Multimedia Tools and Applications

Adaptive Multi-View Video Streaming using Side Information over Peer to Peer Networks

--Manuscript Draft--

Manuscript Number:	MTAP-D-17-02695R1
Full Title:	Adaptive Multi-View Video Streaming using Side Information over Peer to Peer Networks
Article Type:	Manuscript
Keywords:	Multi-view plus-depth-map (MVD); peer-to-peer (P2P); Compression; adaptation; streaming
Corresponding Author:	Cagri Ozcinar University of Dublin Trinity College Dublin, IRELAND
Corresponding Author Secondary Information:	
Corresponding Author's Institution:	University of Dublin Trinity College
Corresponding Author's Secondary Institution:	
First Author:	Cagri Ozcinar
First Author Secondary Information:	
Order of Authors:	Cagri Ozcinar
	Erhan Ekmekcioglu
	Gholamreza Anbarjafari
	Ahmet Kondozi
Order of Authors Secondary Information:	
Funding Information:	

Dear Editor and Reviewers

We had submitted a revised version of our manuscript with the title of “Adaptive Multi-View Video Streaming using Side Information over Peer to Peer Networks” before, and the manuscript was accepted with major revisions.

We have revised the submitted manuscript and now we believe that our revised manuscript addresses all comments from the reviewers.

Below you can see our answers (in blue) to the reviewers' comments (in black). In the revised version of the manuscript, we highlight modifications in red to facilitate the revision.

Best regards,
Authors

COMMENTS FOR THE AUTHOR:

Reviewer #1:

1. The authors did not provide solid achievements in this manuscript since this paper seems to be a somewhat incremental piece of work based on earlier research results.

We would like to thank the reviewer for his/her comment. Based on the reviewer's suggestion in the revised manuscript we have add the following paragraph stating the main contribution of this work:

“The main contribution of this work is conducting comprehensive study on multi-view plus depth streaming and exploiting results on state-of-the-art techniques. In addition, this work contributes by extending the work in [7] and presenting a more comprehensive performance analysis including additional multi-view plus depth map video sequences, P2P simulations and indicating its advantages and disadvantages.”

/* 7. Ozcinar, C., Ekmekcioglu, E. and Kondo, A., 2014, October. Adaptive 3D multi-view video streaming over P2P networks. In Image Processing (ICIP), 2014 IEEE International Conference on (pp. 2462-2466). IEEE. */

2. The proposed processes should be revised in a more formal pseudocode template.

Based on the suggestion of the reviewer we have added the pseudocode Algorithm 1 to the revised paper.

3. The reader may want to see how this work differs from other previous works (within the last 3 years). If possible, further technical details and explanations should be included.

We agree with the reviewer that it is important to emphasize on our contribution more clearly. In the revised manuscript we have added clearly what is our main contribution in this work and how different it is from the previous works. In this context, we included recent previous works for multi-view video streaming. Technical details, explanations, and their differences from this work can be founded in the related work section of the revised version of the manuscript.

4. Numerical results in this paper are not enough to support the conclusions. The comparison to other improved schemes (within the last 3 years) is required.

Thanks. In order to support the conclusions, we revised the experimental result section by stressing our findings. In addition, we have used all the commonly used metric and possible P2P experiments in this setup. Apart the method that we compared there is no other possible method in the literature for comparison.

5. The references should be in order. Moreover, formats of reference list lack consistency.

We would like to thank the reviewer for his/her careful review. We have now updated the order of references in the revised manuscript.

Reviewer #2:

The authors proposed a framework for adaptive multi-view video streaming over peer-to-peer networks for the purpose of handling high data volume and network congestion. My questions on the proposed algorithm and the experiments are as follows.

1. The contribution of the proposed framework is not clear to me comparing with the authors' previous work [23]. The authors could provide more analysis and experiments showing improvement over the previous system.

With thanks for the reviewer's comment, in the revised manuscript we have add the following paragraph stating the main contribution of this work:

"The main contribution of this work is conducting comprehensive study on multi-view plus depth streaming and exploiting results on state-of-the-art techniques. In addition, this work contributes by extending the work in [7] and presenting a more comprehensive performance analysis including additional multi-view plus depth map video sequences, P2P simulations and indicating its advantages and disadvantages."

/* 7. Ozcinar, C., Ekmekcioglu, E. and Kondo, A., 2014, October. Adaptive 3D multi-view video streaming over P2P networks. In Image Processing (ICIP), 2014 IEEE International Conference on (pp. 2462-2466). IEEE. */

2. I feel that the discussion and the literature survey on the state of the art multi-view streaming system over p2p network is insufficient. What's the unique characteristic of multi-view-plus-depth video stream? Why the proposed strategy works the best for peer-to-peer networks?

Thank you. In the revised manuscript we have improved the related work section with recent research works. Also, we have emphasized the main differences from the previous related works. We have included the following paragraph in the revised version of the manuscript:

3. In section 3.1, how is the code book generated at the first place? Is that generated for each video? Can you elaborate on that?

Thank you for this constructive comment. Codebook is generated for each discarded camera view for each video content. To elaborate this process, we have modified the revised version of the manuscript. In the revised manuscript, we have elaborated the generation of codebook.

4. Is the side information generated for each view in real time? I couldn't find how the weights w_k and w_N in formula (1) are calculated and how the block number x in formula (2) facilitates generating the weights. Can you provide more details on these steps?

We would like to thank to reviewers for his/her valuable comments. The generation of side information for each discarded camera view is done offline. After codebooks and side information are generated, they are stored at the server and transmitted to P2P clients (when they are necessary) for camera view reconstruction.

In order to clarify this steps, we have revised the section 3.1, e.g., adding new descriptions and equations.

5. How is the priority coefficient w_k in section 3.2 formula (4) determined?

Good point. We have checked the the Equations and improved them in order to clearly describe the calculation of priority coefficients.

6. For experiments, as mentioned in 1 and 2, it would be more convincing to compare the results with previous work and the other state of the art systems. And it could be even better to see the experiments

with not only simulations but how the system performs in real environments.

