Noname manuscript No.
(will be inserted by the editor)

Adaptive Delivery of Immersive 3D Multi-View Video over the Internet

Cagri Ozcinar · Erhan Ekmekcioglu · Janko Ćalić · Ahmet Kondoz

Received: date / Accepted: date

Abstract The increase in Internet bandwidth and the developments in 3D video technology have paved the way for the delivery of 3D Multi-View Video (MVV) over the Internet. However, large amounts of data and dynamic network conditions result in frequent network congestion, which may prevent video packets from being delivered on time. As a consequence, the 3D video experience may well be degraded unless content-aware precautionary mechanisms and adaptation methods are deployed. In this work, a novel adaptive MVV streaming method is introduced which addresses the future generation 3D immersive MVV experiences with multi-view displays. When the client experiences network congestion, making it necessary to perform adaptation, the rate-distortion optimum set of views that are pre-determined by the server, are truncated from the delivered MVV streams. In order to maintain high Quality of Experience (QoE) service during the frequent network congestion, the proposed method involves the calculation of low-overhead additional metadata that is delivered to the client. The proposed adaptive 3D MVV streaming solution is tested using the MPEG Dynamic Adaptive Streaming over HTTP (MPEG-DASH) standard. Both extensive objective and subjective evaluations are presented, showing that the proposed method provides significant quality enhancement under the adverse network conditions.

Keywords 3D · multi-view video · multi-view video coding · video coding · video streaming · video adaptation · adaptive streaming · MPEG-DASH

Cagri Ozcinar
LTCI, CNRS, Télécom ParisTech, Université Paris-Saclay, 75013, Paris, France.
Centre for Vision, Speech and Signal Processing, University of Surrey, Guildford, UK.
E-mail: cagri.ozcinar@telecom-paristech.fr E-mail: cagriozcinar@gmail.com

Janco Ćalić
Centre for Vision, Speech and Signal Processing, University of Surrey, Guildford, UK.
E-mail: j.calic@surrey.ac.uk

Erhan Ekmekcioglu · Ahmet Kondoz
Institute for Digital Technologies, Loughborough University London, London, UK.
E-mail: e.ekmekcioglu@lboro.ac.uk, E-mail: a.kondoz@lboro.ac.uk

1 Introduction

Recent advances in video and networked delivery have made it feasible to stream 3D Stereoscopic (3DS) video to homes. In 3D displays, two slightly different views are presented to the eyes and clients can perceive the 3D effect based on the properties of human depth perception [1].

3DS video experience is further enhanced using *Multi-View Video* (MVV) [2, 3] on multi-view displays such as auto-stereoscopic [4]. The MVV technology, simultaneously capturing more than two views of the same scene from different perspectives, allows for motion parallax [5] and a glasses-free immersive 3D experience. Today, a considerable number of multi-view displays are available that require eight-views [6], 16-views [7], or even 28-views [8] as input. However, the Quality of Experience (QoE)[1] associated with such displays is highly dependent on the number of views available at the receiver to render the required virtual viewpoints [9].

Although packet losses are inevitable on the Internet as part of best-effort[2] services, QoE may be degraded significantly, rendering the immersive experience impossible. MVV streaming utilising non-adaptive transmission techniques can cause heavy congestion in the network leading to a congestion collapse [10]. Regardless of the use of the state-of-the-art video coding standards, the delivery of MVV remains a challenging task due to the potentially high number of views required to enable high-quality 3D scene rendering.

Adaptive streaming [11] is the concept of adapting the bandwidth required for the video stream to the throughput available on the network path from the server to the client [12]. Ultimately, adaptive streaming systems provide reliable, high-quality 3D viewing in situations where bandwidths may fluctuate. There have been various commercial developments to support HTTP streaming such as Apple HTTP Live Streaming [13], Microsoft Smooth Streaming [14], and ISO/IEC MPEG Dynamic Adaptive Streaming over HTTP (MPEG-DASH) [15, 16].

Adaptive video streaming using HTTP, which has gained increasing attention nowadays, became a popular alternative to the services that use Real-time Transport Protocol (RTP). Unlike media services such as RTP that uses UDP, HTTP is based on TCP. The use of HTTP/TCP connections allows reusing the existing network infrastructure and overcoming the typical problems of RTP/UDP attributed to firewalls [17]. In TCP, packet re-transmissions are triggered naively to prevent incomplete packet delivery. However, since most multimedia applications are time-sensitive, even through packet re-transmissions are allowed, the packets may be regarded lost unless they arrive within a period.

To investigate the effectiveness of adaptive MVV streaming over the Peer-to-Peer (P2P) network, we proposed an adaptation method based on P2P bandwidth activity in [18], where each P2P client could discard some views based on their buffer occupancy level. Also, we developed an *on-demand* HTTP streaming scenario in [19], where the client receives the *complete* ensemble of encoded MVV.

In this paper, the MPEG-DASH standard [20] and view reconstruction method [19] architecture have been used, and provides the following main contributions:

[1] QoE is addressed from the perspective of a client whose experiencing in a given situation involves a technical application, service or system.

[2] Best-effort describes network services that do not provide any guarantee that the data is delivered.

- Extensive evaluation of the proposed view reconstruction method using different coding standards (HEVC and MVC) and an additional test sequence under various bandwidth conditions suitable for the Internet streaming.
- The results of a formal subjective testing campaign, conducted according to the ITU-T BT.500-13 recommendation for laboratory environments [21].

In this work, our focus is 3D MVV *on-demand* HTTP streaming scenario. To this end, the main goal of this work is to deliver Multi-View video plus Depth (MVD) content to clients in a cost-effective manner by utilising additional metadata. Additional metadata, which is generated by the statistical correlation of the encoded views, comes with a substantially lower overhead compared to the encoded streams. Additional metadata is delivered adaptively *only* when necessary. The proposed adaptive streaming system is completely *independent* from underlying enCOder - DECoder (CODEC). The proposed system also maintains compatibility with the standard decoders allowing MVD content to be decoded and displayed on legacy displays.

The principle of the proposed solution includes deriving multiple adaptation sets with the associated metadata generated at the transmitter side. The receiver makes decisions on when to request and which subset of views needs to be requested. Missing (discarded) views are then recovered at the receiver with the aid of delivered views and the requested additional metadata. In order to test the performance of the proposed adaptive multi-view streaming system, the MPEG-DASH standard was utilised, which allows dynamic adaptive streaming over the Internet using HTTP as its underlying application protocol. Nevertheless, the proposed adaptive 3D multi-view streaming concept can be used with emerging Information-Centric Networking (ICN) delivery system as in [22].

The rest of this paper is organised as follows: Section 2 provides a brief overview of related works. Section 3 presents the overview of the proposed adaptive 3D MVV streaming method. The details of the metadata generation methods at the server, the missing view reconstruction (recovery) method at the receiver, and the proposed HTTP adaptation method are explained in Section 4. Finally, Section 5 presents the experimental results, followed by the concluding remarks in Section 6.

2 Related Works

In this section, the review of the existing works are divided into three categories. The first is state-of-the-art 3D MVV coding techniques followed by the activities in 3D view synthesis using the Depth Image Based Rendering (DIBR) technique. Finally, related works in adaptive 3D video streaming are discussed.

2.1 3D Multi-View Video Coding

Coding of data intensive 3D MVV is challenging. To maximise 3D video coding efficiency, Multi-view Video Coding (MVC) [23] has been standardised as an extension of the H.264/MPEG-4 Advanced Video Coding (AVC) standard [24]. The MVC standard, which takes advantage of different views, uses the inter-view prediction to provide compression (coding) efficient solution. Results show that roughly 50% bitrate saving were observed compared to the AVC standard over a broad range of

objective quality levels [23,25]. Eight views were encoded using AVC – no inter-view prediction coding solution– and MVC standards in [26]. Results present that an average 20% bitrate reduction at the same objective quality is obtained using MVC. However, the increase in visual quality is reported as marginal for some content [27]. The new video coding standard, High Efficient Video Coding (HEVC) [28], provides the same subjective video quality as the AVC standard while requiring approximately 50% less bitrate on average. Wenger *et al.* in [29], used two views, demonstrated that HEVC coding of multiple views (simulcast) results in around a 30% bitrate reduction compared to the MVC extension of AVC standard.

In response to the MPEG Call for Proposals (CfP) on 3D video coding technology [30], several proposal were submitted. For instance, Domański *et al.* in [31] proposed a 3D video coding method that reduces the overall bitrate using disoccluded[3] region coding. The main idea of the proposed coding technique is to exploit view synthesis prediction as much as possible leading to more efficient coding performance. Also, Müller *et al.* in [32] introduced several tools for 3D video coding including new inter-view motion parameter and inter-view residual prediction for the colour texture component of dependent views. Later, the work has been standardised as a 3D extension of HEVC (3D-HEVC). Additionally, the Multi-View extension of HEVC (MV-HEVC) standard [29], uses the same design principle of the MVC standard, utilises neighbouring views at the same time instances for prediction. The prediction is adaptive, hence the optimum predictor among temporal and interview references can be chosen on a block basis [33].

MVV streams may occasionally lose some of the video traffic, which can be related to the rate Variability-Distortion (VD) curve [34]. VD curve illustrates the bit rate variability to the quality level of an encoded video. Video sequences that contain large frame size differences over the time may exceed the network capacity and result in packet losses. For instance, Pulipaka *et al.* in [35] showed that inter-view predicted, *i.e.,* MVC encoded, streams present larger variations in the encoded frame sizes compared to the independently, *i.e.,* no inter-view prediction, encoded video streams. Also, inter-view predicted video streams come at the expense of reduced error robustness due to the massive coding dependencies between the views. Increasing inter-view prediction among views may result in perceivable distortions for the complete set of views if a packet is lost.

2.2 Virtual View Synthesis Using DIBR

MVD representation, which is our focus in this work, provides cost-efficient MVV streaming with the help of MPEG View Synthesis Reference Software (MPEG-VSRS) [30, 36–38]. The MPEG view synthesis method, which is a coordinate conversion process, uses nearest reference views to generate novel views with the help of depth maps and camera parameters. Camera parameters express the relationship between the coordinate system of the camera array and the 3D world coordinates. Depth maps, which are not directly displayed, are only used to provide geometric information of the scene for correct view synthesis.

However, some pixels in the reference views may not be transformed due to an occlusion problem [39], erroneously estimated depth maps [40], incorrectly calibrated

[3] Spatial locations in the virtual view that are not visible from either the left or the right view.

cameras [41], and rounding errors [42][4]. Obviously, those pixels that remain unfilled after the synthesis process may lead to visual distortion. To mitigate view synthesis artefacts, early research works on 3D MVV focused on the depth map pre-/post-processing techniques such as [43]. In the work, Oh *et al.* proposed a novel hole filling method using a depth-based in-painting technique. With the aim of reducing virtual view synthesis artefacts, Cheng *et al.* proposed a spatially and temporally consistent view synthesis method [44]. In their work, an iterative re-weighting framework was introduced by jointly considering texture and depth map temporal consistency. The work does not only achieve temporal consistency, but also reduces the noise disturbance.

MPEG-VSRS can be used to generate a number of different views of the same screen content, especially in limited bandwidth. To achieve cost-efficient MVV streaming performance, Cheung *et al.* proposed a bit allocation method for MVV coding with DIBR [45]. Their goal is to determine how to select the best views for encoding and distribute available bits among them such that to reduce the visual distortion of desired reconstructed views. With the same research line, Sun *et al.* proposed the optimised 3D reconstruction from noise-corrupted multi-view depth videos [46]. In their work, an iterative algorithm was proposed to optimise the scene structure and texture until convergence. Carballeira *et al.* also presented a preliminary study on the Rate Distortion (RD) gain that can be achieved using optimisation techniques at the coding block level [47]. Their work shows substantial bitrate saving as well as an objective quality gain.

Current research activities in view synthesis have achieved important improvements regarding bitrate saving. However, it still causes considerable visual distortions, which remains a challenging task.

