+2}) for rendering the virtual view pair.
3. Calculate the fitness of those camera views for synthesising the virtual view pair. While calculating the fitness of a camera, it is assumed that the virtual view pair is synthesised using only that camera.
4. Assign priority orders to (x_i, x_{i+1}, x_{i+2}) based on the fitness computed in step 3. Higher fitness scores indicate higher ranking.
5. Assign the priority order for the remaining cameras based on the same constraint, such that they continue to be downloaded if there is sufficient bandwidth surplus.

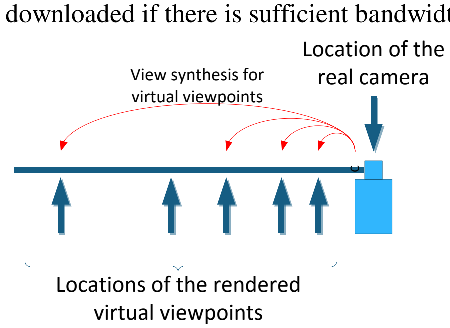

Fig. 7. Quality projection from a real camera to the requested virtual view

The fitness criterion refers to the projection of the objective quality of an existing camera view to the location of the user requested virtual viewpoint as depicted in Fig. 7. Projection was performed by a function derived through a subjective assessment, in which the quality of the rendered views at various selected distances from a real camera position was assessed (in 2cm steps up to 20 cm away from the camera). The results were displayed on a stereoscopic display to 15 participants. The visual quality of the virtual views presented an exponential behaviour with respect to the distance from the real camera. Hence, the perceptual quality of the virtual view is modelled as:

$$Q_{VirtualCam} = Q_{FixedCam} - a \cdot \left(1 - e^{b \cdot x}\right) \tag{2}$$

where x is the distance between the real camera and the rendered virtual view (in cm), which grows in the negative direction (i.e., from 0 to -20). The values for a and b were

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCSVT.2016.2527318, IEEE Transactions on Circuits and Systems for Video Technology

determined empirically using a curve fitting technique based on the MOS results for two of the synthetic multi-view plus depth sequences used in the tests, which are *Dancer* and *T-Rex* described in Section IV.C. The computed a and b values (0.67 and 0.05, respectively) have been found to be in agreement with the MOS results obtained for the other two test multi-view plus depth sequences, verifying the applicability of the distance based perceptual quality estimation model.

The fitness value for a camera is computed by averaging the corresponding $Q_{VirtualCam}$ values within the virtual stereoscopic camera pair.

C.3 Robustness against ungraceful peer departures

The proposed delivery solution also features a robustness mechanism that includes scheduling and downloading certain packets redundantly from multiple peers. The idea is that if ungraceful peer departures (peer churns) happen in the course of chunk transmission or if some peers' network conditions change abruptly, unaffected peers should still be able to proceed with decoding. However, it is important to control the amount of redundancy injected to the network. In the proposed mesh-based topology, each peer has the advantage of deciding when to request a chunk, from whom to request, and which layer to request. In the adopted approach, redundant packets are requested only when necessary in order to keep it under control. For example, a peer is to request a chunk that is available in peers with low upload capacity. If that peer has long buffer duration, then it does not have to schedule redundant packets, because it can retry downloading a chunk from another peer if the current transmission fails. On the other hand, if the buffer has a short length and the deadline is close, then it is safer to request the same chunk from multiple peers to receive at least one copy. We also identify the base layer and the enhancement layer packets as the redundant and the unique data, respectively. To minimise the overhead, only the base layer packets are sent redundantly when needed, whereas the enhancement layer packets are not scheduled from multiple peers at the same time.

The adaptive redundant streaming process runs as follows: A peer continuously ranks its neighbours according to their upload rate. When a chunk is scheduled for download, the candidate peers are sorted based on their upload rate. If a peer with sufficiently high upload capacity is available, then the chunk is requested only from that peer. However, if no such peers are available, then the state of the buffer determines the next step. If the duration of the buffer is long, the risk of scheduling a chunk from a single peer with low upload capacity is taken. However, if the buffer is low, then requests from two different peers are made. The corresponding state diagram is illustrated in Fig. 8.

C.4 Methods to reduce the server's bandwidth requirements

The proposed system features two methods to augment the P2P data distribution and decrease the burden on the MSS. The first one is about adjusting the size of the download window dynamically, and the second one is choosing the chunks in an intelligent way to increase the portion of P2P data distribution.

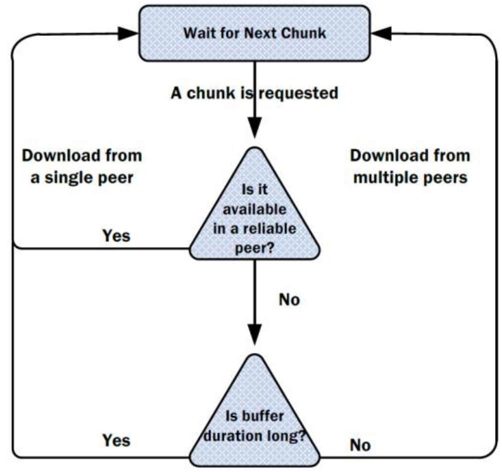

Fig. 8. Adaptive redundant packet streaming scheme

In BitTorrent protocol, a peer can choose an arbitrary chunk to download. Usually a chunk, which is distributed least in the swarm, is requested. By introducing a windowing mechanism for video delivery, the efficiency of this approach is reduced due to the limited number of chunks that the window includes. To compensate for this side effect, dynamically adapted window sizes are used. Each time a window is complete, the P2P software calculates the duration of the buffer. If the duration of the buffer is larger than the previously calculated value, the window size is increased accordingly. The current duration of the available buffer and the current download rate are used to determine the new window size. Each time the window slides, the size of the new window is determined based on the duration of playable video in the buffer. The new window size is calculated as