Thanks. To support the conclusions, we revised the experimental result section by stressing our findings. Also, we believe that we have used all the commonly used metric and possible P2P experiments in this particular setup. Apart from the method that we compared (MPEG VSRS), there is no other possible method in the literature for comparison.

We agree with the reviewer that the real environment setup provides practical results. However, in our setup, this requires code optimization with parallel processing setup, which is out the scope of our current research work.

7. What does sigma mean in figures 4-9?

Sigma represents deviation of PSNR over frames, calculating by standard deviation over frames. Thanks for the reviewer. To clarify this, descriptions are included in red color in the revised version of the manuscript.

Reviewer #4:

This paper propose to use the side information ver peer to peer networks for the adaptive multi-view video streaming. In general this paper is well written. I think this paper can be accpeted for publication.

We would like to thank the reviewer for his/her time and providing positive feedback about our work.

Adaptive Multi-View Video Streaming using Side Information over Peer to Peer Networks

Cagri Ozcinar · Erhan Ekmekcioglu ·
Gholamreza Anbarjafari · Ahmet Kondo

Received: 3 December 2017 / Accepted: date

Abstract Multi-view plus-depth-map (MVD) video streaming with autostereoscopic displays provides multi-user immersive media experiences. In this context, delivery of MVD representation to multiple clients remains a challenging problem because of the high-volume of data involved and the inherent limitations imposed by the delivery networks. To this end, this paper investigates the side information (SI) assisted adaptation algorithm using peer-to-peer (P2P) systems. P2P delivery systems for MVD video can maximize link utilization, preventing the transport of multiple video copies of the same packet for many users. However, the quality of experience (QoE) can be significantly degraded by dynamic variations caused by network congestions. To this end, our solution comprises the extraction of low-overhead metadata at the encoding server that is distributed through the P2P network as SI and used by P2P clients performing network adaptation. In the proposed adaptation strategy, pre-selected views are discarded at times of network congestion and reconstructed with an optimal reconstruction performance using the delivered SI and the delivered neighboring camera views. The experimental results show that the robustness of P2P multi-view streaming using the proposed adaptation scheme is significantly increased in the P2P network.

Keywords Multi-view plus-depth-map (MVD) · peer-to-peer (P2P) · compression · adaptation · streaming

Cagri Ozcinar
Trinity College Dublin (TCD), Dublin 2, Ireland. E-mail: ozcinar@scss.tcd.ie

Erhan Ekmekcioglu · Ahmet Kondo
Institute for Digital Technologies, Loughborough University London, London, UK.

Gholamreza Anbarjafari
iCV Research Group, University of Tartu, Tartu 50411, Estonia.
and Department of Electrical and Electronic Engineering, Hasan Kalyoncu University, Gaziantep, Turkey.

1 Introduction

Multi-view plus depth map (MVD) video representation contains a set of captured color textures and their associated depth maps, enables an immersive user experience from wider viewing angles on multi-view displays. Multi-view displays, for instance, auto-stereoscopic displays, require continuous delivery of a large number of camera views to maintain immersive 3D video experiences [1].

As the number of captured camera views increases, the transmitting of a potentially large number of camera views require considerable bandwidth, which is beyond of the most current Internet connections. MVD streaming with the traditional streaming may cause massive congestion in the network, leading to network collapse [2]. On the other hand, adaptive streaming techniques adjust the required bitrate of the video stream to dynamically changing bandwidth capacity on networks [3–5].

To transmit all captured MVD video sequences to multiple users while assuring consistently high quality requires effective delivery strategy in the context of adaptive MVD streaming. To this end, Peer-to-Peer (P2P) delivery methods play a significant role to distribute a large number of camera views to the end users. In the P2P network, each client in P2P, namely *peer*, downloads data as well as uploads them for the use of other peers. Hence, the bandwidth load of the server is reduced and distributed to peers. The universal data protocol (UDP), which is utilized as the transport layer for most current P2P streaming systems, does not provide a guarantee packet delivery services due to network congestion. Therefore, P2P networks pose severe challenges to service providers that must deliver the desired Quality of Experience (QoE).

To tackle the issues mentioned earlier, our previous works apply a novel view discarding and reconstructing algorithm in [6] to P2P networks [7], where the peer requests an additional bitstream, called side information (SI), during network congestion. The SI can reconstruct the complete set of camera views in a rate-distortion optimal way. In the work, we developed a new adaptive MVD streaming system using the P2P network, where each peer can discard some views based on their buffer occupancy level. The missing camera views are then reconstructed using only a small overhead of the SI, requested by the peers at times of congestion. We further investigated the performance of the proposed view discarding and reconstructing algorithm using the MPEG-DASH standard [8].

The main contribution of this work is conducting comprehensive study on MVD streaming and exploiting results on state-of-the-art techniques. In addition, this work contributes by extending the work in [7] and presenting a more comprehensive performance analysis including additional MVD video sequences, P2P simulations and indicating its advantages and disadvantages. It is shown that the proposed scheme leads to significance benefits in the context of MVD streaming on P2P networks.

The remainder of this paper is organized as follows. Related work is detailed in Section 2. Then, the proposed system overview and its main components

are respectively presented in Section 3. Experiments to demonstrate the performance of the proposed method are presented in Section 4. Finally, Section 5 concludes this paper with a summary and remarks.

2 Related Work

The first multi-view video streaming system using the P2P network was based on the delivery of 3D mesh objects to peers according to user interest [9]. The method allows the peers to obtain video from other peers while minimizing the server resource consumption. Similarly, Chen *et al.* proposed a P2P streaming framework for the multi-view video sequences that categorizes users in terms of a camera view interest [10]. The primary purpose of the work was to achieve short switching delay with an enhanced visual quality for interactive multi-view video streaming. The proposed algorithm separated users with the same interest to improve streaming performance.

Kurutepe *et al.* introduced the delivery of multi-view video over a multi-tree P2P network [11]. In their study, each camera view was streamed over an independent P2P tree. However, ungraceful peer disconnection and transmission losses significantly affect the performance of the tree-based P2P streaming system. Also, to address the packet loss problem in P2P networks, Zhou *et al.* proposed a real-time multi-view video streaming system [12] that can deliver multi-view video over two separate channels. In their work, a packet processing method was employed to hold correlation between views. As a result, lost packets can be replaced by the neighboring camera view in the delivered stream.