2.3 Adaptive Video Streaming

Regardless of the use of 3D MVV coding and view synthesis techniques, robust delivery mechanisms for MVV streaming are vital. Especially, streamed packets may be dropped at a router/switch due to congestion or delayed reception. On the one hand retransmission of lost packets is one of the basic mechanisms to mitigate network failures [48], but, on the other hand, this mechanism is not a feasible solution for delay-sensitive applications due to playback constraints. Frequent retransmissions may lead to severe network congestion, which is undesirable for the Internet streaming.

Advanced 3D video streaming systems may well need to employ intelligent strategies to minimise the effect of inevitable network-related problems. For instance, Thang *et al.* investigated adaptation methods in the context of live video streaming [49]. In their work, experimental evaluations were carried out regarding bitrate as well as the perceptual quality impact. Also, Gürler *et al.* proposed an adaptive stereoscopic 3D video streaming system [50], where bitrate of the stereo view was adapted using Signal-to-Noise Ratio (SNR) scalability option of Scalable Video Coding (SVC). With the same aim and different network application, Savas *et al.* in [51] proposed an adaptive delivery method for MVV over P2P networks. In their work,

[4] Rounding of the pixel position to the closest integer may introduce an incorrect position in the result.

a P2P streaming solution based on BitTorrent P2P protocol was extended with an adaptive windowing mechanism [52]. However, the P2P streaming may lead problems for Internet service providers by consuming significant amounts of bandwidth. Besides, clients need to install additional applications [53] to use P2P streaming protocols.

From the end-user perspective, using a Web browser is more convenient for streaming videos directly over the Internet. For instance, a segmented video streaming was proposed in [22] for small wireless devices and sensors. In the work, each video segment is an independently playable file and does not require any plug-in or back-end engines at the receiver. Also, Oztas *et al.* in [54] introduced a rate adaptation method for the MPEG-DASH standard for streaming MVD by exploring the effects of the number of transmitted views and the quality of views. Later, symmetrically/asymmetrically quality scaling was investigated in [55] to minimise expected network distortions such as network congestion. For this aim, a set of views were delivered and then the undelivered views were reconstructed using the MPEG-VSRS at the receiver. In the work, subjective tests were conducted to examine the best rate adaptation strategy in terms of QoE. The results indicated that transmitted intermediate views[5] should be compressed as much as possible. When the average Peak-Signal-to-Noise-Ratio (PSNR) value for the intermediate views is reached the selected threshold, only the edge views (*i.e.,* the side-most views) should be delivered. Undelivered views are then recovered using the view synthesis method. However, the visual quality performance of the MPEG-VSRS is influenced by the DIBR based view synthesis artefacts outlined in Section 2.2.

Recently, there has been an ongoing research on defining and developing the future of Internet architectures, such as Content-Centric Networking (CCN) [56]. CCN, which is an Information-Centric Networking (ICN) project, is uses a content identifier rather than the classic host identifier. As a result, video sequences can be stored anywhere in the system that can achieve an efficient content distribution. Nowadays, adaptive streaming solutions are emerging for CCN. For instance, Liu *et al.* proposed dynamic adaptive streaming over CCN [57], which shows that the MPEG-DASH standard can be adapted relatively easily to a CCN environment taking advantage of the caching features offered by CCN. Also, Detti *et al.* developed an ICN P2P application for adaptive live video streaming [58]. In their work, enhanced video streaming performance was observed using the combination of the CCN architecture and the MPEG-DASH standard.

However, the current MPEG-DASH standard does not cover the design of the MVV adaptation strategy. To this end, in this paper, an enhanced adaptive MVV streaming technique, which considers the use of MPEG-DASH over HTTP 1.1, is proposed with the adaptation logic performed on the receiver side.

3 System Overview

Fig. 1 shows a schematic diagram of the proposed delivery system. The captured MVD content is divided into equal-length temporal segments, encoded by a standard

[5] In this paper, the term of intermediate views is used as the captured views that are available at the location between the leftmost and rightmost (*see* $View\ N-k$ and $View\ N+k$ in Fig. 2) in MVV content.

encoder[6] at various bitrates, and stored as self-decodable single-layer streams in the HTTP server. The MVV streams exist in the server in two parts: Media Presentation Description (MPD), which is an XML document[7] that contains the manifest of the adaptation strategy that is described in Section 4.3 and MVD segments that contain the MVD streams in the form of video segments. Each segment contains the size of a Group Of Picture (GOP), which is typically around 0.5 second.

The client retrieves MVD segments and, as necessary, the content is adapted to the dynamic network conditions by discarding the pre-determined view(s). View discarding order, which is computed by the MVV encoder, is transmitted with the `MPD file` to the client *each* five-second period. Network adaptation is performed by the client side, which relies on the adaptation rule transmitted by the encoder side, and then retrieves the corresponding Side Information (SI) for MVV reconstruction.

Discarded views are reconstructed by incorporating the low-overhead SI, which is requested by the client at times of adaptation. The SI stream contains the codebook[8] index values of the delivered views' weighting factors, which are calculated using cross-correlation method [59, 60]. The SI, is a raw bit stream (*i.e.,* no compression algorithm is applied to generate it), is delivered to the client using the separate transmission path, as illustrated in Fig. 1. Furthermore, the codebook[9], which is represented as the number of bits, is created as a result of weighting factor estimation at the encoder, and is downloaded by the client during the *start-up* buffering period (*i.e., only* transferred at the beginning of a streaming session).

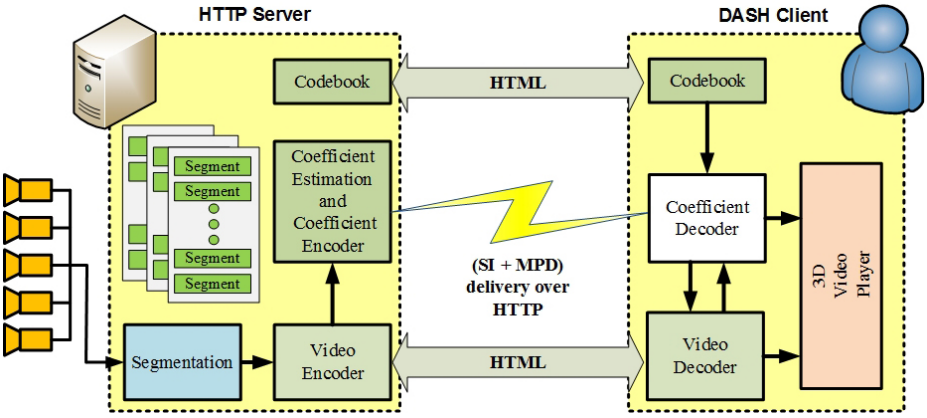


Fig. 1: Overview of the proposed delivery system [19].

After downloading the codebook, to play the MVV content, the DASH client should obtain the pre-estimated `MPD file` through HTTP and fetch the appropriate

[6] Note that the proposed adaptation method is completely independent from the underling CODEC.

[7] Extensible Markup Language (XML) is a metalanguage that defines a set of rules for encoding documents in a format that is both human-readable and machine-readable.

[8] Codebook and SI examples can be found in `https://drive.google.com/open?id=0B_IXpfe4UW-BWFptRTQyZHdheXc`.

[9] The typical 5-bit codebook size is 1.5 KBytes, which is sent once at the start of the sequence.

subset of encoded views. For each view that can potentially be discarded as a result of the instantaneously available network capacity, the DASH client continuously requests the segments from the HTTP server and monitors the network throughput. Depending on the network throughput and discarding order in the `MPD file`, the DASH client decides on whether to adapt by requesting a subset of views or a complete set of views.

The information that is needed to reconstruct the discarded/missing views as a result of adaptation is estimated according to the metadata estimation process with the help of the DIBR technique. The estimated metadata stream is then delivered as the SI stream, the details of which are explained in Section 4.1. The SI is utilised to reconstruct discarded views with high quality at the receiver side, which is delivered to the DASH client using the `HTTP GET` method [61].

4 Adaptive View Recovery for 3D MVV Streaming

In this section, the key components of the proposed adaptive view reconstruction (recovery) model for 3D MVV streaming are described as follows: Section 4.1 presents the metadata (*i.e.,* SI) estimation process. The details of the quadtree-based adaptive block-size selection procedure are explained in Section 4.2. Finally, Section 4.3 presents the dynamic view adaptation strategy.

4.1 Metadata Estimation

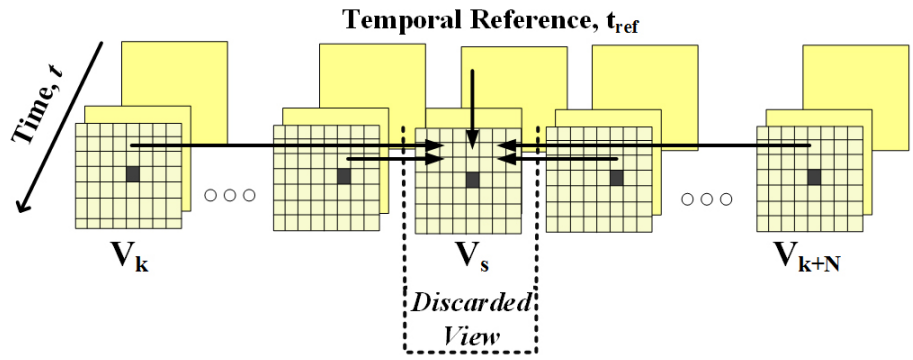


Fig. 2: Cross-correlation method for N number of available views. In the process of recovering of discarded views, its neighbouring delivered views from both directions are utilised.

To calculate the additional metadata for recovering frames within the discarded view(s), the co-located frame from the delivered views, V_{del}, and a previous temporal reference, t_{ref}, within the same discarded view(s) are utilised, as shown in Fig. 2. V_k and V_{k+N} are the most side views of the complete ensemble of MVV content. Hence, $V_{all} = \{ V_k , \ldots , V_{k+N} \}$ where V_{all} is available view domains. The cross-correlation

and DIBR techniques are utilised in the server in such a way that the depth-aided image interpolation quality is superior to that of the encoded view. For each view assumed to be temporarily discarded from streaming, all corresponding blocks are replenished. In order to do so, the weighted sum of all projected corresponding blocks from delivered views is computed as:

$$\widehat{B}_s(x,y,t) = \sum_{c \in V_{del}} [\widetilde{B}_c(x,y,t) \cdot w_c] + \sum_{c \notin V_{del}} [B_c(x,y,t_{ref}) \cdot w_c] \tag{1}$$

where $\widehat{B}_s(x,y,t)$ represents the reconstructed pixel at (x,y,t). x and y are the horizontal and vertical coordinates of the pixel. t is the current frame time. V_{del} is delivered view domains. $\widetilde{B}_c(x,y,t)$ represents the DIBR projected block of the c^{th} view, and $B_c(x,y,t_{ref})$ is the temporal reference of the target view block, which is the last frame's block in the corresponding previous temporal segment. t_{ref} is the temporal reference time, and w_c represents the weighting factors of each block for each view, which corresponds to the SI.

This model, as shown in Fig. 2, recovers the discarded views, V_s where $s \neq \{k, k+N\}$, with the smallest possible pixel error in relation to its uncompressed original representation. In order to estimate discarded view(s), the sum of squared errors, e_s^2, between the reconstructed pixel values, $\widehat{B}_s(x,y,t)$, and original pixel values, $B_s(x,y,t)$, is calculated as shown in Equation (2):

$$e_s^2 = \sum_{x=1}^{X} \sum_{y=1}^{Y} \left[\widehat{B}_s(x,y,t) - B_s(x,y,t) \right]^2 \tag{2}$$

where X and Y represent the width and height of the current block, respectively. In order to minimise e_s^2, it is necessary that the derivative of e_s^2 with respect to each of the weighting factors w is equal to zero, *i.e.,*

$$E\left[2 \cdot e_s \frac{de_s}{dw}\right] = 0 \qquad for \quad every \quad w \quad then, \tag{3}$$

$$\frac{de_s}{dw_s} = \widetilde{B}_s \quad , \quad s \in V_{del} \quad thus, \tag{4}$$

$$2 \cdot E\left[e_s \frac{de_s}{dw}\right] = 2 \cdot E\left[e_s \widetilde{B}_s\right] = 0 \qquad s \in V_{del} \tag{5}$$

Neglecting the constant numbers,

$$\begin{gathered} E\left[\left(\widetilde{B}_k \cdot w_k + \ldots + \widetilde{B}_{k+N} \cdot w_{k+N}\right) \cdot \widetilde{B}_k\right] = 0 \\ \vdots \\ E\left[\left(\widetilde{B}_k \cdot w_k + \ldots + \widetilde{B}_{k+N} \cdot w_{k+N}\right) \cdot \widetilde{B}_{k+N}\right] = 0 \end{gathered} \tag{6}$$

Hence,

$$\begin{bmatrix} E\left[\widetilde{B}_k \cdot \widetilde{B}_k\right] & \cdots & E\left[\widetilde{B}_k \cdot \widetilde{B}_{k+N}\right] \\ \vdots & \vdots & \vdots \\ E\left[\widetilde{B}_{k+N} \cdot \widetilde{B}_k\right] & \cdots & E\left[\widetilde{B}_{k+N} \cdot \widetilde{B}_{k+N}\right] \end{bmatrix} \cdot \begin{bmatrix} w_k \\ \vdots \\ w_{k+N} \end{bmatrix} = \begin{bmatrix} E\left[\widetilde{B}_k \cdot \widetilde{B}_s\right] \\ \vdots \\ E\left[\widetilde{B}_{k+N} \cdot \widetilde{B}_s\right] \end{bmatrix} \tag{7}$$

where $E[\cdot]$ represents the normalised expected value [60].