$$w_s = \frac{d_r \times b_d}{c_s} \tag{3}$$

where, w_s represents the window size (in units of chunks), c_s represents the average chunk size (in bytes), d_r represents the download rate (in bytes/second) and b_d represents the buffer duration (in seconds). The result is rounded down to obtain an integer value. As a result, the chances of peers having distinct chunks in their buffers are also increased, hence increasing the likelihood of P2P activity instead of direct download from the MSS. This feature is enabled only when all streams are active. If a stream is discarded due to bandwidth scarcity and as a result of adaptation, the window size is not increased over time, because it is more difficult to fill the buffer with a large window size.

The adaptation and chunk scheduling decisions are always performed with the state of the buffer in consideration. In the cases of bandwidth surplus, a peer will have long buffer duration. Thus, it can slow its pace of downloading in order to avoid putting more stress on the MSS. In such cases, a peer enters to a specific mode, in which the number of chunks requested from the server is decreased. Each time a chunk is downloaded from the server, it would need to wait for a certain amount of time to be able to schedule another chunk from the

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCSVT.2016.2527318, IEEE Transactions on Circuits and Systems for Video Technology

server. In this particular mode, there is no limitation on video chunk exchange between the peers in the swarm.

C.5 Resynchronisation with DVB

Due to the chaotic nature of the IP it may not keep up the synchronisation with DVB at all times. According to Fig. 3, the video base layer streams of cameras 2 and 3 are delivered on the DVB channel. At times when resynchronisation action needs to be taken, the P2P download windows for each stream slide in an exponentially increasing manner. This process is depicted in Table I. If the last window cannot be completed on time, then it skips one chunk and then slides. If it fails again, it skips two chunks, then next time four chunks, and so on. At the point where the P2P delivered stream catches up with the DVB channel, the exponential window sliding process rewinds. The time it takes to resynchronise with the DVB stream is analysed in Section IV.

TABLE I

PSEUDOCODE FOR THE RESYNCHRONISATION OF THE P2P AND DVB LINKS

Function: Resynchronise P2P and DVB

```
Time_diff = PCR_DVB – PCR_P2P
slide_offset_in_chunks = 0
i = 0
do
{
   Slide_Window(slide_offset_in_chunks)
   slide_offset_offset_in_chunks = 2 ^ i
   i = i + 1
   Time_diff = PCR_DVB – PCR_P2p
}
while (Time_diff > 0)
```

C.6 Improved P2P link utilisation

Most of the previous studies assign a single chunk to a peer at a given time. This brings an inherent inefficiency, where the Round Trip Time (RTT) is not taken into consideration. The inefficiency arises from the fact that the delay in delivering individual chunk requests is very close to the RTT. Considering a typical value for RTT, such as 50ms, and assuming that each chunk carries 500 ms long video segment, then in one tenth of the time the channel is underutilised. Increasing the chunk size can be considered as a remedy. However increasing the chunk size decreases the adaptation capability due to lower granularity. The proposed work uses the "double scheduling" concept to compensate for the inefficient channel usage during chunk scheduling. With the double scheduling policy, a peer can request two chunks from another peer in one go. The sender peer queues these requests and forwards only a single chunk at a time. When the first chunk in the queue is transferred, the peer requests another chunk immediately to maintain a steady flow without delays. In other words, using double scheduling, the proposed P2P solution utilises the IP connection more efficiently.

D. Summary of differences from currently available solutions

The differences of the proposed system from other solutions, such as Tribler and P2PNext can be summarised under two categories, as below.

Chunk generation: Both Tribler and P2PNext use fixed sized chunks. Tribler does not use padding to generate self-decodable video chunks. Thus, Tribler cannot estimate the duration of decodable video in its buffer without parsing the content. In its specifications, this fact has been stated to cause frequent re-buffering when the chunks in the buffer are not useful due to prediction structure of the video [39]. In the P2PNext system, GOP units are split to different chunks using padding. In order to keep the overhead of padding to a minimum, CBR coding is performed [31]. However, to avoid quality losses, usually a high video bit-rate is chosen for CBR coding, which results in inefficient bandwidth usage for IP networks. In the proposed work, chunks are formed using GOP boundaries. It is possible to determine the duration of decodable data in the buffer and perform more informed adaptation.

Chunk picking: When scheduling a chunk, a windowing mechanism is used to enable timely download. The P2PNext system uses layered windows, such that a chunk from a more important stream is more likely to be selected. Similarly, a chunk that has closer play-out deadline is more likely to be picked. This is depicted in Fig. 9, where the entries represent the weight of a chunk that is used to calculate the probabilities. The proposed system in this paper provides dynamic multiple windows with adjustable size over time. The window size is increased if the download rate is high enough, such that the randomisation among chunks is maximised and the P2P activity is boosted. Furthermore, the proposed system utilises adaptive redundant streaming approach by requesting the base layer chunks from two different peers if there is a risk of not receiving a chunk due to poor upload capacity of peers.

	t	t+1	t+2	t+3	t+4	t+5	t+6	t+7	t+8
E3	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
E2	1.0	1.0	1.0	0.7	0.3	0.0	0.0	0.0	0.0
E1	1.0	1.0	1.0	1.0	0.8	0.5	0.0	0.0	0.0
Base	1.0	1.0	1.0	1.0	1.0	0.9	0.5	0.0	0.0

Fig. 9. P2PNext chunk scheduling weightings in sliding window [34]

IV. EXPERIMENTAL RESULTS

The performance of the proposed hybrid MVV content delivery solution has been tested using several test-beds (i.e., a Europe-wide distributed peers' network, PlanetLab platform [40], and a controlled LAN environment). Several test scenarios are used, where each scenario is anticipated to evaluate certain aspects of the proposed content delivery solution. These are depicted in IV.B. Finally, a subjective test is performed using the peer that is with a real-time multi-view media player and a 3D display setup in order to evaluate the subjective performance of the developed platform.