Layer-based video coding is a practical approach to adapting multi-view video content to the available bandwidth [13–16]. For instance, it is possible to discard a portion of the encoded bit-stream and accommodate peers with different bandwidth limitations. Accordingly, Liu *et al.* proposed layer-based video streaming over P2P networks [17]. In that work, various subscriptions were assigned to each user. For example, ordinary peers can download only a base-layer stream to obtain the necessary video playback quality. On the other hand, peers with the premium subscription can download the multiple-layer video to enjoy enhanced video quality. Furthermore, a peer that contributes more upload capacity on P2P networks can receive more layers. However, many digital internet network connections offer less bandwidth for upload than for download. Hence, even if a peer has sufficient download bandwidth to watch the video sequences, it becomes a free rider¹ in the system, which might lead to performance degradation [19]. More recently, the work in [20] integrated broadcast and broadband networks as a hybrid solution for P2P networks. This work contains user-preference-aware multi-view video adaptation, which is a well-defined problem for interactive video streaming concept. With a similar application, Yun *et al.* implemented a DASH-based multi-view video streaming system in [21] to minimize the view-switching delay while providing seamless

¹ A free rider is a peer that uses P2P network services but does not contribute to the network or other peers at an acceptable level [18].

video playback. Differently, in this paper, we focus on the transmission of all available captured camera views with a highest possible quality problem where interactive camera view transmission is not required by servers.

Another adaptation method to maximize the average multi-view video streaming quality is to discard some camera views, *i.e.*, view scaling, for transmission based on the available bandwidth rate. For this purpose, Gurler *et al.* [22] presented an adaptive delivery mechanism of scalable multi-view video. They proposed an adaptation engine determine whether to omit some camera views to be transmitted or decrease their quality based on the available network rate. Similarly, research works in [23] and [24] proposed QoE-aware adaptation strategies for multi-view video over P2P networks. In these studies, symmetric and asymmetric rate scaling methods were evaluated to achieve high 3D perceptual quality. In particular, research in [22] compared different bitrate scaling methods against the view scaling approach. Experiment results found that using the asymmetric rate scaling (*i.e.*, some camera views are scaled to a lower rate than other) provides the best QoE for moderate bitrate reduction. Also, view scaling adaptation provides the best QoE for more aggressive cuts in the multi-view video bitrate. However, this view scaling approach might reduce multi-view video streaming quality because of 3D view synthesis artifacts. For this reason, in this paper, we utilize our previous view scalable multi-view video coding approach in [6] that can reduce view synthesis artifacts.

A recent work in [25] defined the selective multi-view video streaming as an optimization problem by allocating bits of the transmission of MVD for the selected camera views. The algorithm selects the optimum inter-view prediction structures along with the quantization parameters (QPs) under given storage and network bandwidth constraints. In [26], an optimization was formulated to select the optimal set of camera views that client requests for video downloading. The system optimizes the navigation quality experienced by the client using the bandwidth constraint. Also, the transmission of auxiliary information was studied in [27] to reduce the complexity of interactive multi-view video decoders. In the work, a residual frame was encoded at the server side by considering rate-distortion efficiency based on user behavior models. As a difference with these recent studies, in this paper, we focus where all camera views are required to be available with high visual quality at the receiver.

3 System Overview

Figure 1 illustrates the high-level architecture of the proposed MVD delivery scenario. Firstly, multiple cameras capture MVD video sequences with different camera locations. Each captured sequence is then encoded into a collection of transmission chunks that are placed on the data server. The transmission chunk is self-decodable streaming packet which are sent over the mesh-based P2P networks. A peer that experiences network congestion can dispense with receiving some camera views and requesting the SI stream.

Fig. 1 High-level architecture of the proposed MVD delivery scenario using P2P networks [7].

In order to construct mesh-based P2P overlay, each peer first subscribes to the tracker, where information of all peers is recorded in P2P. The tracker periodically delivers the list of online nodes to each peer. The streaming partners (*i.e.*, neighbors) are then chosen randomly using the online peer list to maintain a content-sharing partnership. Hence, a peer can connect to another peer as soon as it detects a link problem.

To achieve reliable MVD streaming performance during network congestions, bitrate of MVD content adapts to dynamic network conditions by incorporating the SI. For this purpose, the SI is requested directly by the peers applying adaptation. **Pseudocode for the proposed algorithm is shown in Algorithm 1.**

Algorithm 1 Proposed adaptation algorithm

```

1: procedure ADAPTATION
   top:
2:    $Bandwidth \leftarrow$  estimated bandwidth
3:    $Strat \leftarrow$  adaptation strategy
4:   if  $bandwidth >$  bitrate for transmission of all camera views then
5:     return Transmit all available camera views
6:   else
7:     return Transmit a subset of camera views with SI
8:   goto top.

```

3.1 Estimation of SI

In order to minimize the quality distortion at times of network congestion, the proposed approach selectively discards some camera views in a pre-determined order and then reconstructs them using the SI stream, which is the comprise of the index values of the codebook and the variable block size structure.

The codebook is designed in the data server for each camera view and then downloaded by all peers at the beginning of streaming (*i.e.*, one-off download), during the start-up buffering period. The SI stream signals the index values of the downloaded codebook which are the corresponding weighting factors for view reconstruction in the downloaded codebook.

Each pixel block for each available camera view's frame is utilized in the estimation procedure of weighting factors. Each camera view is selectively discarded and reconstructed at the data server, considering that it is discarded by a peer, while others continue to be received by that peer. Each b block of the discarded view(s) is reconstructed by the weighted summation of all corresponding blocks from the delivered views using:

$$b_N^t = \sum_{k=1, k \neq N}^M [\tilde{b}_k^t \cdot w_{b_k}] + b_N^{t_{ref}} \cdot w_{b_N}, \quad (1)$$

where N and M represent the discarded camera view and the total number of available camera views, respectively. Each block denotes as b and its projected camera view to N (using DIBR) represents as $\tilde{b}$. Frame number and temporal reference frame are t and t_{ref} , respectively. t_{ref} is the delivered camera view's last frame in the corresponding previous temporal chunk. Weighting factor for each b pixel block represents as w_b .