After estimating the weighting factors per block for all available views, in order to design the codebook, the k-means clustering algorithm [62] is applied to the estimated weighing coefficients. Each weighting factors (w) forms coefficient vectors in the codebook, which is encoded using an l-bit codeword, as described in Equation (8).

$$W_i = \left[w_k \ \ldots \ w_s \ \ldots \ w_{k+N} \right] \tag{8}$$

where coefficient vector W is chosen from a finite set of coefficient vectors in the codebook with size L, and $L = 2^l$. Also, the index number is denoted as i ($1 \leq i \leq L$).

The codebook is downloaded from the HTTP server by the DASH client at the beginning of streaming. The index numbers of each computed coefficient vector corresponding to each computed block are embedded in the SI stream in the HTTP server. The DASH client parses the codebook index values embedded in the SI stream to recover the corresponding coefficient vectors from the codebook. For the recovery of discarded view(s), correctly received and decoded frames of neighbouring views' with the corresponding temporal reference frame are utilised. The reconstruction is performed as per the weighted summation in Equation (1). This reconstruction is applied for each discarded segment of the view, which is the size of a Group Of Picture (GOP).

4.2 Quadtree Based Adaptive Block-Size Selection

Quadtree coding [63, 64], which has been widely used for block partitioning in video processing to take advantage of variable block-size, is employed in the proposed view reconstruction mechanism. This technique segments the regions in such a way that they can be reconstructed at high quality with low overhead. The block partitioning process is entirely independent from the partition used in the underlying CODEC.

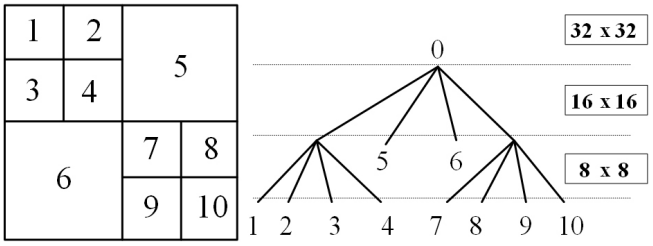


Fig. 3: A variable block-size and the corresponding nested quadtree structure.

Fig. 3 illustrates the exemplary quadtree structure. As can be seen, `8×8` and `32×32` are chosen as the smallest and largest block-sizes, respectively. However, smaller block-sizes, used for regions that can be reconstructed at low quality, increase SI overhead. For this purpose, the trade-off between the discarded views' reconstruction quality and the resulting SI overhead size are optimised using the

Lagrangian optimisation method [65]. At that point, optimum block-sizes for each frame of discarded views are calculated.

The block partitioning method evaluates different block sizes adaptively and assigns an optimum block-size for each region in the frame. Each region is divided into four equal size blocks starting from the largest block-size (*i.e.,* 32×32 block) in a top-down approach [66].

In the block-size optimisation, the overall block distortion, D, is minimised subject to a limited overall SI rate-budget B_{max}. The value of B_{max} was calculated experimentally through subjective training using four different MVV contents (*Bookarrival*, *Newspaper*, *Café*, and *Pantomime*). In this method, the cost of each possible block-size is calculated by Equation (9), and the smallest value is chosen as the optimal for each block.

$$\underset{x}{\operatorname{argmin}}\, J(b) = D(s) + \lambda \cdot B(s), B(s) < B_{max} \tag{9}$$

where J is expressed as the cost value, and b represents each block number in the quadtree structure (*see* Fig. 3). λ is the Lagrangian multiplier, and $B(s)$ is the cost of transmitting the additional metadata. Every partitioning is represented by the corresponding codevector s, which is assigned a variable length codeword from a given quadtree structure. l denotes the code length; when the cost of transmitting the quadtree structure is included, the overall cost function becomes:

$$J(x) = \sum_{x=1}^{P} \sum_{y=1}^{P} D(x,y) + \lambda_1 \cdot l + \lambda_2 \cdot Q \tag{10}$$

where P represents the block size (such as 32, 16, 8) in the view, λ_1 and λ_2 correspond to the Lagrangian multiplier values for the SI and quadtree structure overhead, respectively. λ_1 and λ_2, which are obtained experimentally through subjective training. Four different MVV contents, *Bookarrival*, *Newspaper*, *Café*, and *Pantomime*, were utilised to estimate the optimum λ_1 and λ_2. Q denotes as the number of bits (quadtree code-length) required for the signalling of the quadtree structure. l is the number of bits necessary for the SI, which is explained in Section 4.1.

In this work, the Mean Square Error (MSE) based distortion metric has been utilised in $D(x,y)$. However, other perceptual quality metrics (*e.g.,* Synthesized View Distortion Change (SVDC) [67], Structural SIMilarity (SSIM) [68], Spatial Peak Signal to Perceptual-Noise Ratio (SPSPNR) [69]) are equally compatible with the proposed system.

4.3 View Adaptation Using HTTP

In order to cope with the varying network conditions during streaming, two possible adaptation strategies can be employed in the outlined system model. Firstly, it is possible to reduce the visual quality (*i.e.,* increasing quantisation level) of all available views, V_{all}, to match the Internet bandwidth. However, this method would lead to a reduction in the reconstruction quality of all received views and depth maps, thus resulting in a decrease in the perceptual quality. The second option is to transmit selective view streams through content-aware bandwidth adaptation and to

allow the missing (discarded) view(s) to be reconstructed at the receiver using the delivered views. This strategy may not end up in compromising the delivery quality of some views, as demonstrated in [51,55], unlike the first strategy.

The proposed system, which bases on the second described strategy, reduces the transmitted number of views during the network congestion period. The proposed system aims at enabling the receiver to reconstruct all required views at the highest possible quality at all times by incorporating the low-overhead SI. In contrast to the first explained strategy (*i.e.,* quality reduction of all available views), this method cannot result in compromising the perceptual quality of all encoded views. However, it leads to maintain the highest possible quality at the receiver side, giving the extraction and usage of optimised SI along with the delivered high-quality views and depth maps.

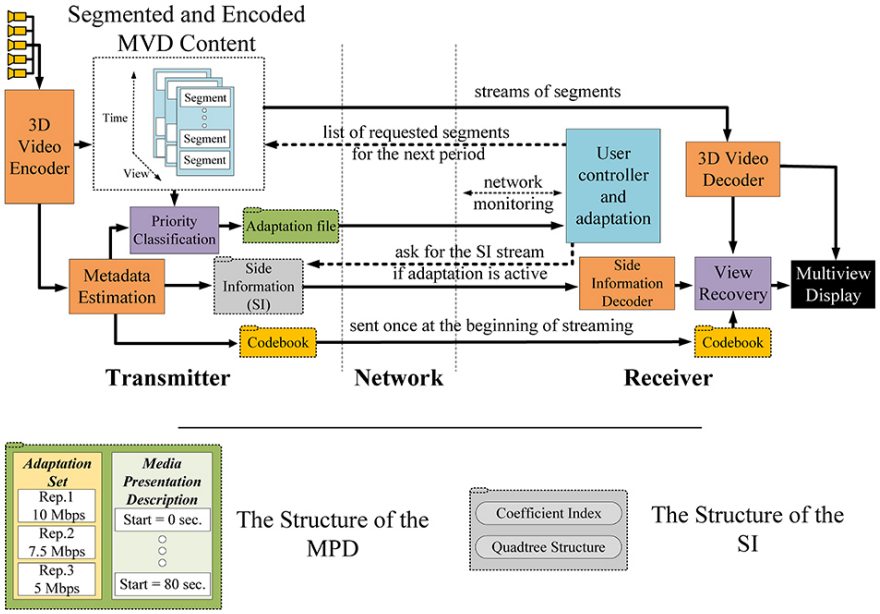


Fig. 4: The proposed adaptive MVV delivery scheme [19]. The SI stream is delivered as the overhead, which contains weighting factors (*Section* 4.1), and quadtree structures (*see* Section 4.2).

Fig. 4 shows the proposed adaptive MVV delivery scheme. All adaptation parameters are prepared according to the possible network bandwidth conditions, *i.e.,* r_{net}, at the server, and they are inserted to the `MPD file`.

`MPD file`, which is prepared at the server, is downloaded by the client before the streaming starts and updates are delivered when necessary (*e.g.,* scene changing). MPD is prepared according to the possible network bandwidth conditions, *i.e.,* r_{net}, at the server. All adaptation parameters are inserted into the `MPD file`. According to the client bandwidth, the receiver fetches the `MPD file` and request optimum video segments.

To prepare the `MPD file`, the number of views is discarded and evaluated at the server. Consequently, optimal discardable view(s) are determined that minimise the overall MVV distortion to meet network bandwidth (r_{net}). This process can be formulated as an optimisation function, $f(\cdot)$, which is defined as in Equation (11):

$$f(Dist_k, \ldots, Dist_{k+N}) \quad and \quad satisfy \quad (R_V + R_{SI}) < r_{net} \tag{11}$$

where, $Dist_k$, ..., $Dist_{k+N}$ are objective distortion of views. R_{SI} is denoted as the additional metadata overhead for the proposed adaptation and the total encoded MVV content bitrate is R_V. The total bitrate requirement of the proposed adaptation, $R_V + R_{SI}$ and the overall distortion, $Dist_k + \ldots + Dist_{k+N}$, are considered for each segment, which is typically around 0.5 second.

Delivered views are determined using the priority information, which is the process to minimise the overall MVV distortion subject to the r_{net}. The priority estimation is described in Equation (12):

$$p_s = \frac{\omega_s}{\sum_{j=k}^{k+N} \omega_j} \tag{12}$$

where, p_s is denoted as the view prioritisation, s belongs to a discrete set of views between (and including) views k and $k+N$. Also, ω_j is expressed as the priority weight of each view.

The priority weight of each view is determined through classification with an aim of minimising the overall reconstruction distortion subject to the bitrate budget (r_{net}). The overall cost minimisation function is described in Equation (13).

$$\underset{k}{\operatorname{argmin}}\, P(k) = D_s(k) + \lambda \cdot R(k), \quad R(k) < R_a \tag{13}$$

where $R(k)$ is the overall transmitted bitrate after discarding some of the views (including the bitrate of the SI, R_{SI}), $D_s(k)$ is the average reconstruction distortion of all discarded view(s) estimated using MSE. Also, λ is the Lagrangian multiplier, which is set through subjective experiments using four different MVV contents (*Bookarrival*, *Newspaper*, *Café*, and *Pantomime*).

Each view is encoded using the similar coding parameters (*e.g.,* similar QPs) and stored in the HTTP server. At this point, the client can manage the streaming session based on the `MPD file`, which contains adaptation information. Hence, the predetermined subset of views can be requested using the `HTTP GET` method when the available network bandwidth cannot accommodate the transmission of the complete collection of encoded MVV. Depending on the measured bandwidth, some views are effectively discarded to be recovered using delivered views with the SI at the receiver. In the most severe condition, to be able to recover all views within the total baseline of the MVV set, the edge views (*i.e.,* side-most views) and their associated depth information need to be delivered.

