Traffic and Statistical Multiplexing Characterization of 3D Video Representation Formats (Extended Version)

Akshay Pulipaka, Patrick Seeling, Martin Reisslein, and Lina J. Karam

***Abstract*—The network transport of 3D video, which contains two views of a video scene, poses significant challenges due to the increased video data compared to conventional single-view video. Addressing these challenges requires a thorough understanding of the traffic and multiplexing characteristics of the different representation formats of 3D video. We examine the average bitrate-distortion (RD) and bitrate variability-distortion (VD) characteristics of three main representation formats. Specifically, we compare multiview video (MV) representation and encoding, frame sequential (FS) representation, and side-by-side (SBS) representation, whereby conventional single-view encoding is employed for the FS and SBS representations. Our results for long 3D videos in full HD format indicate that the MV representation and encoding achieves the highest RD efficiency, while exhibiting the highest bitrate variabilities. We examine the impact of these bitrate variabilities on network transport through extensive statistical multiplexing simulations. We find that when multiplexing a small number of streams, the MV and FS representations require the same bandwidth. However, when multiplexing a large number of streams or smoothing traffic, the MV representation and encoding reduces the bandwidth requirement relative to the FS representation.**

I. Introduction

Multiview video provides several views taken from different perspectives, whereby each view consists of a sequence of video frames (pictures). Multiview video with two marginally different views of a given scene can be displayed to give viewers the perception of depth and is therefore commonly referred to as three-dimensional (3D) video or stereoscopic video [2]–[6]; for brevity we use the term "3D video" throughout. Providing 3D video services over transport networks requires efficient video compression (coding) techniques and transport mechanisms to accommodate the large volume of video data from the two views on bandwidth limited transmission links. While efficient coding techniques for multiview video have been researched extensively in recent years [7], [8], the network transport of encoded 3D video is largely an open research problem.

Technical Report, School of Electrical, Computer, and Energy Eng., Arizona State Univ., November 2012. This extended technical report accompanies [1].

Supported in part by the National Science Foundation through grant No. CRI-0750927.

Please direct correspondence to M. Reisslein.

A. Pulipaka, M. Reisslein, and L.J. Karam are with the School of Electrical, Computer, and Energy Engineering Arizona State University, Tempe, AZ 85287-5706, http://trace.eas.asu.edu, Email: {akshay.pulipaka, reisslein, karam}@asu.edu

P. Seeling is with Central Michigan University, Mount Pleasant, MI 48859, Email: pseeling@ieee.org

Previous studies on 3D video transport have primarily focused on the network and transport layer protocols and file formats [9]–[12]. For instance, [10], [13] examine the extension of common transport protocols, such as the datagram congestion control protocol (DCCP), the stream control transmission protocol (SCTP), and the user datagram protocol (UDP) to 3D streaming, while the use of two separate Internet Protocol (IP) channels for the delivery of multiview video is studied in [14]. Another existing line of research has studied prioritization and selective transport mechanisms for multiview video [15], [16].

In this study, we examine the fundamental traffic and statistical multiplexing characteristics of the main existing approaches for representing and encoding 3D video for long (54,000 frames) full HD (1920×1080) 3D videos. More specifically, we consider (i) multiview video (MV) representation and encoding, which exploits the redundancies between the two views, (ii) frame sequential (FS) representation, which merges the two views to form a single sequence with twice the frame rate and applies conventional single-view encoding, and (iii) side-by-side (SBS) representation, which halves the horizontal resolution of the views and combines them to form a single frame sequence for single-view encoding.

We find that the MV representation achieves the most efficient encoding, but generates high traffic variability, which makes statistical multiplexing more challenging. Indeed, for small numbers of multiplexed streams, the FS representation with conventional single-view coding has the same transmission bandwidth requirements as the MV representation with multiview coding. Only when smoothing the MV traffic or multiplexing many streams can transport systems benefit from the more efficient MV encoding.

In order to support further research on the network transport of 3D video, we make all video traces [17] used in this study publicly available in the video trace library http://trace.eas.asu.edu. In particular, video traffic modeling [18]–[22] requires video traces for model development and validation. Thus, the traffic characteristics of 3D video covered in this study will support the nascent research area of 3D video traffic modeling [23]. Similarly, video traffic management mechanisms for a wide range of networks, including wireless and optical networks, are built on the fundamental traffic and multiplexing characteristics of the encoded video traffic [24]–[27]. Thus, the broad traffic and statistical multiplexing evaluations in this study provide a basis for the emerging research area on 3D video traffic

management in transport networks [28], [29].