To minimize the bitrate overhead of coefficients, a codebook is designed for each camera view which contains a set of coefficients. Coefficient index values, which represent the lookup value of the designed codebook, is then embedded into the bitstream as metadata. Also, the proposed approach optimizes the trade-off between the reconstruction quality of missing view(s) and overhead of metadata. Thus, variable block size based quad-tree algorithm [28–31] is used for each frame of a target view, which is the discarded view. The smallest (8×8) and the largest (32×32) possible block sizes are determined in the quadtree. Each block can be split into four equal sizes of blocks starting from the largest block size in a top-down approach. When the overhead for the transmission of the quad-tree structure is included, the overall Lagrangian cost function becomes:

$$L(x) = D(x) + \lambda_1 \cdot l + \lambda_2 \cdot QT(x), \quad (2)$$

where D is the distortion metric, x contains possible sized blocks, $\{8 \times 8, 16 \times 16, 32 \times 32\}$ for b_N^t , λ_1 and λ_2 correspond to the Lagrangian multiplier values for additional metadata and quadtree overhead, respectively. λ_1 and λ_2 are also obtained following subjective training with three different multi-view sequence. $QT(x)$ denotes the quad-tree codeword length (in bits) required for the signaling of the quad-tree structure for x . l is the size of the resulting metadata. In this work, Mean Squared Error (MSE) metric has been utilized as the distortion metric D . It worth to mention that other perceptual quality metrics are equally employable.

To estimate w , the cross-correlation method in [32] and depth image based rendering (DIBR) [33] techniques are utilized. This model recovers the discarded views by estimating weights with the smallest possible pixel error in relation to its uncompressed original representation, as mathematically described in Eq.3:

$$\begin{bmatrix} w_{b_1} \\ \vdots \\ w_{b_M} \\ w_{b_N} \end{bmatrix} \cdot \begin{bmatrix} E[\tilde{b}_1^t \cdot \tilde{b}_1^t] & \cdots & E[\tilde{b}_1^t \cdot \tilde{b}_M^t] & E[\tilde{b}_1^t \cdot b_N^{t_{ref}}] \\ \vdots & \vdots & \vdots & \vdots \\ E[\tilde{b}_M^t \cdot \tilde{b}_1^t] & \cdots & E[\tilde{b}_M^t \cdot \tilde{b}_M^t] & E[\tilde{b}_M^t \cdot b_N^{t_{ref}}] \\ E[b_N^{t_{ref}} \cdot \tilde{b}_1^t] & \cdots & E[b_N^{t_{ref}} \cdot \tilde{b}_M^t] & E[b_N^{t_{ref}} \cdot b_N^{t_{ref}}] \end{bmatrix} = \begin{bmatrix} E[\tilde{b}_1^t \cdot b_N^t] \\ \vdots \\ E[\tilde{b}_M^t \cdot b_N^t] \\ E[b_N^{t_{ref}} \cdot b_N^t] \end{bmatrix} \quad (3)$$

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

The K-means clustering algorithm is then applied to design codebook, which contains weighting factors. Each weights (w) forms $M + 1$ dimension vectors in the codebook, which is encoded using an l -bit codeword. To obtain the optimum coefficient vectors, each vector's reconstruction quality is compared with all candidate vectors in the codebook. Then, the candidate that results in the highest-quality index value (i) is selected. The index value of each computed coefficient vector that corresponds to each computed block is embedded as side information.

3.2 Adaptive Streaming Strategy

To distribute MVD, the pull-based mechanism is adopted, and a disjoint buffer is designed for each camera view at each peer. After a given period, called the playback delay, peers start decoding the MVD streams from the buffer to play them on the multi-view display. At periodic request intervals, peers exchange chunk availability information with the help of buffer maps. The packet used for signaling contains a buffer map message, request chunks, overlay construction, and adaptation data. A buffer map indicates the chunks that a peer has currently buffered and can share. Each node explicitly requests the missing chunks for each camera view in the request window, which slides forward simultaneously in time.

The primary challenge is how to use the adaptation approach efficiently within the congested network. There are two possible strategies to the outlined system architecture. First, it is possible symmetrically to adapt the bitrate of all streamed camera views to meet the network bandwidth [22]. In this case, all views need to be encoded at several quality levels in the data server. However, this approach leads to a reduction in the reconstruction quality of all received views, thus resulting in a decrease in the view synthesis quality as well [34]. Second, it is possible to transmit the selected camera views through content-aware bandwidth adaptation and allow the missing, *i.e.*, discarded, camera view(s) to be reconstructed at the receiver side. Unlike the first strategy, this approach does not compromise the delivery quality of several views. However, its main limitation is the distortion of the discarded views. It is difficult to achieve an exceptionally high quality of reconstructed pictures caused by errors at object edges and typical occlusion holes.

The proposed system in this work is based on the second strategy, which aims at allowing the receiver to reconstruct all required views at the highest possible quality at all time. This approach uses low-overhead metadata when necessary, allows the discarded views to be estimated using the proposed algorithm. In contrast to the described first strategy, this method does not compromise the perceptual quality of all encoded views. However, it leads to maintains the highest possible quality at the receiver side, giving the extraction and usage of optimized SI along with the delivered high-quality views and depth maps. Also, this approach achieves high chunk distribution performance in the P2P network because the same stream versions are stored on multiple user devices.

An adaptation strategy, *i.e.*, view discarding pattern, is first estimated according to the network bandwidth at the server. The number of views to be discarded is assessed through an evaluation that can be formulated as an optimization problem. The total bitrate requirement of encoded MVD sequences is R_S , and the overall MVD distortion is $Dis(S)$. Consequently to minimize $Dis(S)$ and discard optimal view(s) to meet network bandwidth (r_{max}), the optimization function, $f(\cdot)$, is described as in Equation (4):

$$f(Dis_1, \dots, Dis_M) = \min \|Dis_1 + \dots + Dis_M\| \quad \forall (R_S + R_{SI}) < r_{max}, \quad (4)$$

where, $Dis_1 + \dots + Dis_M$ is the overall distortion of all views, expressed as $Dis(S)$. Furthermore, R_{SI} is denoted as the additional metadata overhead for the proposed adaptation.

