Multimed Tools Appl (2018) 77:22965–22983
https://doi.org/10.1007/s11042-018-5744-8

User-perceived quality aware adaptive streaming of 3D multi-view video plus depth over the internet

Nabin Kumar Karn[1] · **Hongli Zhang**[1] · **Feng Jiang**[1]

Received: 31 July 2017 / Revised: 26 January 2018 / Accepted: 30 January 2018 /
Published online: 12 February 2018

Abstract Video streaming is a foremost and growing contributor in the ever increasing Internet traffic. Since last two decades, due to the enhancement in cameras and image processing technology, we have seen a shift towards multi-view plus depth (MVD) technology from traditional 2D and 3D video technology. This growth comes with deep changes in the Internet bandwidth, video coding and network technologies, which smoothed the mode for delivery of MVD content to end-users over the Internet. Since, MVD contains large amounts of data than single view video, it requires more bandwidth. It is a challenging task for network service provider to deliver such views with the best user's Quality of Experience(QoE) in dynamic network condition. Also, Internet is known to be prone to packet loss, bandwidth variation, delay and network congestion, which may prevent video packets from being delivered on time. Besides that, different capabilities of end user's devices in terms of computing power, display, and access link capacity are other challenges. As consequences, the viewing experiences of 3D videos may well degrade, if the quality-aware adaptation techniques are not deployed. In this article, our work concentrates to present a comprehensive analysis of a dynamic network environment for streaming of 3D MVD over Internet (HTTP). We analyzed the effect of different adaptation of decision strategies and formulated a new quality-aware adaptation technique. The proposed technique is promoting from layer based video coding in terms of transmitted views scalability. The results of MVD streaming experiment, using the proposed approach have shown that the video quality of perceptual 3D improves significantly, as an effect of proposed quality aware adaptation even in adverse network conditions.

Keywords 3D multi-view video plus depth · Adaptive streaming · QoE · MPEG-DASH · Dynamic network environment

✉ Nabin Kumar Karn
karnnabin@hit.edu.cn

Hongli Zhang
zhanghongli@hit.edu.cn

Feng Jiang
fjiang@hit.edu.cn

[1] School of Computer Science and Technology, Harbin Institute of Technology, Harbin, China

1 Introduction

Since the beginning of the era of the Internet based video delivery, the advancement of Internet technologies as well as the meticulous systems for ensuring high quality, high demands of video distribution has lead evolution of the video encoding, distribution and delivery technologies through numerous stages. Video transmitting over Internet is attracting more popularity and engages a significant part of the Internet traffic. Due to the increased demand of Internet video, adoption of the heterogeneous network are on rise, which increases complexity and new challenges to video streaming.

Earlier, two cameras were used to achieve the 3D video representation. User could view the 3D picture by wearing polarized spectacles [3]. Recent improvements in video and networked delivery have made it possible to stream 3D stereoscopic(3DS) video both in theaters and in home entertainment. However, fixed conditions of capturing the 3DS video content provides the user with low flexibility. Development in image processing technology and cameras forced a paradigm shift from conventional video technology to multi-view video technology and evaluation of image emotions by different features [49, 50]. In Multi view video (MVV) technology, video sequences are captured by cameras in predetermined position and angels, and in parallel, captures more than two views of the same scene from distinct perspectives. Contents of MVV video sequences are fixed. Different ways of capturing video contents are shown in Fig. 1. Before transmitting, content is compressed in an appropriate way so that user's viewing machine can easily access the relevant views to interpolate new views. Users are freely permitted to change their viewpoints in multi-view representation. At present, multi-view video coding (MVC) for multi-view video or free-view point video (FVV) is accepted as standard video coding, which is an extension of H.264/MPEG-4 advance video coding (AVC). Several applications such as stereoscopic 3D video, free-view point video and auto-stereoscopic 3D video used MVC technology [5]. To retrieve data efficiently and effectively

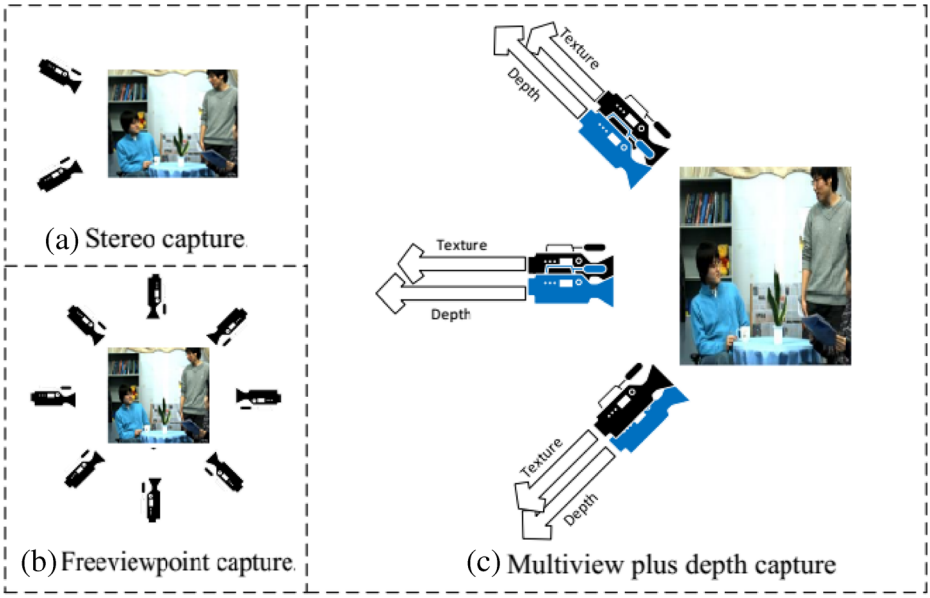


Fig. 1 Different video capturing techniques

that are shared in the Internet in an urgent need, 3D depth data matching strategy is applied before transmitting [45, 46].

Furthermore, FTV(free-view point television) enables users to view a 3D scene by freely changing the viewpoint and generate new views of dynamic scene from any 3D position [38]. Recently developed technology known as multi-view plus depth(MVD) encodes a depth signal from each camera [19]. MVD allows synthesizing virtual views at the client via the depth signals. The signals generated by MVD are sent along MVC. By using depth image-based rendering (DIBR), several virtual views are rendered based on the few real views and their associated depth map [15]. It also provides predicting personalized emotion perceptions of social images as well [47, 48]. Due to the increment in number of captured views of the same scene from different view point, Quality of Experiences(QoE) is highly dependent on the number of views existing at the receiver to render the required virtual view points [37].