A. Test beds

A.1 Europe-wide distributed peers' network

A test network has been formed that consists of 4 sites and 5 nodes. One of the nodes is based in Munich, Germany; one in Ilmenau, Germany; one in Istanbul, Turkey; and two in

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCSVT.2016.2527318, IEEE Transactions on Circuits and Systems for Video Technology

Guildford, UK. The node based in Munich is only used as the content server, i.e., Main Seed Server (MSS). The test node located in Istanbul is open to the Internet over certain ports with no firewall restriction. Hence, other foreign peers from the PlanetLab test bed can connect to this network through the node located in Istanbul in order to increase the number of the peers in the system. All of the nodes have acceptable link capacities as depicted in Table II. The reported link capacity in Table II represent the one-way TCP throughput between two nodes, measured using the iperf tool over a 50 second duration [41]. Measurements are done at the same time as the tests.

TABLE II
LINK CAPACITY BETWEEN THE DISTRIBUTED NODES (MBPS)

	Sender			
Receiver	Ilmenau	Guildford	Istanbul	Munich
Ilmenau	-	3.99	0.74	6.42
Guildford	2.84	-	1.45	3.33
Istanbul	1.35	2.14	-	2.24

A.2 PlanetLab

PlanetLab is a geographically distributed overlay platform designed to support the deployment and evaluation of planetary-scale network services. Researchers are able to request a PlanetLab slice, in which they can experiment with a variety of services, including file sharing, network-embedded storage, content distribution, routing and multicast overlays, QoS overlays, scalable object location, scalable event propagation, anomaly detection, and network measurement tools. PlanetLab consists of 1337 nodes at 681 sites as of May 2015, and over 1000 active research projects run currently on the PlanetLab.

Experimenting over the PlanetLab has some drawbacks too. The available peers are shared by multiple services, making it difficult to obtain stable results. Due to this limitation, multiple trials have been performed to obtain stable results.

A.3 Controlled LAN environment

Although being extremely useful, PlanetLab cannot always provide the best conditions to test a particular implementation. This is mainly because of the inability to control the conditions over the Internet. To test the adaptation capability of a peer, a particular connection bandwidth needs to be set. In order to create the desired conditions that are realistic and can occur over the Internet in long run, software network emulators are utilised. These tools enable restricting the bit-rate to a certain level to trigger the adaptation behaviour of the proposed solution. Hence, it is possible to create a P2P network with multiple work stations in a LAN environment.

B. Objective evaluation scenarios and test results

B.1 Stable and robust content delivery

In this scenario, the efficiency of the adopted adaptive redundant streaming scheme is evaluated that helps sustain an uninterrupted video playback in the events of disgraceful peer departures or network errors. The PlanetLab is used for this purpose with a traced peer within the PlanetLab that has controlled network connection (as denoted in Section IV.A.3). The download bit-rate of the traced peer is limited to just below the total content bit-rate. Therefore, the peer is not able to increase its buffer size beyond a certain level. When the peer schedules a chunk from a peer with low upload capacity, the redundant chunk scheduling feature adopted in the proposed P2P engine is triggered.

TABLE III
THE RESULTS OF USING THE ADAPTIVE REDUNDANT STREAMING SCHEME

	Redundancy Enabled	Redundancy Disabled
Average Delivered Layer	2.5	2.41
Number of Changes in Quality	288	322

This feature is specifically evaluated with a test content compressed with SVC using four quality layers (denoted 0, 1, 2 and 3), where each layer has an average bit-rate of 0.5 Mbps. The average download capacity of the traced peer is adjusted to 1.8 Mbps, which falls between the third and the fourth layers. Besides the traced peer in Istanbul, ten other PlanetLab peers have been selected, where five of the selected peers have the upload capacity of maximum two layers (i.e., less than 1 Mbps). These act as the weak peers. Each peer in the swarm has 5 active connections at a time, where each peer replaces the lowest 2 contributors periodically from the 4 non-active connections as part of the deployed incentive policy to eliminate the effect of fixed peer selection. The peer with the highest available upload capacity acts as the content server for the subsequent trial.

Table III shows the average results of ten trials, where in every trial the selected peers in PlanetLab may change depending on their available capacity based upon their usage by other projects. Each trial takes 5 minutes (600 hundred chunks per stream). The first row in Table III represents the average of the Layer IDs (among 0, 1, 2 and 3) that is downloaded by the P2P streaming engine over time. A higher value means a better overall viewing quality. The second row represents the number of layer change events over time. For example, if Layer ID = 3 at time t = n, and if Layer ID becomes 2 at t = n+1, this is counted as a layer change. A smaller value indicates more stable streaming. The results presented in Table III may seem to suggest limited improvement for this case. However, in more challenging network conditions, losing all layers including the base layer poses a greater problem. Thus, the attempt to maintain the received scalable video quality is an important gain in terms of the overall QoE.