II. Multiview Video Representation, Encoding, and Streaming

In this section, we provide a brief overview of the main representation formats for multiview video [4] as well as the applicable encoding and streaming approaches.

A. Multiview video representation formats

With the full resolution multiview format, which we refer to as multiview (MV) format for brevity, each view $v,\ v = 1, \ldots, V$, is represented with the full resolution of the underlying spatial video format. For instance, the MV format for the full HD resolution of 1920×1080 pixels consists of a sequence of 1920×1080 pixel frames for each view v. Each view has the same frame rate as the underlying temporal video format. For example, for a video with a frame rate of $f = 24$ frames/s, each view has a frame rate of $f = 24$ frames/s.

With the frame sequential (FS) representation, the video frames of the V views (at the full spatial resolution) are temporally multiplexed to form a single sequence of video frames with frame rate Vf. For instance, for $V = 2$ views, the video frames from the left and right views are interleaved in alternating fashion to form a single stream with frame rate $2f$.

Frame-compatible representation formats have been introduced to utilize the existing infrastructure and equipment for the transmission of stereoscopic two-view video [4]. The $V = 2$ views are spatially sub-sampled and multiplexed to form a single sequence of video frames with the same temporal and spatial resolution as the underlying video format [30]. In the side-by-side (SBS) format, the left and right views are spatially sub-sampled in the horizontal direction and are then combined side-by-side. For instance, for the full HD format, the left and right views are sub-sampled to 960×1080 pixels. Thus, when they are combined in the side-by-side format, they still occupy the full HD resolution for every frame. However, each frame contains the left and right views at only half the original horizontal resolution. In the top-and-bottom format, the left and right views are sub-sampled in the vertical direction and combined in top-and-bottom (above-below) fashion. For other formats, we refer to [4], [30]–[32]. We consider the side-by-side (SBS) representation format in our study, since it is one of the most widely used frame-compatible formats, e.g., it is currently being deployed in Japan to transmit 3D content for TV broadcasting over the BS11 satellite channel [30]. The major drawback of these frame-compatible formats is that the spatial sub-sampling requires interpolation (and concomitant quality degradation) to extract the left and right views at their original resolution.

B. Multiview video compression

We now proceed to briefly introduce the compression approaches that can be applied to the representation formats outlined in the preceding subsection. Building on the concept of inter-view prediction [33], multiview video coding [8] exploits the redundancies across different views of the same scene (in addition to the temporal and intra-view spatial redundancies exploited in single-view encoding). Multiview video coding is applicable only to the multiview (MV) representation format since this is the only format to retain distinct sequences of video frames for the views. For the case of 3D video, the recent official ITU multiview video coding reference software, referred to as JMVC, first encodes the left view, and then predictively encodes the right view with respect to the encoded left view.

The frame sequential (FS) and side-by-side (SBS) representation formats present a single sequence of video frames to the encoder. Thus, conventional single-view video encoders can be applied to the FS and SBS representations. We employ the state-of-the-art JSVM reference implementation [34] of the scalable video coding (SVC) extension of the advanced video coding (AVC) encoder in single-layer encoding mode.

For completeness, we briefly note that each view could also be encoded independently with a single-view encoder, which is referred to as simulcasting. While simulcasting has the advantage of low complexity, it does not exploit the redundancies between the views, resulting in low encoding efficiency [4]. A currently active research direction in multiview video encoding is asymmetric coding [10], [35], which encodes the left and right views with different properties, e.g., different quantization scales. For other ongoing research directions in encoding, we refer to the overviews in [4], [8], [10].

C. Multiview video streaming

a) SBS representation: The $V = 2$ views are integrated into one frame sequence with the spatial resolution and frame rate f of the underlying video. For frame-by-frame transmission of a sequence with M frames, frame $m,\ m = 1, \ldots, M$, of size X_m [bytes] is transmitted during one frame period of duration $1/f$ at a bit rate of $R_m = 8fX_m$ [bit/s].