To determine the subset of views to delivered during congestion periods, priority information is derived from priority classification. This process aims to minimize the overall reconstruction distortion of discarded chunks subject to a limited bitrate budget. The priority estimation is described in Equation (5):

$$p_k = \frac{\varpi_k}{\sum_{k=1}^M \varpi_k}, \quad (5)$$

where p_k is the view prioritization, k belongs to a discrete set of views between (and including) views one and M . Furthermore, ϖ_k is expressed as the weighted priority of each camera view. **Each weighted priority is computed as follows.**

$$\varpi_k = \frac{Dis_k}{\max(\{Dis\}_{k=1}^M)}, \quad (6)$$

Depending on the measured bandwidth capacity, lower priority views are intelligently discarded to be recovered at the receiver side. In the most severe condition, in order to estimate all views within the total baseline of the MVD set, the edge views (*i.e.*, see base views in Figure 2) and their associated depth maps are delivered.

Fig. 2 Camera arrangement. Potentially discardable intermediate views are marked in grey, whereas the base views (not discardable) are marked in black.

Following this process, the view discarding pattern is transmitted to each peer. Accordingly, peers can ultimately select the most suitable view(s) to be discarded according to their network conditions. For this purpose, it is necessary to monitor the peer buffer regularly to manage the proposed adaptation method. To this end, occupancy of the playback buffer is the key indicator for accomplishing the described adaptation strategy. For instance, high buffer occupancy indicates that the download rate is greater than the MVD streaming rate, considering that a small size of buffer occupancy means that the link capacity is not sufficient to support MVD streams.

Fig. 3 Relationship between buffer occupancy and buffer size for a sequence.

Figure 3 shows the relationship between buffer occupancy and buffer size, which is measured in each view. According to the size of buffer occupancy, the adaptation engine can decide when to perform the proposed adaptation approach (*i.e.*, view discarding). For instance, if the size of buffer occupancy is lower than a chosen threshold, Υ_b , this indicates that the peer has inadequate bandwidth resources to stream MVD at the current adaptation set. Hence, the

adaptation set is changed, and a subset of views is requested along with the low-overhead metadata, *i.e.*, SI) stream.

In order to change the adaptation set, the P2P engine might discard (or not request) some of the predefined views. The discarded views are estimated at the receiver side using the delivered pictures and the SI. Similarly, if the size of buffer occupancy increases beyond a certain threshold ($\mathcal{T}_b$ in *see* Figure 3), all encoded MVD sequences can be provided without employing view discarding strategy.

4 Experimental results

4.1 Setup

The performance of the proposed adaptive streaming method is evaluated over the the explained P2P overlay. Five neighbouring camera views ($M=5$) from MVD test sequences were encoded and delivered to multiple peers. Six MVD video sequences were chosen from the MPEG test sequences [35]. These sequences are *BookArrival*, *Newspaper*, *Pantomime*, *ChampagneTower*, *Café*, and *PoznanStreet*. To encode each MVD video sequences, the HEVC standard was adopted by selecting QPs to be 20, 26, 32, 38, and 44. In addition, the bitrate for each depth map was fixed at a percentage of 20% of its colour texture bitrate. The size of group of picture (GOP) was set to 16; a single GOP was inserted into each chunk.

In order to evaluate the performance of the proposed adaptation approach further, *MPEG VSRS* software was incorporated as a reference adaptation scheme. This method utilizes the two nearest left and right adjacent camera views of the target camera view to synthesizing the target camera view's color texture view. In our simulations, the performance of the proposed adaptation approach was evaluated experimentally, considering the use of 8-bit codebooks for view estimation purposes. The overhead of additional metadata, *i.e.*, SI, was included for all presented results.

4.2 Adaptive MVV Streaming Performance

Multiple experiments using the parameter settings in Table 1 were performed to investigate the performance of the proposed adaptation method. The test results over different time segments are very similar. To avoid verbosity, only one peer was selected, and its adaptation performance was plotted per content throughout this study.

Figures 4 - 9 demonstrate the overall adaptive P2P streaming quality over time with standard deviation σ for six different MVD video sequences. Deviation of PSNR is represented by σ . The simulation experiments clearly present that the proposed adaptation system consistently outperforms the reference methods, which are *MPEG VSRS adaptation* and *HEVC without adaptation*,

Fig. 4 *BookArrival*: average adaptive P2P streaming quality over time. Video streaming rate is 5.64 Mbps and the peer download/upload capacity is 6 Mbps.

Fig. 5 *Newspaper*: average adaptive P2P streaming quality over time. MVD streaming rate is 3.74 mbps and the peer download/upload capacity is 4.8 mbps.

Fig. 6 *Pantomime*: average adaptive P2P streaming quality over time. MVD streaming rate is 11.52 mbps and the peer download/upload capacity is 12 mbps.

Fig. 7 *ChampagneTower*: average adaptive P2P streaming quality over time. MVD streaming rate is 6.23 mbps and the peer download/upload capacity is 7 mbps.

Fig. 8 *Café*: average adaptive P2P streaming quality over time. MVD streaming rate is 6.79 *mbps* and the peer download/upload capacity is 7 *mbps*.

Fig. 9 *PoznanStreet*: average adaptive P2P streaming quality over time. Video streaming rate is 5.84 *mbps* and the peer download/upload capacity is 6 *mbps*.

Table 1 Configuration of the proposed adaptive P2P streaming simulator.

MVV content	Content bitrate /mbps	Download/Upload capacity /mbps
<i>BookArrival</i>	5.64	6
<i>Newspaper</i>	3.74	4.8
<i>Pantomime</i>	11.52	12
<i>ChampagneTower</i>	6.23	7
<i>Café</i>	6.79	7
<i>PoznanStreet</i>	5.84	6

at times of network congestion. Moreover, the average PSNR gain with respect to both references (with and without adaptation) is depicted in Table 2 along with the additional overhead that is increased proportionally to the video resolution size as well as the size of the congestion rate. As indicated in Table 2, the proposed adaptation method requests high amount of metadata size in the *BookArrival*, *Café*, and *PoznanStreet* sequences because of their higher resolution video sizes and amount of congestion experiences on the network.