B.2 Synchronised streaming with the DVB channel

In this scenario, the performance of the synchronisation methodology is evaluated by measuring how fast the P2P catches synchronisation with the DVB broadcast stereoscopic content. For this purpose, the distributed peers' network is utilised, where the peer in Istanbul is traced for evaluation and the node in Munich acts as the content server. The DVB broadcast is based on the DVB-T technology that uses orthogonal frequency division multiplexing (OFDM). In our implementation, the QPSK modulation with a code rate 3/4 and guard interval 7/8 has been used such that 2 different

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCSVT.2016.2527318, IEEE Transactions on Circuits and Systems for Video Technology

stereoscopic contents could be transmitted simultaneously and comfortably. All nodes remain active (no disconnection) and dedicated for the P2P streaming throughout the trials. The traced peer tunes in to the on-going DVB channel, which enforces it to synchronise its P2P video retrieval with the ongoing broadcast. The average duration of re-synchronization for different network capacities are measured.

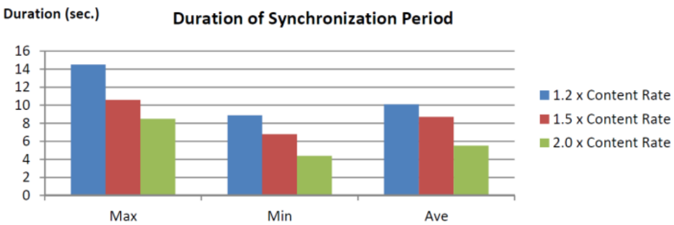

Fig. 10. Duration of re-synchronisation with the DVB channel under different network conditions

The download capacity of the traced peer is adjusted to a value that is higher than the MVV content bit-rate. The time that elapses until the when the P2P streaming engine starts delivering all associated transport streams (i.e., quality enhancement layers and depth maps) in synchronization with the DVB broadcast is measured. Tests are repeated ten times to obtain stable results. The results shown in Fig. 10 indicate that when the available bit-rate is twice the content bit-rate, the re-synchronisation takes 5.5 seconds on average. If the available network capacity is closer to the bit-rate of the MVV content (i.e., the available network capacity is about 1.2 times of the multi-view bit-rate), then re-synchronisation takes place in 10 seconds on average. It is seen that the synchronisation is linked to the network bandwidth capacity. Note that the presented values correspond to the average time elapsed until all associated transport streams of the MVV content start arriving together in synchronicity. This includes the quality enhancement layers. Therefore, the given figures correspond to the average time elapsed until the retrieval of the broadcast content at the highest permitted quality. On the other hand, according to the adaptation rules applied, the P2P streaming engine attempts to start downloading the base quality layers' chunks first. In the conducted tests, the time that elapsed until the full retrieval of base layer transport streams was found to be under a second. Hence, a user can start experiencing the MVV in a very short time after tuning in to the broadcast content.

B.3 P2P link utilisation

In this test scenario, specifically the effect of the double scheduling explained in Section III.C.6 is evaluated. For this purpose, the distributed peers' network is utilised without introducing any regulations regarding the network capacity, since the purpose is not to evaluate the adaptation capability. The four other peers in the swarm are used to increase the availability of chunks for the traced peer in Istanbul. The traced peer enters the session after all the chunks are downloaded by the other peers. Therefore, the only factor that affects the streaming performance is the utilisation of the links between traced peer and the other peers. Tests are conducted under two conditions, one with double scheduling and one without. The evaluation metrics are the download statistics over time, the overall bit-rate of downloading, and the duration of the streaming session. The graph in Fig. 11 presents the obtained results. In this graph, the blue line represents the download rate when double scheduling is enabled, and the red line represents the download rate when double scheduling is disabled. The double scheduling policy grants almost 20% better throughput and therefore the download process finishes earlier. When it is disabled, the peer needs an additional ~100 second interval to download all the chunks.

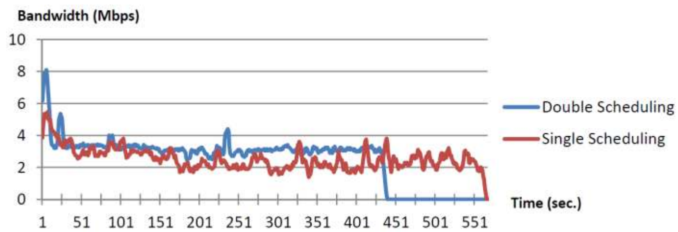

Fig. 11. Download rate of the traced peer with and without double scheduling

It should be noted that most of the existing pull-based P2P video delivery systems use large chunks sizes in order to decrease the overhead of chunk requests. However, using larger size chunks result in coarser adaptation capability. On the other hand, with use of the double scheduling policy, the developed P2P streaming solution delivers finer adaptation capability and also achieves higher throughput rates.

B.4 Server load reduction

In this scenario, the effect of employing variable download window size on the bandwidth load of the MSS is evaluated. For this particular test, relatively low bit-rate content and the PlanetLab platform are used. Using lower bit-rate content allows the peers in PlanetLab to increase their buffer sizes easily and increase their window sizes accordingly. The major contribution of the reduced server load is seen when peers forward chunks directly to other peers without letting them schedule that chunk from the MSS. A scalable test stream with three layers, each at 330 Kbps, is used so that the peers have enough capacity to forward the chunks to each other in PlanetLab.

Two sets of tests are conducted. In the first one, the window size is fixed to 5 seconds. In the second one, the window size is dynamically adjusted based on Equation 3 in Section III.C.4. Tests are conducted with various numbers of peers within the PlanetLab platform (1, 2, 4, 6 and 8 peers). The pre-buffering duration in each test is set to 5 seconds. The bandwidth consumption at the peer acting as the content server (MSS) is measured as the primary performance criterion. Each trial is repeated ten times and the average results are reported.