While MVV and MVD technologies are very promising, heavy congestion occurs in network leading to a network collapse, if non-adaptive transmission techniques are utilized during MVV streaming [14]. The delivery of MVV remains challenging task due to large number of views, no matter the use of the state-of-the-art video coding standard. The first major challenge is, MVV traffic requires more bandwidth to transmit, because it captures multiple video sequences by multiple cameras in comparison than traditional video. Due to bandwidth variation, packet loss, and delay in Internet, transmission of MVV data in such network lag is more challenging. Secondly, capabilities of end user's devices are different in terms of computing power, display, and access link capacity. This needs an adaptive system to regulate the difference of video characteristic while traversing the network's path.

Due to the heterogeneity of today's communication networks, adaptivity is the most important requirement for any streaming client. Adaptive streaming [6] is a concept to flow video with throughput available on Internet after adapting the bandwidth, which required passing the video from server to end-user [18]. End-user can get reliable and high-quality 3D views in situation where bandwidth is unstable with the help of adaptive streaming systems. Also, detecting events from massive social media data in social networks can facilitate browsing, search, and monitoring of real-time events by corporations, governments, and users [9, 51]. To support HTTP streaming, several commercial applications have been built, such as Apple HTTP Live Streaming [2], Microsoft Smooth Streaming [41] and ISO/IEC MPGE Dynamic Adaptive Streaming over HTTP (MPEG-DASH) [33, 35].

Adaptive video streaming using HTTP has increasingly been gaining attention these days. It is a strong service against those who use Real-time transport protocol (RTP) service. RTP works fine in managed IP network. Nonetheless, in today's Internet structure, content delivery networks replaced the managed networks, and many of them don't support RTP streaming. HTTP is based on TCP service where as RTP uses UDP service. Use of TCP as the dominant transport protocol is due to presence of network address translators(NATs). Also, it is firewall friendly because almost all firewalls are configured to support its outgoing connections [34]. TCP prevents the incomplete packet delivery by triggering naively the packet retransmission. Most of multimedia applications are time sensitive. So, if the packets do not arrive within time period, they are called lost, even though packet re-transmission is permitted.

In this proposed work, we looked user perceived quality aware adaptation way out to increase smooth MVD playback quality. End-user can up/down the 3D visual quality in this manner. We adopted MPEG-DASH standard [33, 35], and high efficiency video coding (HEVC) [19, 36] for our proposed method, nonetheless, our contribution is as follows:

1. We proposed a new layer-based quality, with the view scalability technique to create multiple layers of video for Internet consumer to address the issue of varying network bandwidth. View scalability reduces the number of views transmitted to support lower bandwidth.
2. Our evaluation is extended using more and diverse MVV sequences and accomplished formal subjective testing operation according to the ITU-T BT.500–13 reference for laboratory environment [13].

The compression efficiency of several generations of video coding standards are compared [20]. High efficiency video coding provides approximate 50% bitrate reduction compared to H.264 for same video quality. HEVC was introduced by Video Coding Experts Group of ITU-T and ISO/IEC Moving Picture Experts Group(MPEG). In addition, HEVC 3D extension provides flexibility for different user by generating proper bit stream format [8]. Video contents are divided in equal length and stored in a server in MPEG-Dynamic Adaptive Streaming over the HTTP(DASH)system [1, 33]. Furthermore, to maintain the different qualities and resolutions, video segment copies are encoded with different bitrates as shown in Fig. 2.

In MPEG-DASH system, media presentation description(MPD) file which is an XML, are available to access all segments that are stored in server. Users are responsible to manage all streaming sessions, that means, based on network condition end-user chooses the bitrate.

The rest of this paper is organized as follows: Section 2 provides a brief overview of background and related works. Section 3 presents a detailed description of the proposed adaptation system. Section 4 presents the simulation, objective and subjective test results of proposed user-perceived quality aware adaptation. Concluding remarks presented in Section 5.

2 Background and related work

In this section, we first review the background of techniques and then related work regarding the quality of experience aware adaptation strategies.

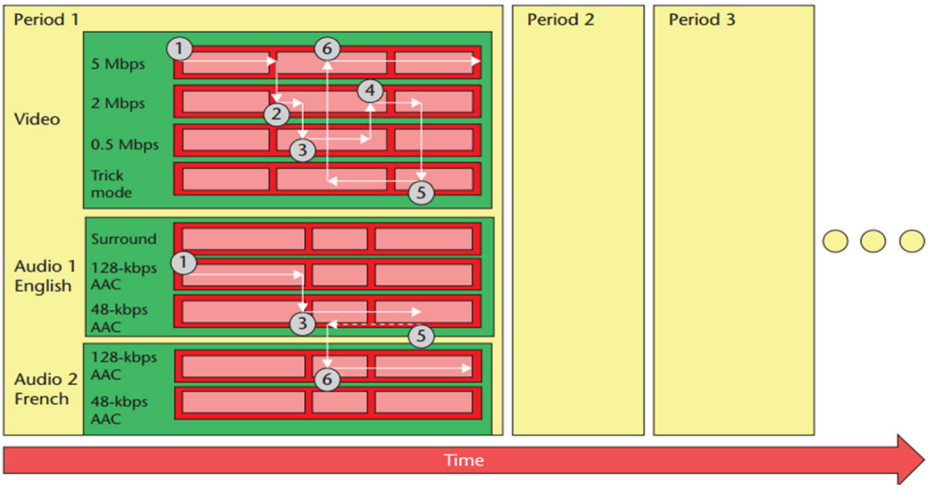


Fig. 2 Simple example of dynamic adaptive streaming. Numbered circles demonstrate the action points taken by the device

2.1 Background

2.1.1 3D multi-view video coding system

Coding of data intensive 3D multi-view video is challenging. There are three main categories of 3D video coding, which are explained below:

H.264/MPEG-4 Advanced Video Coding (AVC) -This standard has been established jointly by the ISO/IEC MPEG and ITU-T video coding experts group (VCEG) [44]. This coding standard reported approximately 50% bit-rate saving compared to earlier MPEG-2 standard for same perceptual quality. In this coding, each view encode independently before 3D multi-view video transmission, which is also called AVC simulcast coding.

Multi-view video coding (MVC) -Many views are transmitted to the recipient in free-view point video system. Due to the increment in number of views and characterizing the same views from varying perspectives, there will be an increment in inter-view statistical dependency. Inter-view prediction method helps to minimize the redundant information between views. MVC is an extension of H.264/MPEG-4 Advance video coding (AVC) [44], which takes the advantage of inter-view statistical dependencies in order to increase the efficiency of 3D video coding. MVC usually provides higher compression efficiency compared to AVC simulcast coding. Experimental results show that compared to AVC, MVC saves approximately 50% bitrate over a wide range of objective quality levels [12, 43]. However, in some experiments, visual quality increment is reported as marginal for some content [31]. Introduction of HEVC, a new video coding technology, presents the same subjective video quality as the AVC standard while bitrate requisite is roughly 50% on average [36].