Table 2 Average PSNR gain compared to the reference methods and incurred overhead because of the additional metadata.

MVV sequence	Compare Proposed to		Proposed
	HEVC (in dB)	MPEG VRSR (in dB)	Total overhead kbps
<i>BookArrival</i>	1.89	0.16	163.04
<i>Newspaper</i>	0.15	0.03	12.33
<i>Pantomime</i>	2.30	0.18	32.1
<i>ChampagneTower</i>	0.73	0.18	22.52
<i>Café</i>	1.82	0.27	274.47
<i>PoznanStreet</i>	1.09	0.08	223.87

Tables 3-8 present P2P streaming simulation results for each MVD content. In this experiment, the adaptation ratio is expressed as the total number of adapted chunks to the total number of successfully delivered chunks. Test results present that as the number of peers rises in P2P streaming, the adaptation ratio increases because of network congestion. The reason is that a large number of users in P2P networks lead resource reduction to distribute data on time. This problem can be solved by increasing the link capacity of servers and peers. Eventually, chunk delay accumulates hop by hop and exceeds the waiting timeout that drops the chunks, and subsequently, the throughput.

As seen in Table 4, the *Newspaper* sequence demonstrates low adaptation ratio and slight chunk delay. The reason is that it has a reasonable streaming rate (3.74 Mbps) to distribute streaming chunks successfully to each peer. There are simply sufficient resources available (50 Mbps data server plus 4.8 Mbps upload rate for each peer) to receive streaming content on time. On the other hand, the *Pantomime* sequence, which has 11.52 Mbps streaming rate, shows the highest adaptation ratio among all tested sequences. Without

employing adaptation methods, this congestion problem can be solved using a higher link capacity of nodes. However, as illustrated in Figure 6, the *proposed adaptation* method achieves higher transmission in terms of total usage link capacity. In addition, smoother playback is observed in terms of PSNR variation over time.

Table 3 P2P simulation results for the transmission of the 5.64 *mbps BookArrival* sequences.

# of peers	Avg. downloading rate / <i>mbps</i>	Avg. uploading rate / <i>mbps</i>	Avg. chunk delay / <i>sec.</i>	Hop count	Adaptation ratio /%
250	5.05	5.84	10.6	3.6	14.7
150	4.92	5.62	8.77	3.26	13.8
50	5.2	5.6	7	2.58	9
σ	0.14	0.13	1.8	0.52	3.06

Table 4 P2P simulation results for the transmission of the 3.74 *Mbps Newspaper* sequences.

# of peers	Avg. downloading rate / <i>mbps</i>	Avg. uploading rate / <i>mbps</i>	Avg. chunk delay / <i>sec.</i>	Hop count	Adaptation ratio /%
250	4.02	4.47	6.01	3.23	0.7
150	3.99	4.26	3.93	2.79	0.02
50	3.84	3.36	3.09	2.02	0
σ	0.09	5.59	1.50	0.61	0.39

Table 5 P2P simulation results for the transmission of the 11.52 *Mbps Pantomime* sequences.

# of peers	Avg. downloading rate / <i>mbps</i>	Avg. uploading rate / <i>mbps</i>	Avg. chunk delay / <i>sec.</i>	Hop count	Adaptation ratio /%
250	7.64	8.07	9.23	3.29	34.15
150	8.38	9.19	8.37	3.14	29.2
50	9.99	10.92	7.12	2.78	17.4
σ	1.20	1.44	1.06	0.26	8.60

Table 6 P2P simulation results for the transmission of the 6.23 *Mbps ChampagneTower* sequences.

# of peers	Avg. downloading rate / <i>mbps</i>	Avg. uploading rate / <i>mbps</i>	Avg. chunk delay / <i>sec.</i>	Hop count	Adaptation ratio /%
250	6.52	6.73	8.64	3.58	12.5
150	5.65	6.27	6.42	3.01	6.28
50	6.1	6.48	5.28	2.53	8.66
σ	0.43	0.23	1.71	0.53	3.14

Table 7 P2P simulation results for the transmission of the 6.79 Mbps *Café* sequences.

# of peers	Avg. downloading rate /mbps	Avg. uploading rate /mbps	Avg. chunk delay /sec.	Hop count	Adaptation ratio /%
250	5.53	6.46	10.5	3.69	22.05
150	5.64	6.55	10.65	3.41	17.99
50	5.85	6.44	9.68	2.72	15.23
σ	0.16	0.06	0.52	0.49	3.43

Table 8 P2P simulation results for the transmission of the 5.84 mbps *PoznanStreet* sequences.

# of peers	Avg. downloading rate /mbps	Avg. uploading rate /mbps	Avg. chunk delay /sec.	Hop count	Adaptation ratio /%
250	4.98	5.68	10.25	3.54	15.2
150	4.94	5.67	9.65	3.27	13.3
50	5.29	5.62	6.68	2.53	7.87
σ	0.19	0.03	1.91	0.52	3.80

5 Conclusion

This paper introduced a novel adaptive multi-view-plus-depth map (MVD) video streaming over peer-to-peer (P2P) networks for delivery of next-generation 3D multi-view video services. The proposed approach demonstrated better performance under network congestion, and provides consistent quality at a broad range of network conditions. The new approach relies on an adaptive MVD streaming technique, where some views are discarded by affected peers in a pre-determined order. The missing views are then reconstructed using only a small amount of additional metadata requested by the affected peers at times of congestion. The proposed system outperformed the reference methods in terms of overall RD and the performance in adaptation against ungraceful network conditions experienced in P2P systems.