The average server activity (i.e., bit-rate consumption) with the increasing number of peers is presented in Table IV. It can be seen that even without the utilisation of the adaptive window size, significant server bandwidth is saved against the increasing number of peers. Eight peers consume only 2.16 Mbps in total, instead of 8 Mbps. It can also be observed that the percentage of server bandwidth gain as a result of applying adaptive window size scheme increases with the increasing

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCSVT.2016.2527318, IEEE Transactions on Circuits and Systems for Video Technology

number of peers (12% when number of peers is 8). This fact hints the improved scalability of the proposed delivery system.

TABLE IV
COMPARISON OF ADAPTIVE AND NON-ADAPTIVE WINDOW SIZE FOR THE BANDWIDTH LOAD AT SERVER

	Bitrate Consumption at MSS (Mbps)	
	Adaptive Window Size	Non-Adaptive Window Size
# Peer	Average Bitrate	Average Bitrate
1	0.98	0.98
2	1.10	1.09
4	1.23	1.27
6	1.50	1.61
8	2.16	2.45

C. Subjective evaluation of the proposed system

Subjective tests have also been performed to evaluate the QoE-aware adaptation capability of the deployed hybrid multi-view video delivery system using the distributed peers' network. 24 subjects have been invited to participate in the tests with an age range from 22 to 45. Fig. 12 provides a sketch of the subjective test environment. ITU Recommendation BT.500-11 [42] has been followed in various environmental settings. Peak luminance of the display has been measured as 200 cd/m^2.

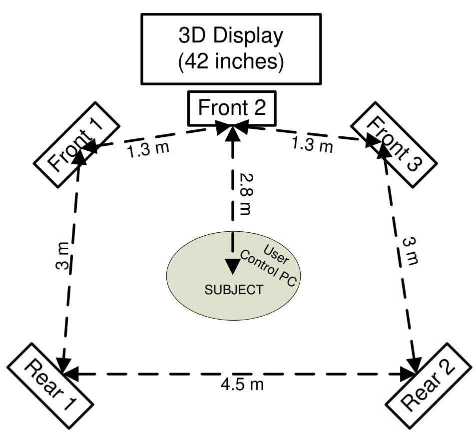

Fig. 12. Subjective test setup with 5 loudspeakers and the 3D passive stereoscopic display (surround audio is broadcast over the DVB network)

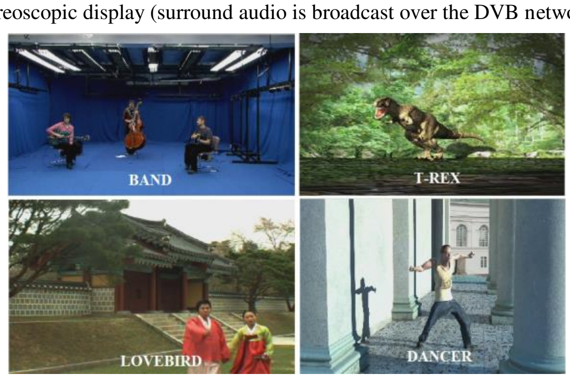

Fig. 13. Sample images from the used MVD contents

The measured ambient illumination has been 180 lux, which is slightly below the recommended value for home environment (200 lux). The viewers' distance from the screen has been set to 2.8 meters that is in compliance with the preferred viewing distance indicated in the recommendation for the according screen height. The user control PC is equipped with a custom Graphical User Interface (GUI), which allows subjects to freely use the "free-viewpoint navigation" (left-right and inwards-outwards directions), "viewing angle" (azimuth and zenith), and "depth adjustment" controls in real-time, while the P2P streaming is activated in addition to DVB broadcast. All four camera views with the associated depth maps have been made available to the peers through the P2P. However, only the peer that is used of the subjective test has been equipped with a DVB receiver. The test peer doesn't retrieve the streams it gets from the DVB network through the IP network to prevent redundancy. All peers in the network remain active throughout the trials without any disconnection. Four different HD-MVD sequences have been used for the test. These are *Band*, *T-Rex*, *Lovebird* and *Dancer* sequences. *Band* sequence comprises three musicians with a simple and static background. *Lovebird* sequence is an outdoor sequence with a static, but fairly complex textured background and slowly moving actors. *T-Rex* sequence is synthetic, with static but complex textured background and a walking dinosaur. *Dancer* sequence is also synthetic, comprising more complex motion (both the imaginary cameras and the actor are moving). The latter two have ground truth depth maps for all four cameras, since they are computer generated. For all contents, the 5.1 surround audio has been broadcast via DVB only. Fig. 13 shows a sample snapshot from all MVD contents. The average bit-rate for the used content is as follows: *Band* – 12 Mbps, *T-rex*– 17.5 Mbps, *Lovebird*– 18.5 Mbps, and *Dancer*– 20 Mbps.

In order to trigger the adaptation process in the test peer, the available bandwidth to the user terminal has been throttled using the software utility called NetLimiter [43]. In each subjective test session and for each test sequence, three different bandwidth throttling patterns have been applied. The patterns are depicted in Fig. 14. The non-shaded intervals correspond to the times with no bandwidth restrictions. In the shaded intervals, the available bandwidth to the user terminal is dropped to about 40% of the total multi-view video bit-rate (i.e., four views plus four depth maps), which forces a majority of the transport streams to be dropped from the downloaded (active) group of streams.