5 Experiment Results

In this section, experiment results are presented to demonstrate the enhancement effects of the proposed adaptation mechanism for MVV streaming. Thus, both objective and subjective results are depicted to highlight the positive impact of the proposed approach over the congested network.

5.1 Experiment Setup

Four different MPEG video test sequences were considered for experiments. These test sequences are also selected as test material for the MPEG 3D video coding standardisation activities. Our experiments were conducted using five adjacent colour texture views and depth maps (*i.e., M=5*) from selected different MVV test contents, which are *Bookarrival* [70,71], *Newspaper* [72,73], *Café* [73,74], and *Pantomime* [75]. The properties of these contents are summarised in Tab. 1.

Fig. 5 shows the Spatial-Information (S-I) and Temporal-Information (T-I) indexes on the luminance component of each content, as described in the ITU-R P.910 [76]. Fig. 5a presents the S-I and T-I indexes for colour texture view. From the figure, it can be inferred that *Bookarrival*, *Newspaper*, and *Pantomime* have large S-I values for texture view, whereas *Café* has a small value, *i.e.,* low spatial details. Also, the *Bookarrival*, *Café*, and *Pantomime* sequences contains higher T-I value for both texture and depth maps, which indicates high level temporal of detail.

Table 1: MVV Test Sequences Specifications.

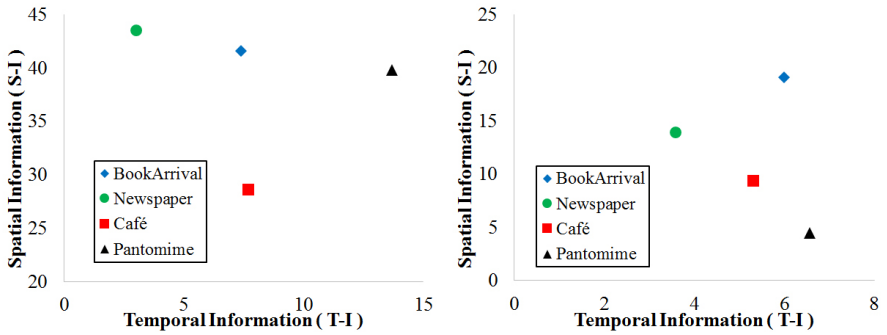
Sequence	*Camera Spacing (cm)*	*Resolution*	*Selected Views*
Bookarrival	6.5	1024×768	6,7,8,9,10
Newspaper	5	1024×768	2,3,4,5,6
Café	6.5	1920×1080	1,2,3,4,5
Pantomime	5	1280×960	37,38,39,40,41

HEVC and MVC standards were utilised to analyse the performance of the proposed adaptation system. Therefore, `HM v10.1` and `JMVM v8.1` were used to encode each MVV content. Various QPs were selected, *i.e.,* 20, 26, 32, 38, and 44, for each colour texture sequence. Also, the depth maps bitrates were chosen to be equal to the percentage of 20% of the colour texture bitrates, as suggested in [77]. Hence, appropriate compressed depth maps were chosen by trials. Hierarchical B pictures were used with a GOP length of 16, and a single GOP was inserted into each transmission segment. In the case of MVC, the inter-view prediction structure was determined based on the view discarding pattern.

To further evaluate the performance of the proposed approach, `MPEG-DASH` [78] was incorporated in the sever-client setup [79]. Three regular PCs were used in this setup. A PC was used as a transmitter (*i.e.,* streaming server), one regular PC as a receiver (*i.e.,* streaming client), and one regular PC as a router based on the `Dummynet` tool [80] to emulate network environment.

To evaluate the performance of the proposed approach, MPEG View Synthesis Reference Software `VSRS v3.5` [36] was incorporated as an adaptation reference. In MPEG-VSRS, the two nearest left and right adjacent views are utilised in this reference to synthesise discarded view(s). MPEG-VSRS was used as the base to estimate the views that are not delivered along with the received SI and codebooks of various sizes. The additional overhead from the proposed method (*i.e.,* SI) is *included* in all reported results.

Experiment results were compared using the Bjøntegaard Delta (BD) method [81], which describes the distance between two RD curves. In this manner, PSNR difference, namely ΔP, in deciBel (dB) averaged over the whole range of bitrates,



(a) For texture (b) For depth map

Fig. 5: Spatial-Information (S-I) versus Temporal-Information (T-I) indexes of the selected MVV test sequences.

and bitrate difference, namely ΔR, in percentage averaged over the entire range of PSNR, were identified.

Furthermore, subjective tests were performed [82] in accordance with the ITU-R BT.500-13 [21]. In total, 18 non-expert observers (12 males and six females), aged between 26 and 45, participated in the test. The observers watched the prepared test sequences at a distance of 3.75 m away from the display, which is approximately 5.5 times picture height. Each test session started after a short training and instruction session. The observers were introduced to the test environment, grading scale, and were presented with an example of the training sequences. Each assessment session lasted up to half an hour.

The Double Stimulus Continuous Quality Scale (DSCQS) was utilised using a scale that ranged from 0 (Bad) to 100 (Excellent). Observers were provided the freedom to view video pairs, *i.e.,* original reference and processed, over and over again as they wish before reaching a verdict. It was observed that the test repetition rate was below 5%.

The collected Mean Opinion Scores (MOS) were then analysed for each subject across different test conditions, which verified MOS distribution consistency. The outlier detection procedure was applied outlined in ITU-R BT.500-13 [21], there were no outliers in the reported scores.

Finally, the Different Mean Opinion Score (DMOS), which is obtained by subtracting the MOS of the processed (reconstructed/synthesised) sequence from that of the original one. DMOS measurements can show how much differences introduced in test videos degrade subjective picture quality. A higher DMOS indicates bigger quality degradation of the processed sequence.

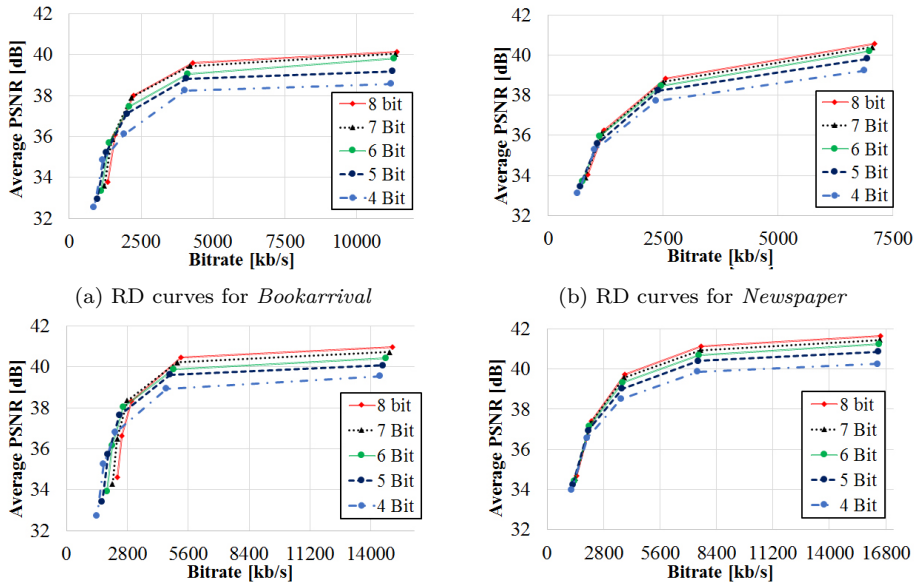
5.2 Results and Discussion

In this section, the experiment results are given along with a discussion. Also, both objective and subjective results are demonstrated. In the Subsections 5.2.1 -5.2.3, only one intermediate view was discarded at a time, and analysed. The results are

presented in a comparative manner, which clearly shows the proposed approach performance compared with the references.

5.2.1 Effect of Codebook Size Used on Reconstructed View

The view reconstruction (recovery) performance of the proposed approach using different codebook sizes is demonstrated using RD curves. Fig. 6 shows the resulting average view reconstruction performance using different codebook sizes for *BookArrival*, *Newspaper*, *Café*, *Pantomime* respectively. In this experiment, only the same view was discarded for each tested codebook at a time, and the average view estimation quality in terms of PSNR was calculated from the discarded views. The estimated image quality was calculated with respect to the uncompressed original view. The bitrate includes the transmitted MVV content and the overhead caused by transferring the additional SI. Three different codebooks created for each discarded view, which contains a varying number of coefficient vectors (W), are described by the number of bits (l-bit). Each view was assumed to be discarded, and all corresponding blocks were estimated using different codebooks. The objective reconstruction quality for each view was compared to others in order to determine the optimum views to be discarded for each period.



(a) RD curves for *Bookarrival*

(b) RD curves for *Newspaper*

(c) RD curves for *Café*

(d) RD curves for *Pantomime*

Fig. 6: Missing view reconstruction performance using different codebook sizes.

As can be seen in the results presented in Fig. 6, the performance of the proposed approach tends to saturate each of the content pieces when the size of the used codebook increases. However, a relatively large performance gap is obtained between the five-bit and four-bit codebooks. Accordingly, a five-bit codebook achieves view

reconstruction performance closer to that of the largest codebook sizes, and benefits from the advantage of lower SI overhead. To this end, five-bit codebook was used in the remaining experiments.

5.2.2 Effect of Block-Size Used on Estimated View

Fig. 7 illustrates the performance comparison of the proposed reconstruction method using different block-sizes. The aim of this analysis is to demonstrate the effectiveness of the variable block-size selection process (*see* Section 4.2).

In this experiment, 32×32, 16×16, and 8×8 fixed block-sizes were compared to the quadtree-based variable block-size. As can be seen in the figure, for the *BookArrival* and *Newspaper*, (1024 × 768) sequences, 32×32, 16×16, and 8×8 require 153.6, 614.4, and 2457.6 *kbps*, respectively. The high-resolution (1920 × 1080) *Café* sequence requires 405, 1620, and 6480 *kbps* in order to reconstruct a discarding view for 32×32, 16×16, and 8×8, respectively. Furthermore, the *Pantomime* sequence, which is 1280 × 960 resolution, needs 245, 947, and 3845 *kbps* to reconstruct a discarding view for 32×32, 16×16, and 8×8, respectively.

As can be seen in Fig. 7, the resulting overhead from the quadtree-based variable block-size demonstrates an increasing bitrate performance, whereas video distortion increases. The reason for this is that increasing video distortions present poor reconstruction performance, and thus the proposed approach requires higher amount of metadata (*i.e.,* SI) to construct a missing view with high quality. Moreover, it is observed that corrupted depth maps affect the required metadata overhead size. For instance, the *Café* sequence, which contains inaccurate depth maps compared with others, requires high amount of metadata.

Furthermore, the proposed view reconstruction quality performance in Figures 7 (b), (d), and (f) support the analysis of view estimation overhead in Figures 7 (a), (c), and (e). As the block-size increases, the objective quality is also enhanced. However, the increase in the overhead of the view estimation metadata, *i.e.,* SI, reduces the coding performance in the RD curves. In the proposed view reconstruction approach, each block is assigned a weighting coefficient and increased SI overhead decreases the degree of coding efficiency. For this reason, a quadtree-based variable block-size selection mechanism employs cost-quality optimisation in an exchange between SI overhead (cost) and the view reconstruction quality.

Experiment results for three MVV sequences indicate that the quadtree-based variable block-size shows optimum performance compared with the fixed block sizes (*e.g.,* 32×32, 16×16, and 8×8). This occurs primarily because of the variable block-size selection process that exploits the quadtree-structure and reduces SI overhead.