High efficiency video coding (HEVC)-ITU-T video coding experts group and the ISO/IEC moving picture experts group launched a hybrid video codec design known as HEVC. HEVC has following improvements over H.264/AVC [43, 44] and MPEG-2 [24].

- Coding tree units and coding tree block (CTB) structure: HEVC supports a partitioning of the CTBs into smaller blocks using a tree structure and quad tree structure. Coding unit contains prediction units, which is used for inter-and intra-prediction. Prediction unit further processed the transformation like DCT and quantization [28].
- Coding tree unit consists of a luma and chroma coding tree block (CTB), which are analogous to Macroblock in H.264/AVC [36], but are larger in size. Coding tree block size can be up to 64×64 compared to H.264 and MPEG-2 Macroblock of size 16×16. The larger size of coding tree unit produces better compression performance for high definition video.
- Parallelization design: HEVC introduces tiles for Parallelization design. This design enhances the speed of codec to tackle improvement issues of the computation complexity.
- Support for 3D extension: Besides supporting video compression for 2D high definition video and 2 K video, HEVC supports the views plus depth format of 3D multi-view video [19].

2.1.2 MPEG dynamic adaptive HTTP streaming

HTTP by using the transmission control protocol(TCP) and Internet protocol (IP) plays an important role for delivery of video streaming [16]. HTTP streaming offers the facility of reusing the existing Internet infrastructure as well as better scalability and cost effectiveness to

the service provider. Packet losses are prevented by using TCP's packet re-transmission property in HTTP streaming. Earlier approaches used by HTTP streaming was progressive download, which allows end-users to play out online video contents [1]. Nonetheless, this did not maintain the real aspect of streaming, like quality of video and resolution in dynamic network conditions, because abrupt changes in network conditions may cause users to experience network freeze and interruption.

To overcome the limitations of progressive download, adaptive streaming has been proposed by streaming providers to use the available network resources efficiently and provide high-quality video viewing experience. Highest video quality is guaranteed in adaptive streaming system without long buffering [29].

To reduce the overall transmission delay, HTTP streaming provides a facility to end-users to request different video segment content [35] and adapts to the best video quality based on viewer experiences on dynamic network conditions [29]. Several adaptive streaming techniques have been developed; most of them provide the best video play back quality and minimize start-up delay [29]. MPEG-DASH is one of the adaptive streaming techniques, in which each video is encoded and compressed into a variety of video bitrates corresponding to different resolutions and qualities. Most suitable bitrate [7] or video quality [52] is selected to play according to the bandwidth available or client's hardware specifications. These compressed versions represent the video being fragmented into several segments and then stored in common web servers of organization. After that, it generates the XML-based file call media representation description (MPD). The server then sends the MPD to client to determine the available representations and corresponding URLs. DASH client can individually start to play the video by asking MPD file using HTTP GET request, since DASH is a pull-based method. After receiving the MPD file, the client intially parses it, then sends request for fetching and downloading the suitable segment based on the condition of network, available bandwidth, and buffering.

2.1.3 Video quality assessment in 3D multi-view video plus depth

The main objective of video quality assessment is to approximate the viewer's understanding and satisfaction over a video. There are two ways to do quality assessment viz.: subjective quality assessment and objective quality assessment. In subjective quality assessment, viewer explicitly scores a sequence according to its perceived quality. It gives a more approximation of a user's experiences. Although, conducting the subjective test practically in real time in all video applications is difficult. Thus, objective video quality evaluation methods are employed to estimate the quality of video by taking into account of mathematical model which estimate the subjective quality assessment results.

The objective quality assessment metrics are categorized in three major groups. Objective quality assessment's metrics for different scalable modalities in 3D multi-view video are presented in [26]. Subjective tests were conducted to see the significance of the number of views on the quality of synthesized views for streaming MVD [23], which is close to our work. Results show that acceptable subjective quality can be maintained by decreasing the baseline and the number of transmitted views. Constant quantized parameter(CQP) method was employed to set quantized parameter precisely to encode the different perceptual quality. However, CQP technique is not a part of MPEG-DASH standard [16]. In Contrast with R-Lambda model [20], which used constant bitrate (CBR), it is difficult to ensure the best quality at one particular bitrate given the variation of the bandwidth.

2.2 Related work

The goal of adaptation in video streaming is to match the over all bitrate to the available network rate in a graceful manner, so as to improve user's perceived quality of experience(QoE). QoE in 3D refers to the quality of the received video and includes the depth sense and visual comfort factors in addition to the metrics like fidelity in monoscopic video. To evaluate the quality of experience in 3D MVV, subjective assessment tests method is the most reliable. An adaptive MVV streaming over peer-to-peer network is proposed by Saves [27]. This work deals with whether reducing the quality of all views or keeping a subset of views and synthesize the missing views generates better results. But, quality of visual experience sever artifact on the synthesized view in overall [32]. Toni [42] presented the user satisfaction by picking best possible encoder parameters. Both the compression and spatial scaling artifacts are minimized in this method. Similarly,Cagri [21] proposed a technique to delivery multi-view video along with depth content by employing additional metadata. However, this approach requires extra bandwidth, storage in sever, and processing power. Although the above works achieve good quality of multi-view video streaming, none of them focus on the adaptation of decision strategy and analysis over a dynamic network environment.

3 Proposed system

Figure 3 shows a schematic diagram of the proposed Quality Aware Adaptive streaming of 3D multi-view video over the Internet. The architecture shows the DASH server side and the client side. The main purpose of the proposed architecture is to adaptively transmit the optimized quality of 3D multi-view plus depth map content video over the Internet.

3.1 HTTP server

The HTTP server consists of four different fundamentals: the Media Presentation Description (MPD), the chunks, the LookUp Table (LUT), and the Side Information (SI). The MPD includes the manifest of the adaptation strategy. MPEG-DASH server provides a variety of