b) MV representation: There are a number of streaming options for the MV representation with V views. First, the V streams resulting from the multiview video encoding can be streamed individually. We let $X_m(v),\ m = 1, \ldots, M,\ v = 1, \ldots, V$, denote the size [bytes] of the encoded video frame m of view v and note that $R_m(v) = 8fX_m(v)$ [bit/s] is the corresponding bitrate. The mean frame size of the encoded view v is

$$\bar{X}(v) = \frac{1}{M} \sum_{m=1}^{M} X_m(v) \tag{1}$$

and the corresponding average bit rate is $\bar{R}(v) = 8f\bar{X}(v)$. The variance of these frame sizes is

$$S_X^2(v) = \frac{1}{M-1} \sum_{m=1}^{M} \left[X_m(v) - \bar{X}(v)\right]^2. \tag{2}$$

The coefficient of variation of the frame sizes of view v [unit free] is the standard deviation of the frame sizes $S_X(v)$ normalized by the mean frame size

$$CoV_X(v) = \frac{S_X(v)}{\bar{X}(v)} \tag{3}$$

and is widely employed as a measure of the variability of the frame sizes, i.e., the traffic bitrate variability. Plotting the CoV as a function of the quantization scale (or equivalently, the average PSNR video quality) gives the bitrate variability-distortion (VD) curve [36], [37].

Alternatively, the V streams can be merged into one multiview stream. We consider two elementary merging options, namely sequential (S) merging and aggregation (combining). With sequential merging, the M frames of the V views are temporally multiplexed in round-robin fashion, i.e., first view 1 of frame 1, followed by view 2 of frame 1, ..., followed by view V of frame 1, followed by view 1 of frame 2, and so on. From the perspective of the video transport system, each of these VM video frames (pictures) can be interpreted as a video frame to be transmitted. In this perspective, the average frame size of the resulting multiview stream is

$$\bar{X} = \frac{1}{V} \sum_{v=1}^{V} \bar{X}(v). \tag{4}$$

Noting that this multiview stream has V frames to be played back in each frame period of duration $1/f$, the average bit rate of the multiview stream is

$$\bar{R} = 8Vf\bar{X}. \tag{5}$$

The variance of the frame sizes of the sequentially (S) merged multiview stream is

$$S_S^2 = \frac{1}{(M-1)(V-1)} \sum_{m,v=1}^{M,V} \left[X_m(v) - \bar{X}\right]^2 \tag{6}$$

with the corresponding $\mathrm{CoV}_S = S_S/\bar{X}$.

With combining (C), the V encoded views corresponding to a given frame index m are aggregated to form one *multiview frame* of size $X_m = \sum_{v=1}^{V} X_m(v)$. For 3D video with $V = 2$, the pair of frames for a given frame index m (which corresponds to a given capture instant of the frame pair) constitutes the multiview frame m. Note that the average size of the multiview frames is $V\bar{X}$ with $\bar{X}$ given in (4). Further, note that these multiview frames have a rate of f multiview frames/s; thus, the average bit rate of the multiview stream resulting from aggregation is the same $\bar{R}$ as given in (5). However, the variance of the sizes of the (combined) multiview frames is different from (6); specifically,

$$S_C^2 = \frac{1}{(M-1)} \sum_{m=1}^{M} \left[X_m - V\bar{X}\right]^2 \tag{7}$$

and $\mathrm{CoV}_C = S_C/(V\bar{X})$.

c) FS representation: Similar to the MV representation, the FS representation can be streamed sequentially (S) with the traffic characterizations given by (4)–(6). Or, the V encoded frames for a given frame index m can be combined (C), analogous to the aggregation of the MV representation, resulting in the frame size variance given by (7).

d) Frame size smoothing: The aggregate streaming approach combines all encoded video data for one frame period of playback duration $1/f$ [s] and transmits this data at a constant bitrate over the $1/f$ period. Compared to the sequential streaming approach, the aggregate streaming approach thus performs smoothing across the V views, i.e., effectively smoothes the encoded video data over the duration of one frame period $1/f$. This smoothing concept can be extended to multiple frame periods, such as a Group of Pictures (GoP) of the encoder [38]. For GoP smoothing with a GoP length of G frames, the encoded views from G frames are aggregated and streamed at a constant bitrate over the period G/f [s].

III. Evaluation Set-up

In this section, we describe our evaluation set-up, including the employed 3D video sequences, the encoding set-up, and the video traffic and quality metrics.