Subjects have been asked to compare the perceptual free-viewpoint viewing quality against a reference adaptation scheme that tends to drop quality enhancement layers regardless of other parameters, such as the user context (instantaneous viewing preferences). The Double Stimulus Continuous Quality Scale (DSCQS) method has been used and the subjects have been allowed to navigate in the scene using the controls throughout each test. The viewing quality after each test is rated on a categorical scale from 1 to 5 (1 for bad quality and 5 for excellent quality). In the analysis stage, the difference in the subjective ratings for the distorted sequence (i.e., the sequence for which the bandwidth is throttled) and the reference sequence (i.e., the sequence delivered always with the highest permitted quality) is calculated. The difference is then scaled into a linear scale ranging from 0 (excellent) to 1

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCSVT.2016.2527318, IEEE Transactions on Circuits and Systems for Video Technology

(bad).

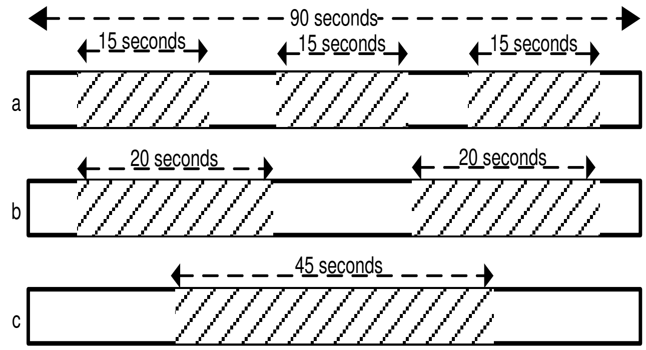

Fig. 14. Applied bandwidth throttling patterns

TABLE V
MOS FOR THE TESTED ADAPTATION SCHEMES

Content	MOS with 95% confidence interval	
	QoE-aware adaptation	Reference adaptation
Band	0.35 +/- 0.10	0.52 +/- 0.10
T-Rex	0.33 +/- 0.13	0.55 +/- 0.07
Lovebird	0.42 +/- 0.09	0.59 +/- 0.11
Dancer	0.38 +/- 0.17	0.67 +/- 0.19

The MOS for each test sequence is obtained by averaging the opinion scores over all subjects. Table V depicts the MOS for the used test multi-view contents. The values depicted in Table V clearly show the positive effect of the proposed bandwidth adaptation scheme for all contents. The relatively large confidence interval is due to the limited number of subjects. Besides, the scores also depend on the amount of user navigation activity performed in the shaded intervals, where the effects of bandwidth throttling on the picture quality are evident. To balance it, tests have been performed using various throttling patterns as shown in Fig. 14.

V. CONCLUSION

This paper has outlined the features of a multi-view video content delivery system operating over the DVB and IP networks, where for the IP delivery part a P2P overlay network is used. The major challenges include the synchronisation of the delivery paths, QoE-aware adaptation against changing network conditions, reduction of the server load for improved system scalability, and stable delivery in the presence of network failures or frequent disgraceful peer departures. The applied novel mechanisms in the P2P delivery network have addressed these challenges and this is verified via performed tests over a number of test beds. A scalable content generation scheme is introduced for the hybrid delivery system. The test results have suggested that it is possible to effectively distribute multi-view video plus depth content over the Internet adaptively, utilising a P2P architecture that limits the load on the content servers and can quickly re-synchronise itself with the on-going DVB broadcast, which carries a sub-set of the multi-view video content.

REFERENCES

[1] Dimenco 3D multi-view display technology, http://www.dimenco.eu/our-technology/3d-image-processing/ (accessed on 28/04/2015).

[2] C. Fehn, "Depth-image-based rendering (DIBR), compression, and transmission for a new approach on 3D-TV", Proc. of the SPIE Stereoscopic Displays and Virtual Reality Systems XI, San Jose, CA, USA, pp. 93–104, Jan. 2004.

[3] Digital Video Broadcasting, DVB Plano-stereoscopic 3DTV Part 2: Frame-Compatible Plano-stereoscopic 3DTV, ETSI TS 101 547-2, 2012.

[4] ITU-T Recommendation H.264: Advanced Video Coding for Generic Audiovisual Services (Apr. 2013).

[5] A. Vetro, T. Wiegand, G. Sullivan, "Overview of the Stereo and Multi-view Video Coding Extensions of the H.264/MPEG-4 AVC Standard," Proc. of the IEEE, vol. 99, no. 4, pp. 626-642, Apr. 2011.

[6] P. Merkle, A. Smolic, K. Muller, T. Wiegand, "Efficient Prediction Structures for Multiview Video Coding", IEEE Transactions on Circuits and Systems for Video Technology, vol.17, no.11, pp.1461-1473, Nov. 2007.

[7] Y. Chen, M. Hannuksela, T. Suzuki, S. Hattori, "Overview of the MVC+D 3D video coding standard", Journal of Visual Communication and Image Representation, vol. 25, no. 4, pp. 679-688, May 2014.

[8] G. Sullivan, J. Ohm, W. Han, T. Wiegand, "Overview of the High Efficiency Video Coding (HEVC) Standard", IEEE Transactions on Circuits and Systems for Video Technology, vol.22, no.12, pp.1649-1668, Dec. 2012.

[9] H. Schwarz, et. al., "Extension of High Efficiency Video Coding (HEVC) for multiview video and depth data", IEEE International Conference on Image Processing (ICIP), vol., no., pp. 205-208, Sep. 2012.

[10] I. Sodagar, "The mpeg-dash standard for multimedia streaming over the Internet", IEEE MultiMedia, vol. 18, no. 4, pp. 62-67, Nov. 2011.