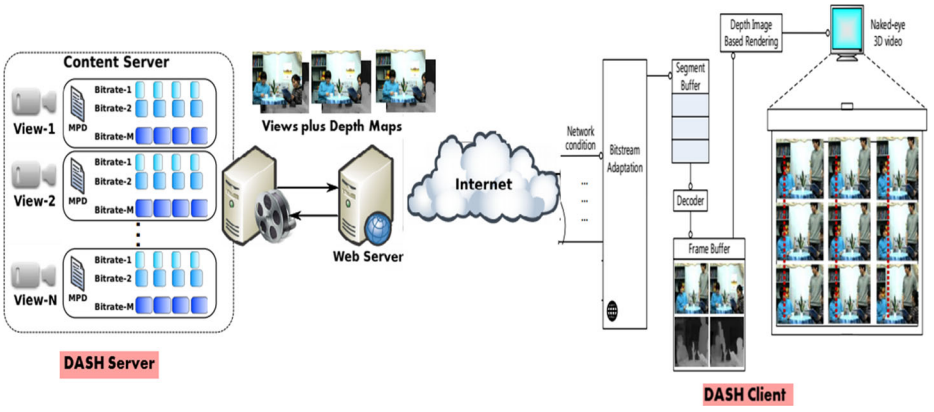


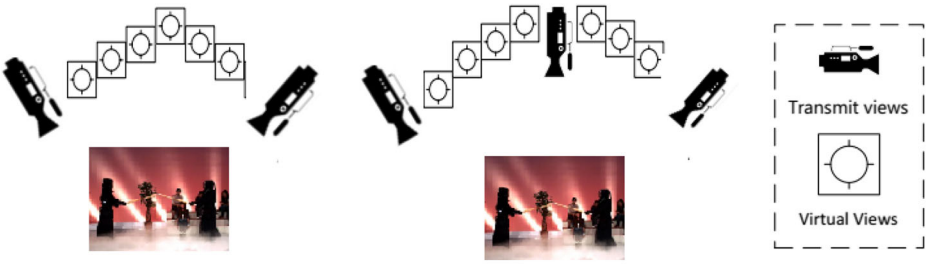
Fig. 3 The Overview of proposed 3D multi-view video transmission system over Internet based on MPEG-DASH standard

video versions according to network condition so that each end-user can adaptively choose the suitable segment of video representation. Server's selection approach depends upon several factors such as congestion of network, availability of the bandwidth, buffer size capacity on the end-user side, and the end-user display resolution. At the time of establishment of the transmission phase, it is not possible to choose the video segment due to stochastic nature of bandwidth and diversity nature of video content over most excellent network. Therefore, in MPEG-DASH system, server breakdowns the whole video into smaller segment temporarily in which each segment ranges from 2 s to 10 s. Each segment then encodes into various bitrates and clusters them into an adjustment set. The final target is to permit the DASH client to switch among various video bitrates according to the offered bandwidth. We use Multi-view video plus depth in our server to characterize the multi-view video. In other words, a view is captured by a series of cameras from different view point and its coupled depth information [3]. In addition, HEVC 3D extension encoder is used in our DASH server encoding engine [8] due to the enhanced compression efficiency over H.264/AVC as shown in [20]. Rate lambda model gives the highest compression quality at the target bitrate in HEVC [17]. Rate lambda model assigns the bit at different levels called Group of Picture (GOP), Picture Level, and Large Coding Unit (LCU) level. We adjust the HEVC 3D extension encoder to encode the various versions of segments in accurate target bit and then store them in the server. The raw video sequences are sliced into small segment with the same duration of time. Then, these small segments are provided to encoding engine to generate the MVD video segment having different bitrates. After that, these segments are kept in server along with MPD.

View scalability is the most important scalable modality in multi-view video, since it has a functionality that enables the receiver to select the number of required views according to its limitations and resource constraints [30]. Our technique uses view scalability along with other scalabilities in order to support receivers whose bandwidths are lower than the total bitrate of the overall MVD video stream. Decreasing the distance between cameras produce more number of views than that of using the larger distance between cameras as shown in Fig. 4. Virtual views generated by using the short distance between cameras produce higher quality. With this technique, according to available bandwidth, clients adapt the video segment with diverse number of views, which has significant effect on the perceived quality of experience.

3.2 HTTP client

HTTP client is responsible to handle all the transmission session when the client picks the video bit stream. The client in DASH sends the HTTP GET request to get MPD file. MPD file



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

Fig. 4 Views Scalability setting

is prepared at server for downloading by the clients before the streaming starts and updating when necessary. The client knows the representation of the content that resides in the server after parsing the MPD file, because all the adaptation parameters are inserted into the MPD file. According to the client bandwidth, the receiver fetches the MPD file and requests optimum video segments. In MPEG-DASH system, client has direct power to manage the streaming session and handles the adjustment of video bitrates in order to cope with the varying network conditions, incoming bandwidth, and status of playback buffer. Client needs to frequently switch among different versions of the temporal segments due to adaptation algorithm. In Fig. 3, buffer segment is provided for safe margin to prevent in case network bandwidth decreases abruptly [10] . This buffer also helps to predict the available bandwidth along with throughput of the downloaded segment.

3.2.1 Selection of bitstream

In this proposed system, bitstream is selected based on the current network conditions to maximize the perceptual quality of rendered views. The choice of bitstream among available bandwidth solely depends upon the client. In our simulation section, we will show the objective measurement test result. This simulation also presents the effect of different number of transmitted views at the same total bitrate on the final perceptual quality of rendered views and transmitted views. These results help to set a policy further to switch the bitrate selection from different number of views. Obviously, the generated total bitrate for bitstream should be lower than the available bandwidth. In addition, we suppose that each of the views share the total bitrate equally. So, the total bitrate used in MVD can be stated as

$$\text{TotalBitrate} = \text{Bitrate}_{\text{view}} \text{x} i \tag{1}$$

where,

$\text{Bitrate}_{\text{view}}$ = The bitrate per views.
Totalbitrate = Total bitrate of the bitstream.
i = Number of transmitted views.

Conventional DASH player chooses the most suitable video bitrate among the offered representation for adaptation. It selects that total video bitrate which is slightly less than available network bandwidth. It is significant to mention that our proposed technique complements conventional methods. In this technique, we used conventional technique to choose the largest representations that are less than the available network bandwidth. In MVD, it is promising to have the same video stream in terms of bitrate with various numbers of views. So, we use Eq. 2 to select the more appropriate video segment according to corresponding computed structure similarity index (SSIM).

$$\text{TotalBitrate} = \begin{cases} Bitrate_{view} \times \ i & \text{if AvgSSIM}(i) > \text{AvgSSIM}(i-1) \\ Bitrate_{view} \times \ (i-1) & \text{if AvgSSIM}(i-1) > \text{AvgSSIM}(i) \end{cases} \tag{2}$$

AgvSSIM(i) stands for the average of the structure similarity value of all rendered views compared to the raw views, without quality loss from compression. In some specific case, we can see that user can ask the server to decrease or increase the number of transmitted views according to available bandwidth or total bitrates.

3.2.2 Bandwidth prediction

We used throughput based method called smoothed throughput method to calculate available bandwidth. It predicts the available bandwidth by moving average of the observed throughputs. This algorithm decides the best possible quality level considering the moving average of the throughput of downloading section measurement T_i. The expected throughput can be expressed as [40]:

$$T_e(t+1) = \begin{cases} (1-\delta) \times T_e(t) + \delta \times T_i(t), & if\ t > 0 \\ T_i(t) & if\ \ t = 0 \end{cases} \quad (3)$$

where, $T_i(t)$ is measured as the segment throughput of the last segment(instant throughput), T_e is estimated throughout, δ is a weighing value, and t represents the downloaded order of sequence.