References

1. C. Ozcinar, E. Ekmekcioglu, J. Čalić, and A. Kondo, "Adaptive delivery of immersive 3d multi-view video over the internet," *Multimedia Tools and Applications*, vol. 75, no. 20, pp. 12 431–12 461, 2016.
2. V. Jacobson, "Congestion avoidance and control," in *ACM SIGCOMM computer communication review*, vol. 18. ACM, 1988, pp. 314–329.
3. K. Miller, E. Quacchio, G. Gennari, and A. Wolisz, "Adaptation algorithm for adaptive streaming over HTTP," *2012 19th International Packet Video Workshop (PV)*, pp. 173–178, May 2012. [Online]. Available: <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6229732>
4. W. E. Castellanos, J. C. Guerri, and P. Arce, "Svceval-ra: an evaluation framework for adaptive scalable video streaming," *Multimedia Tools and Applications*, vol. 76, no. 1, pp. 437–461, 2017.
5. C. Ozcinar and G. Anbarjafari, "Dynamic bitrate allocation of interactive real-time streamed multi-view video with view-switch prediction," *Signal, Image and Video Processing*, pp. 1–7, 2017.

6. C. Ozcinar, E. Ekmekcioglu, and A. Kondo, "Dynamic adaptive 3d multi-view video streaming over the internet," in *Proceedings of the 2013 ACM international workshop on Immersive media experiences*. ACM, 2013, pp. 51–56.
7. —, "Adaptive 3d multi-view video streaming over p2p networks," in *Image Processing (ICIP), 2014 IEEE International Conference on*. IEEE, 2014, pp. 2462–2466.
8. I. Sodagar, "The mpeg-dash standard for multimedia streaming over the internet," *IEEE MultiMedia*, vol. 18, no. 4, pp. 62–67, 2011.
9. S.-Y. Hu, T.-H. Huang, S.-C. Chang, W.-L. Sung, J.-R. Jiang, and B.-Y. Chen, "Flod: A framework for peer-to-peer 3d streaming," in *INFOCOM 2008. The 27th Conference on Computer Communications*. IEEE, 2008, pp. 1373–1381.
10. Z. Chen, L. Sun, and S. Yang, "Overcoming view switching dynamic in multi-view video streaming over P2P network," in *3DTV-Conference: The True Vision-Capture, Transmission and Display of 3D Video (3DTV-CON), 2010*. IEEE, 2010, pp. 1–4.
11. E. Kurutepe and T. Sikora, "Feasibility of multi-view video streaming over p2p networks," in *3DTV Conference: The True Vision-Capture, Transmission and Display of 3D Video, 2008*. IEEE, 2008, pp. 157–160.
12. Y. Zhou, C. Hou, Z. Jin, L. Yang, J. Yang, and J. Guo, "Real-time transmission of high-resolution multi-view stereo video over ip networks," in *3DTV Conference: The True Vision-Capture, Transmission and Display of 3D Video, 2009*. IEEE, 2009, pp. 1–4.
13. H. Schwarz, D. Marpe, and 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, 2007.
14. K. Müller, A. Smolic, K. Dix, P. Merkle, and T. Wiegand, "Coding and intermediate view synthesis of multiview video plus depth," in *Image Processing (ICIP), 2009 16th IEEE International Conference on*. IEEE, 2009, pp. 741–744.
15. J. M. Boyce, Y. Ye, J. Chen, and A. K. Ramasubramanian, "Overview of shvc: scalable extensions of the high efficiency video coding standard," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 26, no. 1, pp. 20–34, 2016.
16. E. Erçelebi *et al.*, "A new teleconference system with a fast technique in hevc coding," *Multimedia Tools and Applications*, vol. 76, no. 2, pp. 1775–1800, 2017.
17. Z. Liu, Y. Shen, K. Ross, S. Panwar, and Y. Wang, "LayerP2P: Using layered video chunks in P2P live streaming," *Multimedia, IEEE Transactions on*, vol. 11, no. 7, pp. 1340–1352, Nov 2009.
18. M. Karakaya, I. Korpeoglu, and Ö. Ulusoy, "Free riding in peer-to-peer networks," *IEEE Internet computing*, vol. 13, no. 2, pp. 92–98, 2009.
19. A. Habib and J. Chuang, "Service differentiated peer selection: an incentive mechanism for peer-to-peer media streaming," *IEEE Transactions on Multimedia*, vol. 8, no. 3, pp. 610–621, 2006.
20. E. Ekmekcioglu, C. G. Gurler, A. Kondo, and A. M. Tekalp, "Adaptive multiview video delivery using hybrid networking," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 27, no. 6, pp. 1313–1325, June 2017.
21. D. Yun and K. Chung, "Dash-based multi-view video streaming system," *IEEE Transactions on Circuits and Systems for Video Technology*, pp. 1–1, 2017.
22. C. G. Gurler, B. Gorkemli, G. Saygili, and A. M. Tekalp, "Flexible transport of 3-d video over networks," *Proceedings of the IEEE*, vol. 99, no. 4, pp. 694–707, 2011.
23. C. G. Gurler, S. S. Savas, and A. M. Tekalp, "Quality of experience aware adaptation strategies for multi-view video over p2p networks," in *Image Processing (ICIP), 2012 19th IEEE International Conference on*. IEEE, 2012, pp. 2289–2292.
24. C. G. Gurler and M. Tekalp, "Peer-to-peer system design for adaptive 3d video streaming," *IEEE Communications Magazine*, vol. 51, no. 5, pp. 108–114, 2013.
25. A. De Abreu, P. Frossard, and F. Pereira, "Optimizing multiview video plus depth prediction structures for interactive multiview video streaming," *IEEE Journal of Selected Topics in Signal Processing*, vol. 9, no. 3, pp. 487–500, Apr. 2015.
26. X. Zhang, L. Toni, P. Frossard, Y. Zhao, and C. Lin, "Adaptive streaming in interactive multiview video systems," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. PP, no. 99, pp. 1–1, 2018.