5.2.3 View Reconstruction Performance

In order to further investigate and evaluate the performance of the proposed view reconstruction, another view estimation method is incorporated as an additional reference: MPEG-VSRS without using SI. In this reference, the nearest left and right neighbouring views are projected to the target view's position, one from the nearest left and another from the nearest right. The two projected images are then blended to form a synthesised image. The pixel values in the synthesised image are created by blending the respective pixel values in the projected images with unequal weights, where the weights are inversely proportional to the distance from the target

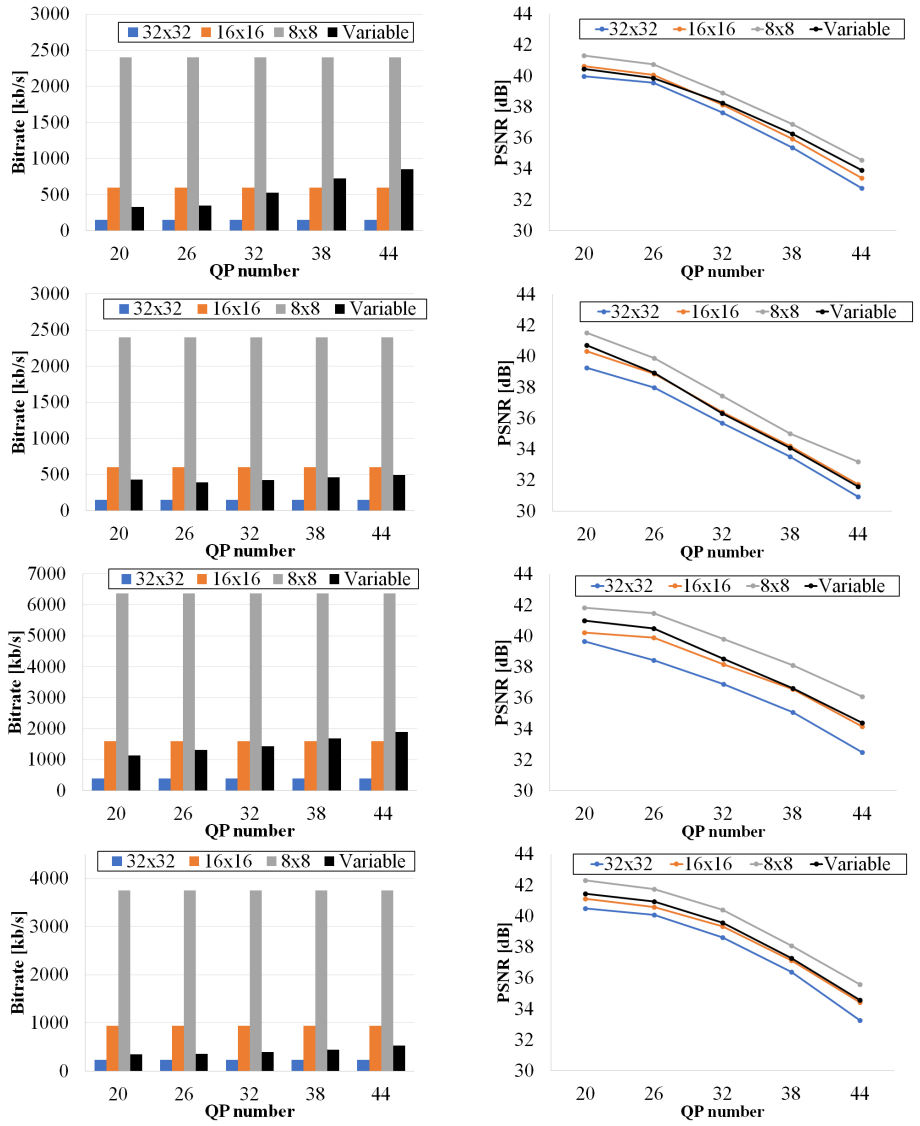
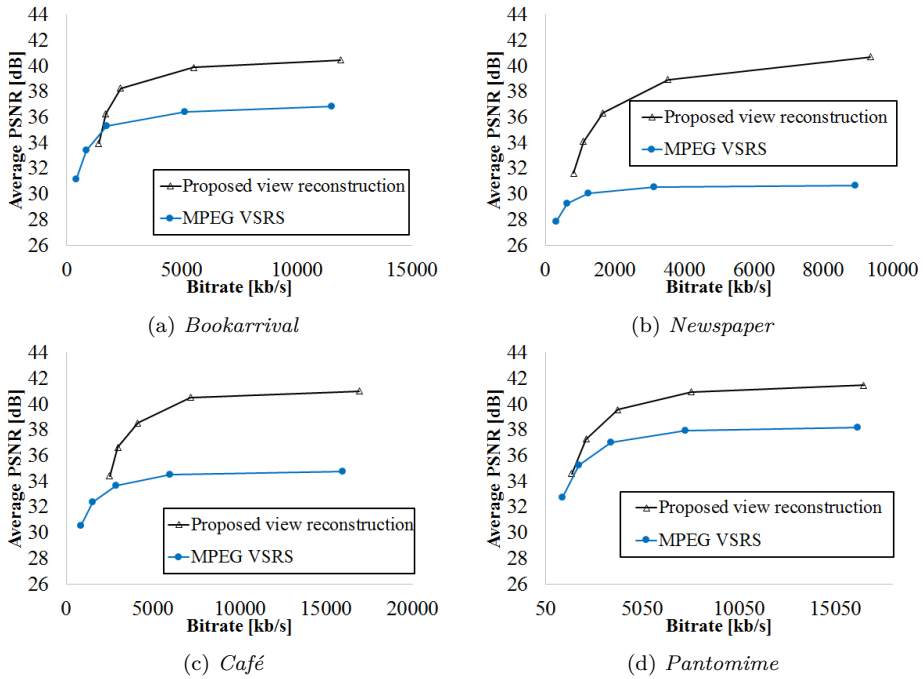


Fig. 7: Proposed view reconstruction overhead and quality performance. The colours represent calculated bitrate and quality performance for a block-size of 32×32, 16×16, 8×8, and variable block-size.

view. For fair comparison, this reference approach employs the same view discarding approach (*i.e.,* priority method) as the proposed system and utilises MPEG-VSRS to estimate discarding views. Average reconstruction quality, which was calculated in terms of PSNR, was estimated using only discarded views for each piece of MVV content.



(a) *Bookarrival* (b) *Newspaper*

(c) *Café* (d) *Pantomime*

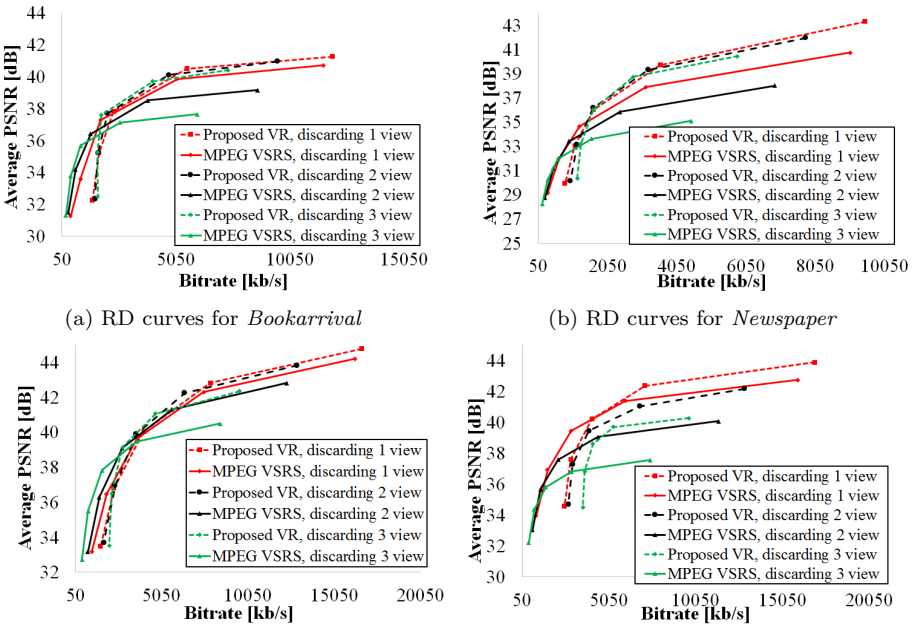
Fig. 8: RD comparison curves for the proposed view reconstruction and the MPEG-VSRS algorithms.

Fig. 8 shows average PSNR scores for the proposed view reconstruction and the MPEG-VSRS algorithms for the *BookArrival*, *Newspaper*, *Café*, and *Pantomime* video sequences. The transmission bitrate included the total transmitted data and SI. As seen in the RD curves, 2.8, 7.13, 5.04, and 5.26 dB average PSNR gains (ΔP) are achieved over the entire range of bitrates with respect to the MPEG-VSRS for *BookArrival*, *Newspaper*, *Café*, and *Pantomime* sequences, respectively.

Experiment results demonstrate that the proposed view reconstruction algorithm presents significant coding efficiency compared to the MPEG-VSRS method. The overall gains can be explained by two intrinsic novelties of the proposed reconstruction approach: 1) exploiting all available views (*see* Fig. 2) during the view reconstruction process. This feature provides high quality occlusion filling, which enhances the overall view estimation quality at the decoder. High numbers of reference views minimise the occlusion problems on the estimated view and improve the view estimation performance. 2) optimisation process in the variable block-size selection. This process alternates between metadata overhead and view estimation quality, which allows the transmission of an optimum overhead of metadata without significantly decreasing the overall quality.

5.2.4 Effect of The Number of Discarded Views

Fig. 9 depicts the average PSNR versus the total bitrate requirement for three different MVV contents. In this experiment, several views were discarded and then reconstructed using other available neighbouring views at the receiver side.



(a) RD curves for *Bookarrival*

(b) RD curves for *Newspaper*

(c) RD curves for *Café*

(d) RD curves for *Pantomime*

Fig. 9: Effect of the number of discarded views on the overall MVV streaming quality.

The reported results clearly demonstrate that a significant PSNR gain is achieved by the proposed approach against MPEG-VSRS for all tested MVV content. This advantage is due to the high-quality view estimation and decreasing bitrate with SI.

For the *Newspaper* sequence, the gain is more visible because of their high spatial complex scene (*see* SI value in Fig. 5). Moreover, the coding gain gap between the proposed approach and MPEG-VSRS is smaller for the *BookArrival* and *Pantomime* sequences relative to other sequences. The main reason is that both sequences contain complex object motion, which produces new occlusion areas on the projected views. This problem can be solved by transmitting a new codebook when the existing codebook is not sufficient for estimating new occlusion areas with optimum performance. Also, performance reduction links to the accuracy of depth maps. The *Café* sequence, contains inaccurate depth maps in comparison to other sequences, demonstrates low performance below 5000 *kbps*. The reason is the proposed approach requires a greater number of weighting coefficients to reconstruct with erroneous depth maps with higher quality.

Furthermore, it is clearly observed that as the number of discarded views increases, PSNR values for each content severely decreases. The overall MVV stream-

ing performance depends on the quality of the estimated views; hence, it is also linked to the accuracy of the depth maps and the efficiency of the view estimation algorithm. Obviously, it also depends on the complexity of the scene that needs to be estimated. This experiment clearly demonstrates that MPEG-VSRS weaknesses affect adaptation performance. In addition, the results show that the proposed approach helps enhance adaptation performance by a notable margin consistently over a different number of discarded views, *e.g.,* 1, 2, and 3.

5.2.5 Objective and Subjective Coding Efficiency

The objective (average PSNR versus total bitrate) and subjective (DMOS versus bitrate) evaluation results are demonstrated in Fig. 10. In this assessment, the performance of the proposed adaptive streaming method was compared with the reference adaptation method (based on MPEG-VSRS) using two different video coding standards. It is noted that only discarded views were considered to calculate the average PSNR. DMOS were also calculated based on the described test setup (*see* Section 5.1). The bitrate requirement was calculated using all transmitted sequences with the SI stream bitrate.

As can be seen in Fig. 10, both the objective and subjective evaluation results lead to similar conclusions. The delivery of MVV sequences with the proposed approach provides a significant advantage for both coding standards in terms of bitrate gain compared with the MPEG-VSRS method. Experiment results also show that as SI constitutes a larger portion of the overall transmission at low operating bitrates, the relative coding gain decreases.