Algorithm: Smoothed bitstream selection by moving average throughput.

Predication bandwidth has following features:

Input:

Throughput Instantaneous =T_i
Playlist = p
Level of video representation= V_n
(Lowest quality level =1)
Transmitted views number per video representation = V_n(i)
Bitrate of representation for N^{th} video representations =$Totalbitrate_n$
Counter

Start
if counter>0
Download from the lowest video bitrate: V_1
Update the estimated throughput: $T_e(p+1) \leftarrow T_i(p)$
Counter--
else
Estimate the available bandwidth based on Lookup table:
$T_e(p+1) \leftarrow (1-\delta) \times T_e(p) + \delta \times T_i(p)$
Find the suitable representations in server for $T_e(p+1)$:
$Totalbitrate_{n-1} \leq T_e(p+1) \leq Totalbitrate_n$
Download V_n
While
Number of candidates for N^{th} representation
Number ($Totalbitrate_n$)> 1
do
Decide number of transmitted views based on (2)
if*AgvSSIM*(*i*)>*AgvSSIM*(*i*-1)
DownloadV_n(*i*)
else
Download V_n(*i*-1)
end if
Counter - -
end if

- It starts with lowest quality of segments for the first time.
- To calculate the available network bandwidth, it starts moving average of the throughput from the last downloaded segments.
- Based on the moving average of the estimated throughput, the selected quality level can be adjusted up and down which requires multiple step switching. End user selects different numbers of transmitted views to download based on the calculated structure similarity values when specific total video bitrates are provided by the server. The switching operations are controlled by while loop in the algorithm.

This algorithm has two main advantages, first one is that it efficiently utilizes available bandwidth and is aware to changes in estimated available bandwidth. Second, it uses layer based method for MVD content of 3D video in terms of the number of transmitted views. This ultimately enhanced the perceptual quality of virtual views subsequently by rendering, thereby increasing the user's quality of experience.

3.2.3 Reconstruction based on the MVD format

After decoding process, MVD representation for potential 3D content is reconstructed as shown in Fig. 4. The rendering software commenced by Fruanhofer [8] has already shown better performance than MPEG VSRS when compared both in terms of Structure Similarity and Peak Signal to Noise Ratio for rendering the synthesized views from the MVD video. To avoid the transmission cost of large number of virtual views, we render the virtual views in client side. Sending all the virtual views is not an optimal solution in the case of best effort networks. Increasing number of virtual views plus multi-views linearly increase the bandwidth burden.

4 Simulation

4.1 Simulation setup

The experiments were conducted using three adjacent color texture views and depth maps from different MPEG test sequences [39]. Test sequences used for evaluation were as follows: *BookArrival*(resolution: 1024 × 768), *Newspaper* (resolution: 1024 × 768), *Balloons* (1024 × 768). Furthermore, picture group length and intra-period were to set 8 and 24 respectively, for all the test sequences. Each segment's duration was 10 s and we repeated each segment for 10 times to consider as sequences, because longer sequences were not feasible with the above mentioned sequences.

Due to high compression performance in comparison to the H.264/AVC encode, we selected HEVC 3D extension Encoder HM 11.0 [8]. We prepared two layers of stream as adaptation decision strategy over varying dynamic network. First layer consists of 2 views and their corresponding depths (2 V + D). Second layer consists of 3 views and their corresponding depths (3 V + D). We used views number 1 and 5 in *BookArrival*, views 2 and 6 in *Newspaper*, and views 3 and 5 in *Balloons*. Rest of the views we used were for second layer of stream to emulate the streaming of higher number of views.

Each segment is encoded using constant bitrate per views 300, 500, 800, 1000, 1200 and 1500 kbps. View and its depth have equal bitrate as pointed out in [15]. First layer consist of 2

views and its depth. Therefore, the total bits of first layer streams are 1200, 2000, 3200, 4000, 4800 and 6000 kbps respectively. We used the same encoding method for second layer streams in similar way and bitrates range from 1800 kbps to 9000 kbps. All the representations segments of first and second layers streams as well as MPD file were stored in the Internet information service of HTTP server [11]. We employed the DummyNet tool [25] on the end user side to increase the bandwidth. Initial bandwidth was set to 2 Mbps. After every 2 segments, bandwidth increased by 1Mbps. The software introduced in [4] was selected for the rendering purpose to generate virtual view points based on the MVD content received by the end user. The Blend Mode and Hole Filling Mode parameters were enabled. It is significant to note that in both layer experiments, aggregate numbers of views transmitted plus virtual views are same. For first layer, seven virtual viewpoints were rendered between transmitted views. While for second layer six virtual viewpoints were rendered, because 3 viewpoints are between two transmitted views as shown in Fig. 4.

4.2 Objective quality measurement

We utilize objective metrics PSNR and SSIM to demonstrate how the differences in number of views transmitting influence over the quality of virtual views, which eventually influence the quality of the end-user experience. The experimental result helps to decide which adaptation strategy of transmitting MVD content for auto-stereoscopic 3D display would be the best in terms of quality of experience under diverse situations.

Figures. 5, 6 and 7 explain the results of average PSNR, while Figs. 8, 9 and 10 explain the results of average SSIM for three sequences of different layers.

We calculated the PSNR of different layers as Quality of Experience metric for same total bitrate. Second layer that contained 3 views plus depth format input always had a higher PSNR than first layer which contained 2 views plus depth. Therefore, according to PSNR metric for fixed video bitrate, second layer produces better quality. However, when we used SSIM metric for calculating the quality of virtual views of different total bitrates, for same case, we saw the different scenario. When total video bitrate is lower, the SSIM of first layer is higher than