[11] T. Stockhammer, "Dynamic adaptive streaming over http - standards and design principles", Proc. of the second annual ACM conference on Multimedia systems (MMSys '11), pp. 133-144, NY, USA, 2011.

[12] Adobe HTTP Dynamic Streaming (HDS), http://www.adobe.com/devnet/hds.html (accessed on 20/4/2015).

[13] Apple, Using HTTP live streaming, https://developer.apple.com/streaming/ (accessed on 20/4/2015).

[14] Microsoft Smooth-Streaming, http://www.iis.net/downloads/microsoft/smooth-streaming (accessed on 20/4/2015).

[15] P. Aflaki, M. Hannuksela, M. Gabbouj, "Subjective quality assessment of asymmetric stereoscopic 3D video", Signal, Image and Video Processing, vol. 9, no. 2, pp. 331-345, Feb. 2015.

[16] A. Aksay, et. al., "Temporal and spatial scaling for stereoscopic video compression", Proc. 14th Eur. Signal Process. Conf. (EUSIPCO), Italy, Sep. 2006.

[17] H. Schwarz, D. Marpe, T. Wiegand, "Overview of the Scalable Video Coding Extension of the H.264/AVC Standard", IEEE Transactions on Circuits and Systems for Video Technology, vol.17, no.9, pp.1103-1120, Sep. 2007.

[18] G. Saygili, C. G. Gurler, A. M. Tekalp, "Evaluation of Asymmetric Stereo Video Coding and Rate Scaling for Adaptive 3D Video Streaming", IEEE Transactions on Broadcasting, vol.57, no.2, pp. 593-601, Jun. 2011.

[19] E. Ekmekcioglu, V. De Silva, P. T. Pesch, A. Kondoz, "Visual Attention Model Aided Non-Uniform Asymmetric Coding of Stereoscopic Video", IEEE Journal of Selected Topics in Signal Processing, vol.8, no.3, pp. 402-414, Jun. 2014.

[20] J. Chakareski, "Adaptive multiview video streaming: challenges and opportunities", IEEE Communications Magazine, vol.51, no.5, pp. 94-100, May 2013.

[21] E. Kurutepe, M. R. Civanlar, A. M. Tekalp,"Client-driven selective streaming of multiview video for interactive 3DTV", IEEE Transactions on Circuits Syst. Video Technology, vol. 17, no. 11, pp. 1558-1565, Nov. 2007.

[22] G. Cheung, A. Ortega, N. Cheung, "Interactive Streaming of Stored Multiview Video Using Redundant Frame Structures", IEEE

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCSVT.2016.2527318, IEEE Transactions on Circuits and Systems for Video Technology

Transactions on Image Processing, vol.20, no.3, pp.744-761, Mar. 2011.

[23] D. Ren, S.-H.G. Chan, G. Cheung, P. Frossard, "Coding Structure and Replication Optimization for Interactive Multiview Video Streaming," IEEE Transactions on Multimedia, vol.16, no.7, pp. 1874-1887, Nov. 2014.

[24] Z. Liu, G. Cheung, J. Chakareski, Y. Ji, "Multiple Description Coding and Recovery of Free Viewpoint Video for Wireless Multi-Path Streaming", IEEE Journal of Selected Topics in Signal Processing, vol.9, no.1, pp.151-164, Feb. 2015.

[25] B. Macchiavello, C. Dorea, E. M. Hung, G. Cheung, W. Tan, "Loss-Resilient Coding of Texture and Depth for Free-Viewpoint Video Conferencing", IEEE Transactions on Multimedia, vol.16, no.3, pp.711-725, Apr. 2014.

[26] C. G. Gürler, B. Görkemli, G. Saygili, A. M. Tekalp, "Flexible Transport of 3-D Video Over Networks," Proc. of the IEEE, vol.99, no.4, pp. 694-707, Apr. 2011.

[27] N. Magharei, R. Rejaie, Y. Guo, "Mesh or multiple-tree: A comparative study of live P2P streaming approaches", Proc. of IEEE International Conference on Computer Communications (INFOCOM'07), vol., no., pp. 1424-1432, May 2007.

[28] B. Cohen, "Incentives build robustness in BitTorrent", Workshop on Economics of P2P systems, Jun. 2003.

[29] J. Pouwelse, et. al., "Tribler: A social-based peer-to-peer system", Concurrency and Computatation: Practice and Experience, vol. 20, no.2, pp. 127-138, Feb. 2008.

[30] N. Capovilla, et. al., "An Architecture for Distributing Scalable Content over Peer-to-Peer Networks", Proc. of International Conferences on Advances in Multimedia (MMEDIA), vol., no., pp. 1-6, Jun. 2010.

[31] M. Eberhard, et al., "Knapsack problem-based piece picking algorithms for layered content in peer-to-peer networks", Proc. of ACM Workshop on Advanced Video Streaming Techniques for P2P Networks and Social Networking, vol., no., pp. 71-76, Oct. 2010.

[32] C. Ozcinar, E. Ekmekcioglu, A. Kondoz, "Adaptive 3D multi-view video streaming over P2P networks", IEEE International Conference on Image Processing (ICIP), vol., no., pp. 2462-2466, Oct. 2014.

[33] A. Lykourgiotis, et. al., "Hybrid broadcast and broadband networks convergence for immersive TV applications", IEEE Wireless Communications Magazine, vol.21, no.3, pp.62-69, Jun. 2014.

[34] E. Kurutepe, T. Sikora, "Feasibility of Multi-View Video Streaming Over P2P Networks", 3DTV-Conference 2008 , vol., no., pp. 157-160, May 2008.