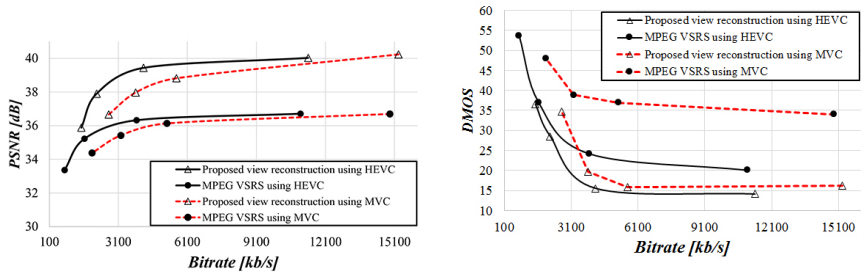
In addition, for HEVC bit-stream, the proposed method outperforms the reference method at the similar bitrate by 2.72, 7.75, and 5.27 dB on average for the *BookArrival*, *Newspaper*, and *Café* sequences, respectively. For MVC bit-stream, an average 2.63, 7.77, and 5.72 dB quality enhancements are achieved (ΔP) over the entire range of bitrates for *BookArrival*, *Newspaper*, and *Café* sequences, respectively. With respect to the overall transmission bitrate, 2.02%, 2.76%, and 2.85% SI overhead (ΔR) on average is obtained over the tested operating points.

Tab. 2 lists the bitrate saving percentage with the proposed adaptive streaming method subject to symmetrical compression of HEVC and MVC.

Table 2: Bitrate saving percentage (ΔR) with the proposed adaptive 3D MVV streaming method.

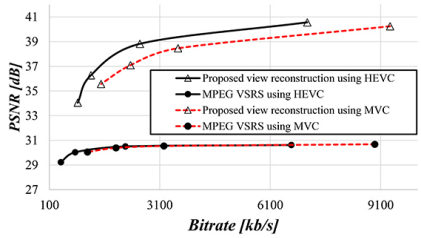
Sequence	*HEVC*	*MVC*
Bookarrival	79.14%	49.99%
Newspaper	25.46%	21.94%
Café	14.49%	49.27%

As can be seen in Tab. 2, the proposed method provides significant bitrate saving against two reference coding standards (HEVC and MVC) in terms of bitrate. Various coding performance is observed with MVC (*e.g.,* different trend in *Café*). The reason is the illumination mismatches between views and content types. As demonstrated in [83], these characteristics cause performance degradation in MVC that uses inter-view prediction. Moreover, because of symmetrical compression, blurred

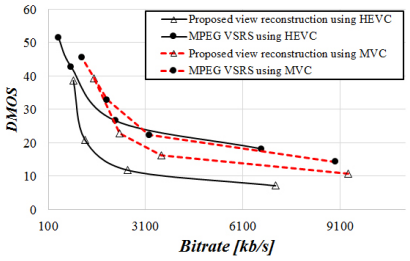


(a) RD curves for *Bookarrival*

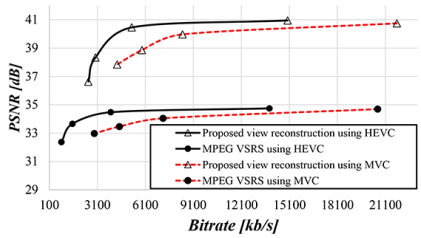
(b) DMOS curves for *Bookarrival*



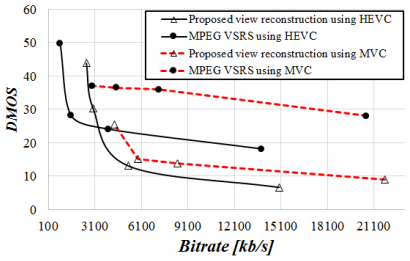
(c) RD curves for *Newspaper*



(d) DMOS curves for *Newspaper*



(e) RD curves for *Café*



(f) DMOS curves for *Café*

Fig. 10: Objective (left) and subjective (right) evaluation results for *Bookarrival*, *Newspaper*, and *Café* MVV contents.

images and blocking artifacts were observed for each MVV content when high QPs are applied to match the network bandwidth. Subjective experiments suggest discarding some views for transmission and recovering them with the proposed view reconstruction approach at the receiver side.

5.2.6 Performance Analysis Over The Dynamic Network Environment

In order to evaluate the impact of the adaptation pattern (*i.e.,* view discarding order in the `MPD file`), a test pattern with instantaneous network throughput changes was used as depicted in Fig. 11. Due to the lack of test pattern for MVV, we create these challenging patterns for MVV streaming. `Test 1` corresponds to the varying link

capacity case, whereas `Test 2` and `Test 3` correspond to the fixed link capacity case throughout the streaming. MVV temporal segments were selectively discarded based on the available bandwidth capacity.

In this experiment, the HEVC standard was used with the MPEG-DASH. All streams were divided into segments, encoded with various QP (*see* Section 5.1), and stored in the DASH server. Avoiding inter-view dependencies within the proposed framework will offer increased flexibility in view combinations that can be discarded independently and are replaced with the corresponding SI stream. Also, it prevents potential inter-view error propagation. In addition, the proposed adaptation system may reduce the quality of all the transmitted views based on the available bandwidth.

Furthermore, MPEG-VSRS was used as an adaptation reference, in which MVV temporal segments were selectively discarded based on the available bandwidth. Also, this adaptation reference may also increase/decreases QP of the all transmitted views.

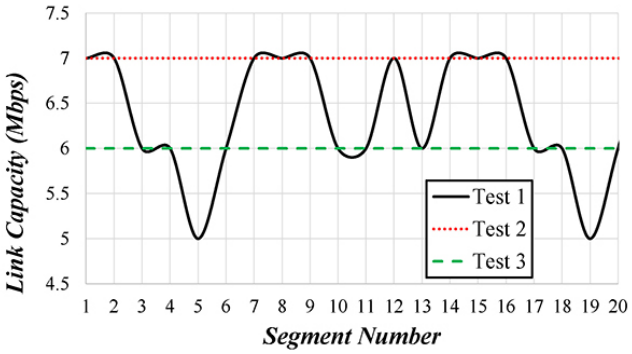


Fig. 11: Link capacity of the client in the network.

Table 3: View discarding pattern between 1st segment and 7th segment for varying link capacity test case.

Segment Range	(0.5sec./segment)	*1-2*	*2-3*	*3-4*	*4-5*	*5-6*	*6-7*
Method	**Sequence**	**Discarded viewpoint number(s)**					
Proposed View Reconstruction	*Bookarrival*	-	8	-	8	7	-
MPEG View Synthesis		-	8	-	8	8	-
Proposed View Reconstruction	*Newspaper*	5	4,3	3,5	3,4	4,5	5
MPEG View Synthesis		3	5,3	5,3	5,3	5,3	5
Proposed View Reconstruction	*Café*	-	-	4	3	-	4
MPEG View Synthesis		-	-	4	4	-	4
Proposed View Reconstruction	*Pantomime*	-	4	3	4	-	-
MPEG View Synthesis		-	4	4	3	-	-

The view discarding order (*i.e.,* which views were discarded) in `Test 1` for both adaptation methods is shown in Tab. 3.

In addition, Tab. 4 shows the comparison in terms of PSNR and subjective scores that are reported as an average of all views (delivered and discarded/estimated

views). For subjective experiments, MOS values are converted to quality scales (Bad, Poor, Fair, Good, and Excellent) based on the ITU-R BT.500-13 recommendation.

The results depicted in the Tab. 4 show that the proposed adaptation method consistently outperforms the reference method, both objectively and subjectively, in all test conditions.

Table 4: Comparison of the adaptation methods.

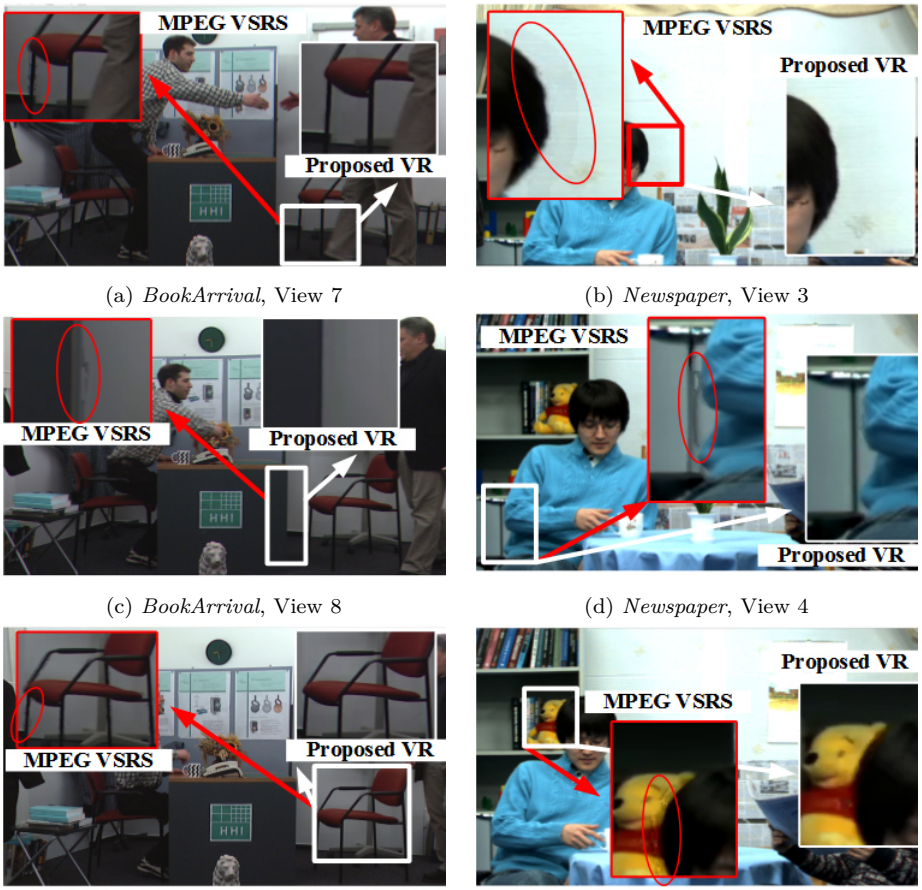
	Method	**Quality**	**Test Conditions**		
			Test 1	*Test 2*	*Test 3*
Bookarrival	Proposed View Reconstruction	PSNR (dB)	39.37	39.74	39.68
		Subjective	Excellent	Excellent	Excellent
	MPEG View Synthesis	PSNR (dB)	37.61	39.74	39.52
		Subjective	Good	Good	Good
Newspaper	Proposed View Reconstruction	PSNR (dB)	42.53	42.41	41.65
		Subjective	Excellent	Excellent	Excellent
	MPEG View Synthesis	PSNR (dB)	40.52	41.17	39.18
		Subjective	Good	Good	Fair
Café	Proposed View Reconstruction	PSNR (dB)	42.03	42.22	41.72
		Subjective	Excellent	Excellent	Excellent
	MPEG View Synthesis	PSNR (dB)	41.59	42.22	41.28
		Subjective	Excellent	Excellent	Good
Pantomime	Proposed View Reconstruction	PSNR (dB)	42.36	42.82	41.41
		Subjective	Excellent	Excellent	Excellent
	MPEG View Synthesis	PSNR (dB)	41.12	41.71	40.89
		Subjective	Excellent	Good	Good

5.2.7 Visual Quality Performance

The presented results demonstrate that the video coding standards with the proposed approach performs objectively (PSNR) better than the MPEG-VSRS. These improvements are also visible in Figures 12 and 13, which illustrate the visual quality of the *BookArrival*, *Newspaper*, *Café*, and *Pantomime* test sequences.

In this analysis, thumbnails of the reconstructed discarded view images were captured and illustrated. Around the object edges, the occluded areas that do not exist with the proposed method are clearly shown. It is clear that particular object boundaries that look distorted in the MPEG-VSRS are better conserved with the proposed View Reconstruction (*proposed VR*) method. The reason is the edge pixels are not scattered, given that the local high-frequency components are conserved successfully with an additional SI. Consequently, a sharper and more robust perception is achieved with the proposed method.