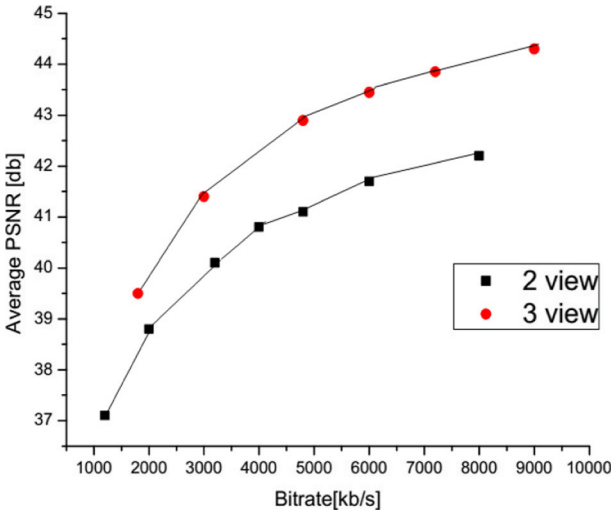


Fig. 5 PSNR for BookArrival

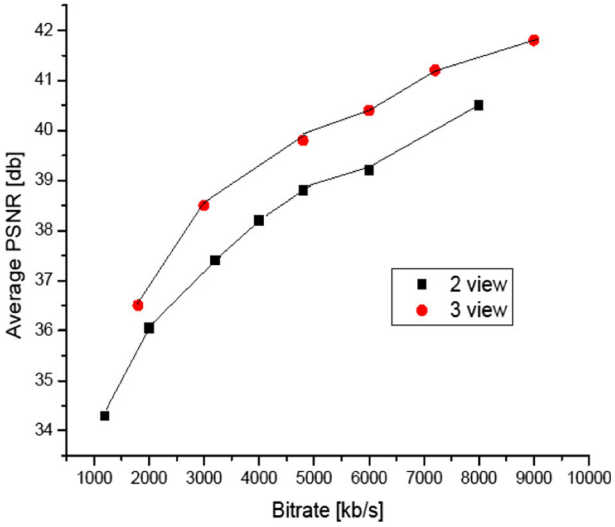


Fig. 6 PSNR for Newspaper

second layer. After increasing the total video bitrate, beyond the certain point, in this case 5000 kbps, second layer outperforms first layer which can be seen in Figs. 6, 8 and 10. In other words, total video bitrate determines number of transmitting views and quality of each view after rendered virtual views. In SSIM metric, we saw that transmitting lower number of views with optimized quality is better if the bitrate is lower than 5000 kbps. If the total video bitrate is higher than 5Mbps, then firstly increase the number of views and then increase the quality of each view. In comparison to PSNR, SSIM can imagine the quality of rendered virtual views better because it has better correlation with human perception. This helps us to put more number of views by exposing the network bandwidth threshold.

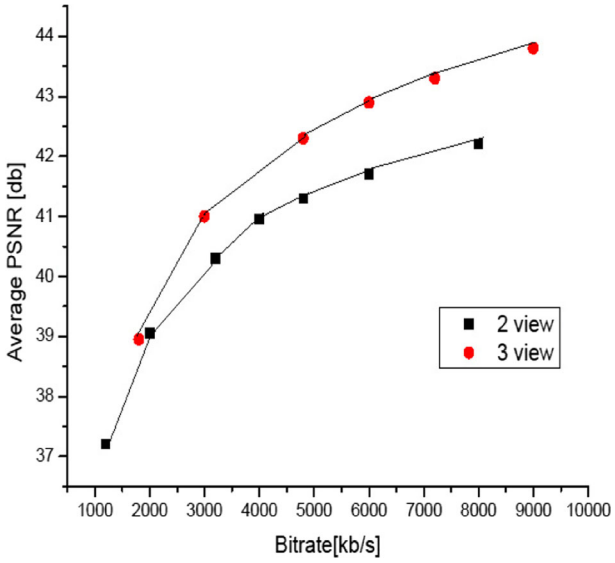


Fig. 7 PSNR for Balloons

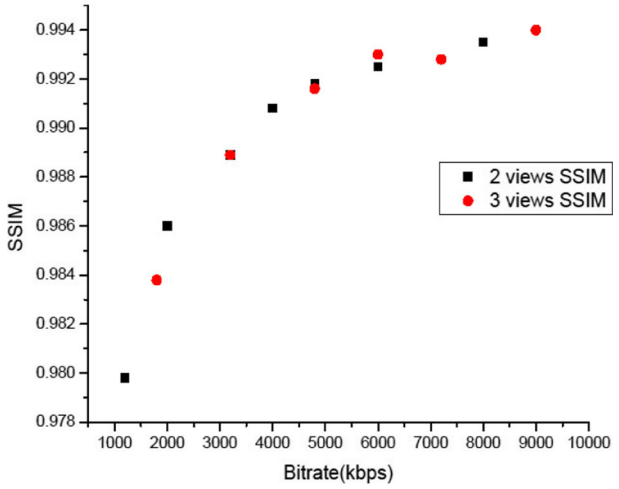


Fig. 8 SSIM for BookArrival

4.3 Subjective quality assessment

Subjective tests give the perceived quality of generated views. It looks into comparision among different subjective quality of different layers containing the different number of input view plus depth at various bitrates. We performed subjective test according to ITU-R recommendation BT.500–13 [13]. In the setup of subjective test, 18 non-expert observer, twelve males and six females, age range from 25 to 45, participated in the test. All the test session began after a short training and instructions. During the training phase, the observers were introduced to the test environment, grading scale, what they had to evaluate and how to express their opinions. Each assessment session lasted upto half an hour.