27. T. Maugey and P. Frossard, "Interactive multiview video system with low complexity 2d look around at decoder," *IEEE Transactions on Multimedia*, vol. 15, no. 5, pp. 1070–1082, Aug 2013.
28. H. Samet, "The quadtree and related hierarchical data structures," *ACM Computing Surveys (CSUR)*, vol. 16, no. 2, pp. 187–260, 1984.
29. Y. Morvan, D. Farin, and P. H. de With, "Depth-image compression based on an rd optimized quadtree decomposition for the transmission of multiview images," in *Image Processing, 2007. ICIP 2007. IEEE International Conference on*, vol. 5. IEEE, 2007, pp. V–105.
30. P. Merkle, Y. Morvan, A. Smolic, D. Farin, K. Mueller, P. de With, and T. Wiegand, "The effects of multiview depth video compression on multiview rendering," *Signal Processing: Image Communication*, vol. 24, no. 1, pp. 73–88, 2009.
31. P. Merkle, K. Müller, and T. Wiegand, "3d video: acquisition, coding, and display," *IEEE Transactions on Consumer Electronics*, vol. 56, no. 2, pp. 946–950, 2010.
32. Y. Sugiyama, "An algorithm for solving discrete-time Wiener-Hopf equations based upon Euclid's algorithm," *Information Theory, IEEE Transactions on*, vol. 32, no. 3, pp. 394–409, May 1986.
33. C. Fehn, "Depth-image-based rendering (DIBR), compression, and transmission for a new approach on 3D-TV," *Proceeding SPIE*, vol. 5291, pp. 93–104, 2004. [Online]. Available: <http://dx.doi.org/10.1117/12.524762>
34. 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*, Nov 2009, pp. 721–724.
35. M. Tanimoto, T. Senoh, S. Naito, S. Shimizu, H. Horimai, M. Domański, A. Vetro, M. Preda, and K. Mueller, "Proposal on a new activity for the third phase of ftv," in *The 105th meeting of MPEG*, no. MPEG2011/N12036, 2013.

Cagri Ozcinar is a research fellow within the V-SENSE project at Trinity College Dublin, Ireland, since July 2016. Before he joined the V-SENSE team, he was a postdoctoral research fellow in the Multimedia group at Institut Mines-Telecom ParisTech, Paris, France. Cagri received the M.Sc. (Hons.) and the Ph.D. degrees in electronic engineering from the University of Surrey, UK, in 2011 and 2015, respectively. His Ph.D. topic was about multi-view video communication. His current research interests include multi-view, HDR and 360° video compression and streaming techniques. Cagri has been serving as a reviewer for a number of journals and conferences.

Erhan Ekmekcioglu received his PhD degree from the University of Surrey, UK, in 2010, and worked there as a post-doc researcher until 2014. Since 2014 he has been working as a research associate in the Institute for Digital Technologies at Loughborough University London, a post graduate teaching, research and enterprise institute. His research interests include video processing and compression, machine learning applications on image and video, video streaming over constrained networks, mobile video communication, Quality of Experience, immersive and 3D media systems, augmented reality, and virtual reality. He has participated in a number of European Union funded collaborative research projects and conducted research mainly on multi-view video acquisition, compression, distribution, and Quality of Experience. He is the author/co-author of several research articles, book chapters, a book on 3D-TV systems, and the guest editor of several special issues published by the IEEE Multimedia Communications Technical Committee. He has been a regular reviewer of prestigious IEEE Transactions including Circuits and Systems for Video Technology, Multimedia, and Image Processing. Since January 2017 he has been co-chairing the 3D rendering, processing and communications Interest Group (3DIG) of the IEEE Multimedia Communications Technical Committee.

Gholamreza Anbarjafari heads the intelligent computer vision (iCV) research group in the Institute of Technology at the University of Tartu. He is also the supervisor of Philosopher, the Estonian Robocup team of the University of Tartu. He is an IEEE Senior member and the Vice Chair of the Signal Processing / Circuits and Systems / Solid-State Circuits Joint Societies Chapter of the IEEE Estonian section. He received the Estonian Research Council Grant (PUT638) in January 2015 and has been involved in many international industrial projects such as AIDesign, A.G.E., iRental and Virtual Fitting Room by Fits.Me Rakutan. He is expert in computer vision, human-robot interaction, graphical models and artificial intelligence. His work in image super resolution was selected for the best paper award in

2012 by the Electronics and Telecommunications Research Institute (ETRI) Journal, South Korea. He has supervised 7 MSc students and 5 PhD students. He has published over 90 scientific works. He has been in the organizing committee and technical committee of the IEEE Signal Processing and Communications Applications Conference in 2013, 2014 and 2016 and the TCP of conferences such as ICOSST, ICGIP, and SIU. He is organizing a challenge and a workshop on in FG17, CVPR17, and ICCV17.

Ahmet Kondo (M'91, SM'11) received his PhD degree in 1987 from University of Surrey, UK. From 1986 to 1988, he was employed as a research fellow in the communication systems research group. He became a lecturer in 1988, reader in 1995, and in 1996 he was promoted to professor in multimedia communication systems. He was the founding head of I-LAB, a multi-disciplinary multimedia communication systems research lab at the University of Surrey. Since January 2014, he has been appointed as the founding Director of the Institute for Digital Technologies, at Loughborough University London, a post graduate teaching, research and enterprise institute. He is also serving as the Associate Dean for Research at Loughborough University London. His research interests include digital signal processing and coding, fixed and mobile multimedia communication systems, 3D immersive media applications for the future Internet systems, smart systems such as autonomous vehicles and assistive technologies, big data analytics and visualisation and related cyber security systems. He has over 400 publications, including six books, several book chapters, and seven patents, and graduated more than 75 PhD students. He has been a consultant for major wireless media industries and has been acting as an advisor for various international governmental departments, research councils and patent attorneys. Dr. Kondo has been involved with several European Commission FP6 & FP7 research and development projects, such as NEWCOM, e-SENSE, SUIT, VISNET, MUSCADE, etc. involving leading universities, research institutes and industrial organisations across Europe. He coordinated FP6 VISNET II NoE, FP7 DIOMEDES STREP and ROMEO IP projects, involving many leading organisations across Europe which deals with the hybrid delivery of high quality 3D immersive media to remote collaborating users including those with mobile terminals. He co-chaired the European networked media advisory task force, and contributed to the Future Media and 3D Internet activities to support the European Commission in the FP7 programmes.

Cagri Ozcinar

Erhan Ekmekcioglu

Gholamreza Anbarjafari

Ahmet Kondoz