In the extensive evaluation, the proposed method demonstrates the highest QoE video streaming performance. Especially, results show that exploiting more views with metadata provide high quality 3D MVV reconstruction. However, for inaccurate depth maps, the proposed method requires important amount of metadata to reconstruct discarding views with high quality. Additionally, the performance reduction was observed for some sequences that contain complex object motion (*e.g.,* fast



(a) *BookArrival*, View 7 (b) *Newspaper*, View 3

(c) *BookArrival*, View 8 (d) *Newspaper*, View 4

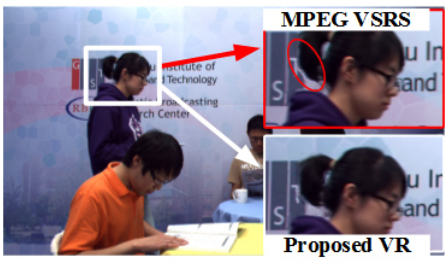
(e) *BookArrival*, View 9 (f) *Newspaper*, View 5

Fig. 12: Visual view reconstruction performance comparison between the *MPEG-VSRS* and the proposed view reconstruction (*proposed VR*) method. The resulting reconstruction views are shown for two different MVV contents: *BookArrival* (left) and *Newspaper* (right). Three successive views for each content are shown and the most representative distortion areas are marked with red.

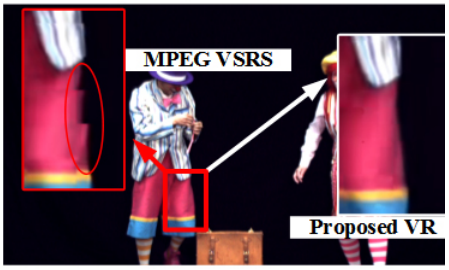
moving objects, quick scene changes). In order to solve these issue, as a future work, the codebook will be transmitted adaptively (*e.g.,* based on the characteristics of scene change and object motion) in order to further enhance the streaming quality.

6 Conclusions

To maintain the perceived 3D MVV quality in congested networks, this paper suggests a novel adaptive delivery scheme. The proposed method yields a superior performance over a wide range of channel conditions. In this system, some views are discarded at times of network congestion in an intelligent way to maximise the re-

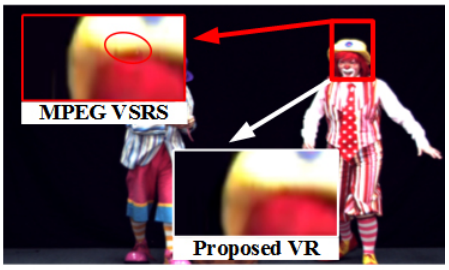


(a) *Café*, View 2

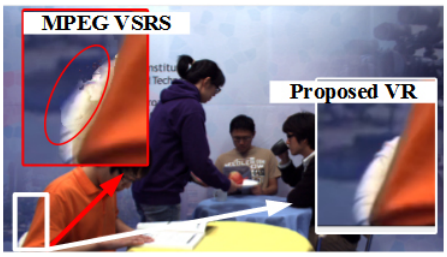


(b) *Pantomime*, View 38

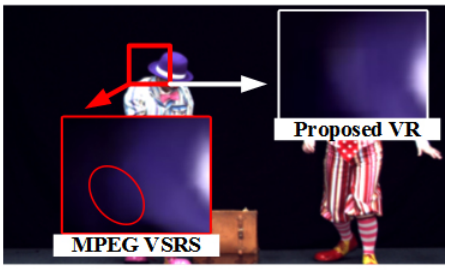
(c) *Café*, View 3



(d) *Pantomime*, View 39



(e) *Café*, View 4



(f) *Pantomime*, View 40

Fig. 13: Visual view reconstruction performance comparison between the *MPEG-VSRS* and the proposed view reconstruction (*proposed VR*) method. The resulting reconstruction views are shown for two different MVV contents: *Café* (left) and *Pantomime* (right). Three successive views for each content are shown and the most representative distortion areas are marked with red.

sultant reconstruction performance on the client side. The discarded views are reconstructed using only a small amount of additional metadata that is estimated in the server and sent to the client.

In the proposed method, the additional metadata is calculated using adjacent views in the server and delivered to the client at times of congestion in the network. Also, a novel view reconstruction method is designed to take into account the received Side Information (SI) for improved view reconstruction performance. In order to help facilitate a quality-aware bandwidth adaptation mechanism, the best sets of views to be discarded are calculated for various network throughput levels, such that the best overall MVV reconstruction quality is achieved on the client side.

The proposed adaptive 3D MVV streaming method was evaluated using a prototype HTTP streaming client and two different state-of-the-art coding standards. These were compared to the corresponding reference techniques that use MPEG's reference view synthesis software (VSRS). The experiment results have shown that significant quality improvements are obtained under challenging network conditions.

For some content types, the performance reduction has been reported. In doing so, as a future work, perceptual MVV metric, adaptive codebook transmission, and advanced loss-resilient coding will be integrated in the proposed adaptive delivery scheme. Also, the proposed adaptive MVV streaming mechanism will be adapted to the emerging information centric networks.

Acknowledgements This work was supported by the ROMEO project (grant number: 287896), which was funded by the EC FP7 ICT collaborative research programme. This paper is an extended version of the original paper [19] which appeared in the Proceedings of the 2013 ACM International Workshop on Immersive Media Experiences [84]. Special thanks to the anonymous reviewers and program chairs in the workshop and the journal for their constructive comments and suggestions that assisted in enhancing the paper.

References

1. L. Onural. Television in 3D: What are the prospects? *Proceedings of the IEEE*, 95(6):1143–1145, June 2007.
2. A. Kondoz and T. Dagiuklas. *3D Future Internet Media*. Springer, 2014.
3. F. Dufaux, B. Pesquet-Popescu, and M. Cagnazzo. *Emerging technologies for 3D video: creation, coding, transmission and rendering*. John Wiley & Sons, 2013.
4. P. Benzie, J. Watson, P. Surman, I. Rakkolainen, K. Hopf, H. Urey, V. Sainov, and C.V. Kopylow. A survey of 3DTV displays: Techniques and technologies. *Circuits and Systems for Video Technology, IEEE Transactions on*, 17(11):1647–1658, Nov 2007.
5. C. Zhang, Z. Yin, and D. Florencio. Improving depth perception with motion parallax and its application in teleconferencing. In *Multimedia Signal Processing, MMSP '09. IEEE International Workshop on*, pages 1–6, Rio De Janeiro, Oct 2009.
6. Alioscopy. Alioscopy glasses-free 3D displays, http://www.alioscopy.com/en/3ddisplays.php, Jan 2016.
7. W. Matusik and H. Pfister. 3D TV: A scalable system for real-time acquisition, transmission, and autostereoscopic display of dynamic scenes. *ACM Transactions on Graphics*, 23(3):814–824, Aug 2004.
8. Dimenco. Dimenco displays, http://www.dimenco.eu/3d-displays/displays/, Jan 2016.
9. M. Tanimoto. Overview of FTV (free-viewpoint television). In *Multimedia and Expo, 2009. ICME 2009. IEEE International Conference on*, pages 1552–1553, New York, NY, June 2009.
10. V. Jacobson. Congestion avoidance and control. *ACM SIGCOMM Computer Communication Review*, (1):314–329, 1988.
11. J. Chakareski. Adaptive multiview video streaming: challenges and opportunities. *Communications Magazine, IEEE*, 51(5):94–100, May 2013.
12. K. Miller, E. Quacchio, G. Gennari, and A. Wolisz. Adaptation algorithm for adaptive streaming over HTTP. *2012 19th International Packet Video Workshop (PV)*, pages 173–178, May 2012.
13. Apple. Apple HTTP live streaming, https://developer.apple.com/streaming/, Jan 2016.
14. Microsoft. Microsoft Smooth-Streaming, http://www.iis.net/downloads/microsoft/smooth-streaming, Jan 2016.
15. I. Sodagar. The MPEG-DASH Standard for Multimedia Streaming Over the Internet. *MultiMedia, IEEE*, 18(4):62–67, Apr. 2011.
16. T. Stockhammer. Dynamic Adaptive Streaming over HTTP –: Standards and Design Principles. In *Proceedings of the Second Annual ACM Conference on Multimedia Systems*, MMSys '11, pages 133–144, New York, NY, USA, 2011. ACM.

17. K. Sripanidkulchai, B. Maggs, and H. Zhang. An Analysis of Live Streaming Workloads on the Internet. In *Proceedings of the 4th ACM SIGCOMM Conference on Internet Measurement*, IMC '04, pages 41–54, New York, NY, USA, 2004. ACM.
18. C. Ozcinar, E. Ekmekcioglu, and A. Kondoz. Adaptive 3D multi-view video streaming over P2P networks. In *Image Processing (ICIP), 2014 IEEE International Conference on*, pages 2462–2466, Paris, Oct 2014.
19. C. Ozcinar, E. Ekmekcioglu, and A. Kondoz. Dynamic adaptive 3D multi-view video streaming over the Internet. In *Proceedings of the 2013 ACM International Workshop on Immersive Media Experiences*, ImmersiveMe '13, pages 51–56, Barcelona, Spain, 2013. ACM.
20. A. Vetro and I. Sodagar. Industry and Standards The MPEG-DASH Standard for Multimedia Streaming Over the Internet. *IEEE MultiMedia*, 18(4):62–67, 2011.
21. Recommendation, ITU-R. Recommendation ITU-R BT.500-13, Methodology for the subjective assessment of the quality of television pictures, Jan 2012.
22. A. Seema, L. Schwoebel, T. Shah, J. Morgan, and M. Reisslein. WVSNP-DASH: Name-Based Segmented Video Streaming. *Broadcasting, IEEE Transactions on*, 61(3):346–355, Sept 2015.
23. A. Vetro, T. Wiegand, and G.J. Sullivan. Overview of the stereo and multiview video coding extensions of the H.264/MPEG-4 AVC standard. *Proceedings of the IEEE*, 99(4):626–642, Apr 2011.
24. T. Wiegand, G.J. Sullivan, G. Bjontegaard, and A. Luthra. Overview of the H.264/AVC video coding standard. *Circuits and Systems for Video Technology, IEEE Transactions on*, 13(7):560–576, July 2003.
25. Report of the subjective quality evaluation for MVC call for evidence. Technical Report N6999, ISO/IEC JTC1/SC29/WG11, Hong Kong, China, Jan. 2005.
26. A. Vetro, A.M. Tourapis, K. Müller, and C. Tao. 3D-TV content storage and transmission. *Broadcasting, IEEE Transactions on*, 57(2):384–394, June 2011.
27. A. Smolic, K. Mueller, N. Stefanoski, J. Ostermann, A. Gotchev, G.B. Akar, G. Triantafyllidis, and A Koz. Coding algorithms for 3DTV - a survey. *Circuits and Systems for Video Technology, IEEE Transactions on*, 17(11):1606–1621, Nov 2007.
28. G.J. Sullivan, J. Ohm, W.-J. Han, and T. Wiegand. Overview of the high efficiency video coding (HEVC) standard. *Circuits and Systems for Video Technology, IEEE Transactions on*, 22(12):1649–1668, Dec 2012.
29. K. Wegner, O. Stankiewicz, K. Klimaszewski, and M. Domański. Comparison of multiview compresion performance using MPEG-4 MVC and prospective HVC technology. Technical Report MPEG M17913, ISO/IEC JTC1/SC29/WG11, Geneve, Switzerland, July 2010.
30. Call for proposals on 3D video coding technology. Technical Report MPEG2011/N12036, ISO/IEC JTC1/SC29/WG11, Geneva, Switzerland, March 2011.
31. M. Domański, O. Stankiewicz, K. Wegner, M. Kurc, J. Konieczny, J. Siast, J. Stankowski, R. Ratajczak, and T. Grajek. High efficiency 3D video coding using new tools based on view synthesis. *Image Processing, IEEE Transactions on*, 22(9):3517–3527, Sept 2013.
32. K. Müller, H. Schwarz, D. Marpe, C. Bartnik, S. Bosse, H. Brust, T. Hinz, H. Lakshman, P. Merkle, F.H. Rhee, G. Tech, M. Winken, and T. Wiegand. 3D high-efficiency video coding for multi-view video and depth data. *Image Processing, IEEE Transactions on*, 22(9):3366–3378, Sept 2013.
33. G.J. Sullivan, J.M. Boyce, Ying Chen, J.-R. Ohm, C.A. Segall, and A. Vetro. Standardized Extensions of High Efficiency Video Coding (HEVC). *Selected Topics in Signal Processing, IEEE Journal of*, 7(6):1001–1016, Dec 2013.
34. P. Seeling and M. Reisslein. The rate variability-distortion (VD) curve of encoded video and its impact on statistical multiplexing. *Broadcasting, IEEE Transactions on*, 51(4):473–492, Dec 2005.
35. A. Pulipaka, P. Seeling, M. Reisslein, and L.J. Karam. Traffic and Statistical Multiplexing Characterization of 3-D Video Representation Formats. *Broadcasting, IEEE Transactions on*, 59(2):382–389, June 2013.
36. M. Tanimoto, T. Fujii, K. Suzuki, N. Fukushima, and Y. Mori. Reference softwares for depth estimation and view synthesis. Technical Report MPEG2008/M15377, ISO/IEC JTC1/SC29/WG11, Archamps, Apr 2008.
37. Call for Evidence on Free-Viewpoint Television: Super-Multiview and Free Navigation. Technical Report MPEG2015/N15733, ISO/IEC JTC1/SC29/WG11, Geneva, Switzerland, October 2015.