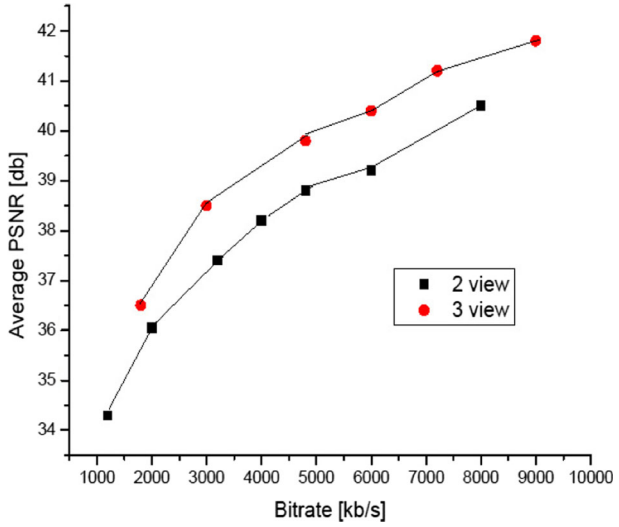


Fig. 9 SSIM for Newspaper

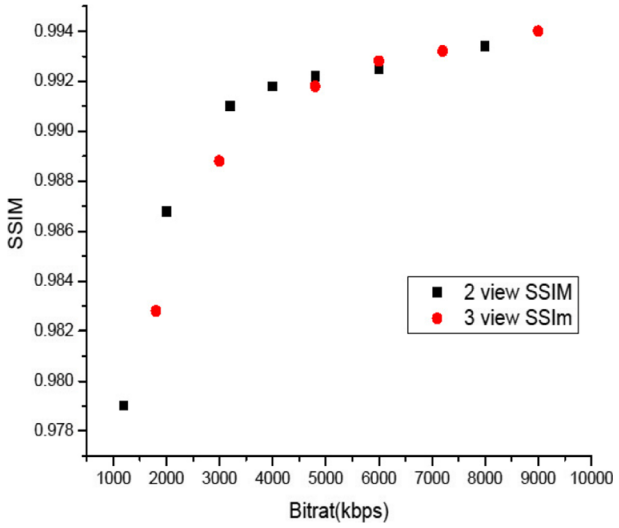


Fig. 10 SSIM for Balloons

The Double Stimulus Continuous Quality Scale (DSCQS) was utilized using a scale thatranged from 0 (Bad) to 100 (Excellent) for 3 view input. The observers were provided to watch the reference video together with generated views from three input views. The network varied from 0.6 Mbps to 4.2 Mbps. One 4 s grey interval is inserted between two 10 s test sequences and repeated again and again as many times as they wished before reaching a verdict. The collected Mean Opinion Score(MOS) were then analyzed for each subject data over various test conditions. The results are shown in Fig. 11. The results were also compared with [22] that used view reconstruction method. The proposed method provides better result than the earlier method [22] consistently as shown in Table 1 in both objectively and subjectively.

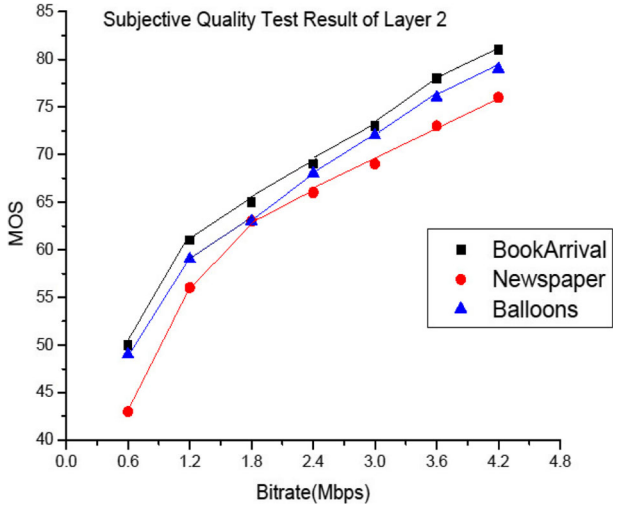


Fig. 11 Subjective test results for the BookArrival,Newspaper and Balloons

Table 1 Subjective test comparison

Sequences	Method	Quality Measure	
		PSNR (db)	MOS
BookArrival	Proposed method	44.31	81.12
	Ozcinar	37.91	73.75
	MPEG VSRS	36.80	58.87
Newspaper	Proposed method	41.31	71.73
	Ozcinar	38.91	63.93
	MPEG VSRS	37.82	57.46
Balloon	Proposed method	43.21	80.32
	Ozcinar	40.46	78.42
	MPEG VSRS	38.77	56.12

5 Conclusion

In this paper, we proposed a new quality aware multi-view video streaming over Internet (HTTP) to concentrate on decision strategy and present a comprehensive analysis in a dynamic network environment. One of the foremost contributions of the proposed technique was a user-perceived quality responsive adaptation stand on client-server model. The proposed adaptation tactic was applied by changing the number of transmitted views to smooth the progress of a quality aware bandwidth adaptation system in auto-stereoscopic 3D display. Two of the state-of-art techniques known as MPEG-DASH and HEVC were used in our system. The simulation results have shown that significant quality improvements are obtained under challenging network conditions.

References

1. (MPEG) IJSW (2010) Dynamic adaptive streaming over http. w11578,CD 23001–6, w11578, CD 23001–6. ISO/IEC JTC 1/SC 29/WG11
2. Apple Developer (2016) Apple HTTP live streaming. https://developer.apple.com/streaming/
3. Benzie P et al (2007) A survey of 3DTV displays: techniques and technologies. IEEE Trans Circuits Syst Video Technol 17(11):1647–1658
4. Bosc E et al (2013) A study of depth/texture bit-rate allocation in multi-view video plus depth compression. annals of telecommunications-annales des télécommunications 68(11–12):615–625
5. Buchowicz A (2013) Video coding and transmission standards for 3D television — a survey. Opto-Electron Rev 21:39
6. Chakareski J (2013) Adaptive multiview video streaming: challenges and opportunities. IEEE Commun Mag 51(5):94–100
7. De Simone F, Dufaux F (2013) Comparison of DASH adaptation strategies based on bitrate and quality signalling. In: Multimedia Signal Processing (MMSP), 2013 I.E. 15th international workshop on. IEEE
8. H.-H.-I. Fruanhofer (2013) HEVC 3D extension test model (3DV HTM) version 11.0. Available from: https://hevc.hhi.fraunhofer.de/svn/svn_3DVCSoftware/tags/HTM-11.0/
9. Gao Y et al (2017) Event classification in microblogs via social tracking. ACM Trans Intell Syst Technol (TIST) 8(3):35
10. Ho YS, Lee EK, Lee C (2008) Multiview video test sequence and camera parameters. Tech. Rep. MPEG2008/M15419 ISO/IEC JTC1/SC29/WG11. Archamps, France
11. Internet Information Service (2014)
12. ISO/IEC JTC1/SC29/WG11 (2005) Report of the subjective quality evaluation for MVC call for evidence. Tech. Rep. MPEG2005/N6999, Hong Kong, China