[35] Z. Chen, L. Sun, S. Yang, "Overcoming view switching dynamic in multi-view video streaming over P2P network", 3DTV-Conference 2010 , vol., no., pp. 1-4, Jun. 2010.

[36] D. Ren, et. al., "Collaborative P2P streaming of interactive live free viewpoint video", ArXiv:1211.4767, vol. abs/1211.4767, Nov. 2012. (http://arxiv.org/pdf/1211.4767v1.pdf)

[37] Xiaojun Hei, Yong Liu, K.W. Ross, "Inferring Network-Wide Quality in P2P Live Streaming Systems", IEEE Journal on Selected Areas in Communications, vol.25, no.9, pp. 1640-1654, Dec. 2007.

[38] Zhengye Liu, et. al., "LayerP2P: Using Layered Video Chunks in P2P Live Streaming", IEEE Transactions on Multimedia, vol. 11, no. 7, pp. 1340-1352, Nov. 2009.

[39] Tribler Protocol Specification, Jan. 2009. (available online: http://svn.tribler.org/bt2-design/proto-spec-unified/trunk/proto-spec-current.pdf)

[40] B. Chun, et. al. "Planetlab: an overlay testbed for broad-coverage services", SIGCOMM Computer Communication Review, vol. 33, no. 3, pp. 3-12, Jul. 2003.

[41] Iperf - The TCP/UDP Bandwidth Measurement Tool, https://iperf.fr/

[42] Methodology for the subjective assessment of the quality of television pictures, Recommendation ITU-R BT.500-11, Jun. 2002. (available https://www.itu.int/rec/R-REC-BT.500-11-200206-S/en)

[43] Internet traffic control and monitoring tool, http://www.netlimiter.com

Erhan Ekmekcioglu received his Ph.D. degree from the University of Surrey, UK, in 2010, and worked there as a post-doc researcher until 2014. Since 2014 he is with Institute for Digital Technologies at Loughborough University London as a research associate. His research interests include 3D video processing and transport techniques. He has participated in a number of collaborative research projects. He is the author/co-author of several research papers, book chapters, and a book on 3D-TV systems.

C. Goktug Gurler received his Ph.D. degree in Electrical and Electronic Engineering from Koc University, Istanbul, in 2013. During his graduate study, he is involved in three European Communication projects, all in different aspects of video delivery over the Internet. He has also published 5 journal and 15 conference papers and co-authored a book chapter about 3D video delivery over P2P networks. After his graduation, he joined Anvato in California as a video encoding engineer. Currently he is leading the department of mobile video streaming technologies at Anvato.

Ahmet Kondoz (M'91–SM'11) received the B.Sc. (Hons.) degree in engineering from the University of Greenwich, UK, in 1983, and the Ph.D. degree in telecommunication from the University of Surrey, UK, in 1987. He became a lecturer in 1988, a Reader in 1995, and then a Professor in Multimedia Communication Systems in 1996, at the University of Surrey. He is the founding Head of the I-Lab Multimedia Communication Research Group. He has been involved with many national and EU projects in the networked media area. He coordinated the EU-FP6 NoE VISNET II, which developed advanced video processing and coding solutions for video surveillance and virtual collaboration applications. He has also coordinated the EU-FP7 DIOMEDES and ROMEO projects, which deal with the distribution of multi-view entertainment using content-aware delivery systems, and remote-collaborative real-time multimedia experience over the future Internet, respectively. Since 2014 he has been leading the Institute for Digital Technologies at Loughborough University London. He has published more than 400 journal and conference papers, authored three books, and holds nine patents.

A. Murat Tekalp received his Ph.D. degree in Electrical, Computer, and Systems Engineering from Rensselaer Polytechnic Institute (RPI), Troy, New York, in 1984. He has been with the Eastman Kodak Company, Rochester, New York, from 1984 to 1987, and with the University of Rochester, Rochester, New York, from July 1987 to June 2005, where he was promoted to Distinguished University Professor. He is currently Professor of Electrical and Computer Engineering at Koc University, Istanbul, Turkey. He served as Dean of Engineering between 2010 and 2013. His research interests are in digital image and video processing, including video compression and streaming, motion compensated filtering, super-resolution, video segmentation and tracking, multi-view and 3D video processing, and video networking. Prof. Tekalp is a Fellow of IEEE. He has been elected a member of Turkish Academy of Sciences and Academia Europaea. He received the TUBITAK Science Award (highest scientific award in Turkey) in 2004. He has been the Editor-in-Chief of the EURASIP journal Signal Processing: Image Communication published by Elsevier between 1999 and 2010. He has served as an Associate Editor for the IEEE Trans. on Signal Processing (1990-1992) and IEEE Trans. on Image Processing (1994-1996). He has been on the editorial board of the IEEE Signal Processing Magazine (2007-2010). He is serving in the Editorial Board of Proceedings of the IEEE since 2014. He has chaired the IEEE Signal Processing Society Technical Committee on Image and Multidimensional Signal Processing (Jan. 1996 - Dec. 1997). He was also a founding member of IEEE Technical Committee on Multimedia Signal Processing in 1997 and served in the committee until 2002. He has been appointed as the Technical Program Co-Chair for IEEE ICASSP 2000 in Istanbul, Turkey, the General Chair of IEEE ICIP at Rochester, NY in 2002, and Technical Program Co-Chair of EUSIPCO 2005 in Antalya, Turkey. He has participated in several EC Framework and COST projects and is serving as a member of ERC Advanced Grants evaluation panel since 2009. Prof. Tekalp holds eight US patents. He has authored the Prentice Hall book Digital Video Processing (1995), the second edition of which was published in 2015.