38. MPEG 3-DV View Synthesis Reference Software. http://wg11.sc29.org/svn/repos/mpeg-4/test/trunk/3d/view.synthesis/, Jan 2016.
39. M. Köppel, P. Ndjiki-Nya, D. Doshkov, H. Lakshman, P. Merkle, K. Müller, and T. Wiegand. Temporally consistent handling of disocclusions with texture synthesis for depth-image-based rendering. In *Image Processing (ICIP), 2010 17th IEEE International Conference on*, pages 1809–1812, Sept 2010.
40. W.-S. Kim, A. Ortega, P. Lai, D. Tian, and C. Gomila. Depth map distortion analysis for view rendering and depth coding. In *Image Processing (ICIP), 16th IEEE International Conference on*, pages 721–724, Nov 2009.
41. Q. Zhang, L. Tian, L. Huang, X. Wang, and H. Zhu. Rendering distortion estimation model for 3D high efficiency depth coding. *Mathematical Problems in Engineering*, 2014(940737):7, Jan 2014.
42. N-M Cheung, D. Tian, A. Vetro, and Huifang Sun. On modeling the rendering error in 3D video. In *Image Processing (ICIP), 2012 19th IEEE International Conference on*, pages 3021–3024, Sept 2012.
43. K.-J. Oh, S. Yea, and Y.-S. Ho. Hole filling method using depth based in-painting for view synthesis in free viewpoint television and 3-D video. In *Picture Coding Symposium, 2009. PCS 2009*, pages 1–4, Chicago, IL, May 2009.
44. C.-M. Cheng, S.-J. Lin, and S.-H. Lai. Spatio-Temporally Consistent Novel View Synthesis Algorithm From Video-Plus-Depth Sequences for Autostereoscopic Displays. *Broadcasting, IEEE Transactions on*, 57(2):523–532, June 2011.
45. G. Cheung, V. Velisavljevic, and A Ortega. On dependent bit allocation for multiview image coding with depth-image-based rendering. *Image Processing, IEEE Transactions on*, 20(11):3179–3194, Nov 2011.
46. W. Sun, G. Cheung, P.A. Chou, D. Florencio, Cha Zhang, and O.C. Au. Rate-distortion optimized 3D reconstruction from noise-corrupted multiview depth videos. In *Multimedia and Expo (ICME), 2013 IEEE International Conference on*, pages 1–6, San Jose, CA, July 2013.
47. P. Carballeira, G. Tech, J. Cabrera, K. Müler, F. Jaureguizar, T. Wiegand, and N. Garcia. Block based Rate-Distortion analysis for quality improvement of synthesized views. In *3DTV-Conference: The True Vision - Capture, Transmission and Display of 3D Video (3DTV-CON), 2010*, pages 1–4, Tampere, June 2010.
48. B.J. Dempsey, J. Liebeherr, and A.C. Weaver. On Retransmission-based Error Control for Continuous Media Traffic in Packet-switching Networks. *Computer Networks and ISDN Systems*, 28(5):719–736, March 1996.
49. T.C. Thang, Q.-D. Ho, J.W. Kang, and A.T. Pham. Adaptive streaming of audiovisual content using MPEG DASH. *Consumer Electronics, IEEE Transactions on*, 58(1):78–85, Feb 2012.
50. C.G. Gürler, K.T. Bagci, and A.M. Tekalp. Adaptive stereoscopic 3D video streaming. In *Image Processing (ICIP), 2010 17th IEEE International Conference on*, pages 2409–2412, Hong Kong, Sept 2010.
51. S.S. Savas, C.G. Gurler, A.M. Tekalp, E. Ekmekcioglu, S. Worrall, and A. Kondoz. Adaptive streaming of multi-view video over P2P networks. *Signal Processing: Image Communication*, 27(5):522 – 531, 2012.
52. B. Cohen. Incentives build robustness in BitTorrent. In *Workshop on Economics of Peer-to-Peer systems*, volume 6, pages 68–72, 2003.
53. B. Pourebrahimi, K. Bertels, and S. Vassiliadis. A survey of peer-to-peer networks. In *Proceedings of the 16th Annual Workshop on Circuits, Systems and Signal Processing*, 2005.
54. B. Oztas, M.T. Pourazad, P. Nasiopoulos, I. Sodagar, and V.C.M. Leung. A rate adaptation approach for streaming multiview plus depth content. In *Computing, Networking and Communications (ICNC), 2014 International Conference on*, pages 1006–1010, Honolulu, HI, Feb 2014.
55. S.S. Savas, C.G. Gurler, and A.M. Tekalp. Evaluation of adaptation methods for multi-view video. In *Image Processing (ICIP), 2012 19th IEEE International Conference on*, pages 2273–2276, Orlando, FL, Sept 2012.
56. V. Jacobson, D.K. Smetters, J.D. Thornton, M.F. Plass, N.H. Briggs, and R.L. Braynard. Networking named content. In *Proceedings of the 5th International Conference on Emerging Networking Experiments and Technologies*, CoNEXT '09, pages 1–12, New York, NY, USA, 2009. ACM.

57. L. Yaning, J. Geurts, J.-C. Point, S. Lederer, B. Rainer, C. Müller, C. Timmerer, and H. Hellwagner. Dynamic adaptive streaming over CCN: A caching and overhead analysis. In *Communications (ICC), 2013 IEEE International Conference on*, pages 3629–3633, June 2013.
58. A. Detti, B. Ricci, and N. Blefari-Melazzi. Mobile Peer-to-peer Video Streaming over Information-centric Networks. *Computer Networks*, 81(C):272–288, Apr 2015.
59. Y. Sugiyama. An algorithm for solving discrete-time Wiener-Hopf equations based upon Euclid's algorithm. *Information Theory, IEEE Transactions on*, 32(3):394–409, May 1986.
60. A.M. Kondoz. *Digital Speech: coding for low bitrate communication systems*. Wiley Online Library, 2004.
61. E. Christensen, F. Curbera, G. Meredith, and S. Weerawarana. Web services description language (WSDL) 1.1, Jan 2016.
62. T. Kanungo, D.M. Mount, N.S. Netanyahu, C.D. Piatko, R. Silverman, and A.Y. Wu. An efficient k-means clustering algorithm: Analysis and implementation. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 24(7):881–892, Jul 2002.
63. G.J. Sullivan and R. Baker. Efficient quadtree coding of images and video. *Image Processing, IEEE Transactions on*, 3(3):327–331, May 1994.
64. P. Helle, S. Oudin, B. Bross, D. Marpe, M.O. Bici, K. Ugur, J. Jung, G. Clare, and T. Wiegand. Block merging for quadtree-based partitioning in HEVC. *Circuits and Systems for Video Technology, IEEE Transactions on*, 22(12):1720–1731, Dec 2012.
65. G.J. Sullivan and T. Wiegand. Rate-distortion optimization for video compression. *Signal Processing Magazine, IEEE*, 15(6):74–90, Nov 1998.
66. M. Lightstone and S.K. Mitra. Quadtree optimization for image and video coding. *Journal of VLSI signal processing systems for signal, image and video technology*, 17(2-3):215–224, 1997.
67. G. Tech, H. Schwarz, K. Müller, and T. Wiegand. 3D video coding using the synthesized view distortion change. In *Picture Coding Symposium (PCS), 2012*, pages 25–28, Krakow, May 2012.
68. Z. Wang, A.C. Bovik, H.R. Sheikh, and E.P. Simoncelli. Image quality assessment: from error visibility to structural similarity. *Image Processing, IEEE Transactions on*, 13(4):600–612, Apr 2004.
69. Y. Zhao and L. Yu. A perceptual metric for evaluating quality of synthesized sequences in 3DV system. In *Proceedings of SPIE Vol*, volume 7744, pages 77440X–1, 2010.
70. I. Feldmann, M. Mueller, F. Zilly, R. Tanger, K. Mueller, A. Smolic, P. Kauff, and T. Wiegand. HHI test material for 3D video. Technical Report MPEG2008/M15413, ISO/IEC JTC1/SC29/WG11, Archamps, France, Apr 2008.
71. 3DV Sequences of HHI. Fraunhofer heinrich hertz institute, berlin, germany, ftp://ftp.hhi.de/hhimpeg3dv, Jan 2016.
72. Y.-S. Ho, E.-K. Lee, and C. Lee. Multiview video test sequence and camera parameters. Technical Report MPEG2008/M15419, ISO/IEC JTC1/SC29/WG11, Archamps, France, Apr 2008.
73. 3DV Sequences of ETRI and GIST. Electronics and telecommunications research institute and gwangju institute of science and technology,gwangju, korea, ftp://203.253.128.142/, Jan 2016.
74. Y.-S. Kang, E.-K. Lee, J.-I. Jung, J.-H. Lee, and I.-Y. Shin. 3D video test sequence and camera parameters. Technical Report MPEG2009/M16949, ISO/IEC JTC1/SC29/WG11, Sian, China, Oct 2009.
75. 3DV Sequences of Nagoya University. Nagoya University, Japan, http://www.tanimoto.nuee.nagoya-u.ac.jp/MPEG-FTVProject.html, Jan 2016.
76. P.910 ITU-T Recommendation. Subjective video quality assessment methods for multimedia applications. 1999.
77. C. Fehn. Depth-image-based rendering (DIBR), compression, and transmission for a new approach on 3D-TV. *Proceedings SPIE*, 5291:93–104, 2004.
78. S. Lederer, C. Müller, and C. Timmerer. Dynamic adaptive streaming over HTTP dataset. In *Proceedings of the 3rd Multimedia Systems Conference*, MMSys '12, pages 89–94, New York, NY, USA, 2012. ACM.
79. C. Müller, S. Lederer, and C. Timmerer. An evaluation of dynamic adaptive streaming over HTTP in vehicular environments. In *Proceedings of the 4th Workshop on Mobile Video*, MoVid '12, pages 37–42, New York, NY, USA, 2012. ACM.
80. L. Rizzo. Dummynet: A simple approach to the evaluation of network protocols. *ACM SIGCOMM Computer Communication Review*, 27(1):31–41, Jan 1997.

81. G. Bjøtegaard. Calculation of average PSNR differences between RD-curves (vceg-m33). Technical Report M16090, VCEG Meeting (ITU-T SG16 Q.6), Austin, Texas, USA„ Apr 2001.
82. F. Lewandowski, M. Paluszkiewicz, T. Grajek, and K. Wegner. Subjective quality assessment methodology for 3D video compression technology. In *Signals and Electronic Systems (ICSES), 2012 International Conference on*, pages 1–5, Sept 2012.
83. J.-H. Hur, S. Cho, and Y.-L. Lee. Adaptive local illumination change compensation method for H.264/AVC-based multiview video coding. *Circuits and Systems for Video Technology, IEEE Transactions on*, 17(11):1496–1505, Nov 2007.
84. T. Chambel, V.M. Bove, S. Strover, P. Viana, and G. Thomas. ImmersiveMe '13: Proceedings of the 2013 ACM International Workshop on Immersive Media Experiences. In *Proceedings of the 2013 ACM International Workshop on Immersive Media Experiences*, Barcelona, Spain, 2013. ACM. 433137.