13. ITU-R Recommendation (2012) ITU-R BT.500–13, Methodology for the subjective assessment of the quality of television pictures. Tech. Rep
14. Jacobson V (1998) Congestion avoidance and control. In: Proceeding of SIGCOMM '88. ACM SIGCOMM Computer Communication Review 18(4):314–329
15. Kauff P et al (2007) Depth map creation and image-based rendering for advanced 3DTV services providing interoperability and scalability. Signal Process Image Commun 22(2):217–234
16. Kuschnig R, Kofler I, Hellwagner H (2011) Evaluation of HTTP-based request-response streams for internet video streaming. In: Proceedings of the second annual ACM conference on Multimedia systems. ACM
17. Li B, Li H, Li L, Zhang J (2012) Rate control by R-lambda model for HEVC. In: JCTVC-K0103, JCTVC of ISO/IEC and ITU-T, 11th Meeting, Shanghai, China
18. Miller K et al (2012) Adaptation algorithm for adaptive streaming over HTTP. In: Packet Video Workshop (PV), 2012 19th International. IEEE
19. Müller K et al (2013) 3D high-efficiency video coding for multi-view video and depth data. IEEE Trans Image Process 22(9):3366–3378
20. Ohm J-R et al (2012) Comparison of the coding efficiency of video coding standards—including high efficiency video coding (HEVC). IEEE Trans Circuits Syst Video Technol 22(12):1669–1684
21. Ozcinar C et al (2016) Adaptive delivery of immersive 3D multi-view video over the internet. Multimed Tools Appl 75(20):12431–12461
22. Ozcinar C, Ekmekcioglu E, Kondoz A (2016) Quality-aware adaptive delivery of multi-view video. In: Acoustics, Speech and Signal Processing (ICASSP), 2016 I.E. International Conference on. IEEE
23. Oztas B et al (2014) A rate adaptation approach for streaming multiview plus depth content. in Computing, Networking and Communications (ICNC), 2014 International Conference on. IEEE
24. Psannis KE, Hadjinicolaou MG, Krikelis A (2006) MPEG-2 streaming of full interactive content. IEEE Trans Circuits Syst Video Technol 16(2):280–285
25. Rizzo L (1997) Dummynet: a simple approach to the evaluation of network protocols. ACM SIGCOMM Comput Commun Rev 27(1):31–41
26. Roodaki, H., M.R. Hashemi, and S. Shirmohammadi, (2012) A new methodology to derive objective quality assessment metrics for scalable multiview 3D video coding. ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM), 2012. 8(3s): p. 44
27. Savas SS, Tekalp AM, Gurler CG (2011) Adaptive multi-view video streaming over P2P networks considering quality of experience. In: Proceedings of the 2011 ACM workshop on social and behavioural networked media access. ACM
28. Schierl T et al (2012) System layer integration of high efficiency video coding. IEEE Trans Circuits Syst Video Technol 22(12):1871–1884
29. Seufert M et al (2015) A survey on quality of experience of HTTP adaptive streaming. IEEE Commun Surv Tutorials 17(1):469–492
30. Shimizu S et al (2007) View scalable multiview video coding using 3-d warping with depth map. IEEE Trans Circuits Syst Video Technol 17(11):1485–1495
31. Smolic A et al (2007) Coding Algorithms for 3DTV—A Survey. IEEE Trans Circuits Syst Video Technol 17(11):1606–1621
32. Smolic A et al (2008) Intermediate view interpolation based on multiview video plus depth for advanced 3D video systems. In: Image Processing, 2008. ICIP 2008. 15th IEEE International Conference on. IEEE
33. Sodagar I (2011) The MPEG-DASH standard for multimedia streaming over the internet. IEEE MultiMedia 18(4):62–67
34. Sripanidkulchai K, Maggs B, Zhang H (2004) An analysis of live streaming workloads on the internet. In Proceedings of the 4th ACM SIGCOMM conference on internet measurement. ACM
35. Stockhammer T (2011) Dynamic adaptive streaming over HTTP–: standards and design principles. In: Proceedings of the second annual ACM conference on Multimedia systems. ACM
36. Sullivan GJ et al (2012) Overview of the high efficiency video coding (HEVC) standard. IEEE Trans Circuits Syst Video Technol 22(12):1649–1668
37. Tanimoto M (2009) Overview of FTV (free-viewpoint television). In: Proceeding of 2009 I.E. International Conference on Multimedia and Expo, New York, pp 1552–1523
38. Tanimoto M (2010) Overview of Free-viewpoint TV (FTV). In: Javidi B, Fournel T (eds) Information optics and photonics. Springer, New York. https://doi.org/10.1007/978-1-4419-7380-1_9
39. Tanimoto M et al (2013) Proposal on a new activity for the third phase of FTV. In: the 105th meeting of MPEG
40. Thang TC et al (2014) An evaluation of bitrate adaptation methods for HTTP live streaming. IEEE J Sel Areas Commun 32(4):693–705
41. The official Microsoft IIS site (2016) Microsoft smooth-streaming. Microsoft Corporation 2016; Available from: https://www.iis.net/downloads/microsoft/smooth-streaming

42. Toni L et al (2014) Optimal set of video representations in adaptive streaming. In: proceedings of the 5th ACM Multimedia Systems Conference. ACM
43. Vetro A, Wiegand T, Sullivan GJ (2011) Overview of the stereo and multiview video coding extensions of the H. 264/MPEG-4 AVC standard. Proc IEEE 99(4):626–642
44. Wiegand T et al (2003) Overview of the H. 264/AVC video coding standard. IEEE Trans Circuits Syst Video Technol 13(7):560–576
45. Zhao S et al (2015) Strategy for dynamic 3D depth data matching towards robust action retrieval. Neurocomputing 151:533–543
46. Zhao S et al (2015) View-based 3D object retrieval via multi-modal graph learning. Signal Process 112:110–118
47. Zhao S, Yao H, Jiang X (2015) Predicting continuous probability distribution of image emotions in valence-arousal space. In: Proceedings of the 23rd ACM international conference on Multimedia. ACM
48. Zhao S et al (2016) Predicting personalized emotion perceptions of social images. In: Proceedings of the 2016 ACM on Multimedia Conference. ACM
49. Zhao S et al (2017) Continuous probability distribution prediction of image emotions via multitask shared sparse regression. IEEE Trans Multimedia 19(3):632–645
50. Zhao S et al (2017) Approximating discrete probability distribution of image emotions by multi-modal features fusion. Transfer 1000:1
51. Zhao S, Gao Y, Ding G, Chua TS (2017) Real-time multimedia social event detection in microblog. IEEE Trans on Cybernetics 99:1–14
52. Zhou C et al (2012) A control-theoretic approach to rate adaptation for dynamic HTTP streaming. In: Visual Communications and Image Processing (VCIP), 2012 IEEE. IEEE

Nabin Kumar Karn is PhD student in School of Computer Science and Technology at Harbin Institute of Technology Harbin China. His research interests include measuring Internet performance, topology measurement and management, Network Security and Traffic classification. Earlier he had worked 4 years in ISP as a network engineer.

Hongli Zhang is the deputy dean of School of Computer Science and Technology, deputy director of State Key Laboratory for Computer Information Content Security and Technical Committee of Security National Engineering Laboratory on computer information content security technologies, a member of Technical Committee of National Computer Network Emergency Response Technical Team/Coordination Center of China (CNCERT/CC), Ministry of Education Teaching Committee on Information Security and Chinese Computer Association. Her research interests include network and information security, network measurement and social network. In network and information security field, her research focuses on anomaly detection of network traffic, peer-to-peer security, protection of data security and privacy in cloud computing and malware detection in virtual machine. She published more than 90 papers, over 59 of which are indexed by SCI, and also obtained more than 10 patents in China.

Feng Jiang received the B.S., M.S., and Ph.D. degrees in computer science from Harbin Institute of Technology (HIT), Harbin, China, in 2001, 2003, and 2008, respectively. He is now an Associated Professor in the Department of Computer Science, HIT and a visiting scholar in the School of Electrical Engineering, Princeton University. His research interests include computer vision, pattern recognition and image and video processing.