A. Video sequences

For a thorough evaluation of the traffic characteristics, especially the traffic variability, the publicly available, relatively short 3D video research sequences [39] are not well suited. Therefore, we employ long 3D ($V = 2$) video sequences of $M = 52,100$ frames each. That is, we employ 51,200 left-view frames (pictures) and 51,200 right-view frames (pictures) for each test video. We have conducted evaluations with *Monsters vs Aliens* and *Clash of the Titans*, which are computer-animated fiction movies, *Alice in Wonderland*, which is a fantasy movie consisting of a mix of animated and real-life content, and *IMAX Space Station*, a documentary. All videos are in the full HD 1920×1080 pixels format and have a frame rate of $f = 24$ frames/s for each view.

B. Encoding set-up

We encoded the multiview representation with the reference software JMVC (version 8.3.1). We encoded the FS and SBS representations with the broadly used H.264 SVC video coding standard using the H.264 reference software JSVM (version 9.19.10) [34], [40] in single-layer encoding mode. We set the GOP length to $G = 16$ frames for the MV and SBS encodings; for the FS encodings we set $G = 32$ so that all encodings have the same playback duration between intracoded (I) frames, i.e., support the same random access granularity. We employ two different GoP patterns: (B1) with one bi-directionally predicted (B) frame between successive intracoded (I) and predictive encoded (P) frames, and (B7) with seven B frames between successive I and P frames. We conducted encodings for the quantization parameter settings 24, 28, and 34.

C. Traffic and quality metrics

We employ the peak signal to noise ratio (PSNR) [41], [42] of the video luminance signal of a video frame m, $m = 1, \ldots, M$, of a view v, $v = 1, \ldots, V$, as the objective quality metric of video frame m of view v. We average these video frame PSNR values over the VM video frames of a given video sequence to obtain the average PSNR video quality. For the MV and FS representations, the PSNR evaluation is conducted over the full HD spatial resolution of each view of a given frame. We note that in the context of asymmetric 3D

video coding [35], the PSNR values of the two views may be weighed unequally, depending on their relative scaling (bit rate reduction) [43]. We do not consider asymmetric video coding in this study and weigh the PSNR values of both views equally.

For the SBS representation, we report for some encodings the PSNR values from the comparison of the uncompressed SBS representation with the encoded (compressed) and subsequently decoded SBS representation as SBS without interpolation (SBS-NI). We also report for all encodings the comparison of the original full HD left and right views with the video signal obtained after SBS representation, encoding, decoding, and subsequent interpolation back to full HD format as SBS with interpolation (SBS-I). Unless otherwise noted, the SBS results are for the SBS representation with interpolation. We employed the JSVM reference down-sampling with a Sine-windowed Sinc-function and the corresponding normative up-sampling using a set of 4-tap filters [34]. We plot the average PSNR video quality [dB] as a function of the average streaming bitrate $\bar{R}$ [bit/s] to obtain the RD curve and the coefficient of variation of the frame sizes CoV as a function of the average PSNR video quality to obtain the VD curve.

IV. Traffic and Quality Results

In this section we present the RD and VD characteristics for the examined 3D video representation formats. We briefly note that generally the encodings with one B frame between successive I and P frames follow the same trends as observed for the encodings with seven B frames; the main difference is that the encodings with one B frame have slightly higher bitrates and slightly lower CoV values, which are effects of the lower level of predictive encoding.

A. Bitrate-distortion (RD) characteristics

In Fig. 1, we plot the RD curves of the multiview representation encoded with the multiview video codec for streaming the left view only (MV-L), the right view only (MV-R), and the merged multiview stream (MV). Similarly, we plot the RD curves for the frame sequential (FS) representation and the side-by-side (SBS) representation encoded with the conventional single-view codec.

From the MV curves in Fig. 1, we observe that the right view has a significant RD improvement compared to the left view. This is because of the inter-view prediction of the multiview encoding, which exploits the inter-view redundancies by encoding the right view with prediction from the left view.

Next, turning to the side-by-side (SBS) representation, we observe that SBS with interpolation can achieve similar or even slightly better RD efficiency than FS for the low to medium quality range of videos with real-life content (*Alice in Wonderland* and *IMAX Space Station*). However, SBS has consistently lower RD efficiency than the MV representation. In additional evaluations for the B7 GoP pattern, we compared the uncompressed SBS representation with the encoded (compressed) and subsequently decoded SBS representation and found that the RD curve for this SBS representation without interpolation (SBS-NI) lies between the MV-L and MV RD curves. We observed from these additional evaluations that the interpolation to the full HD format (SBS-I) significantly reduces the average PSNR video quality, especially for encodings in the higher quality range.

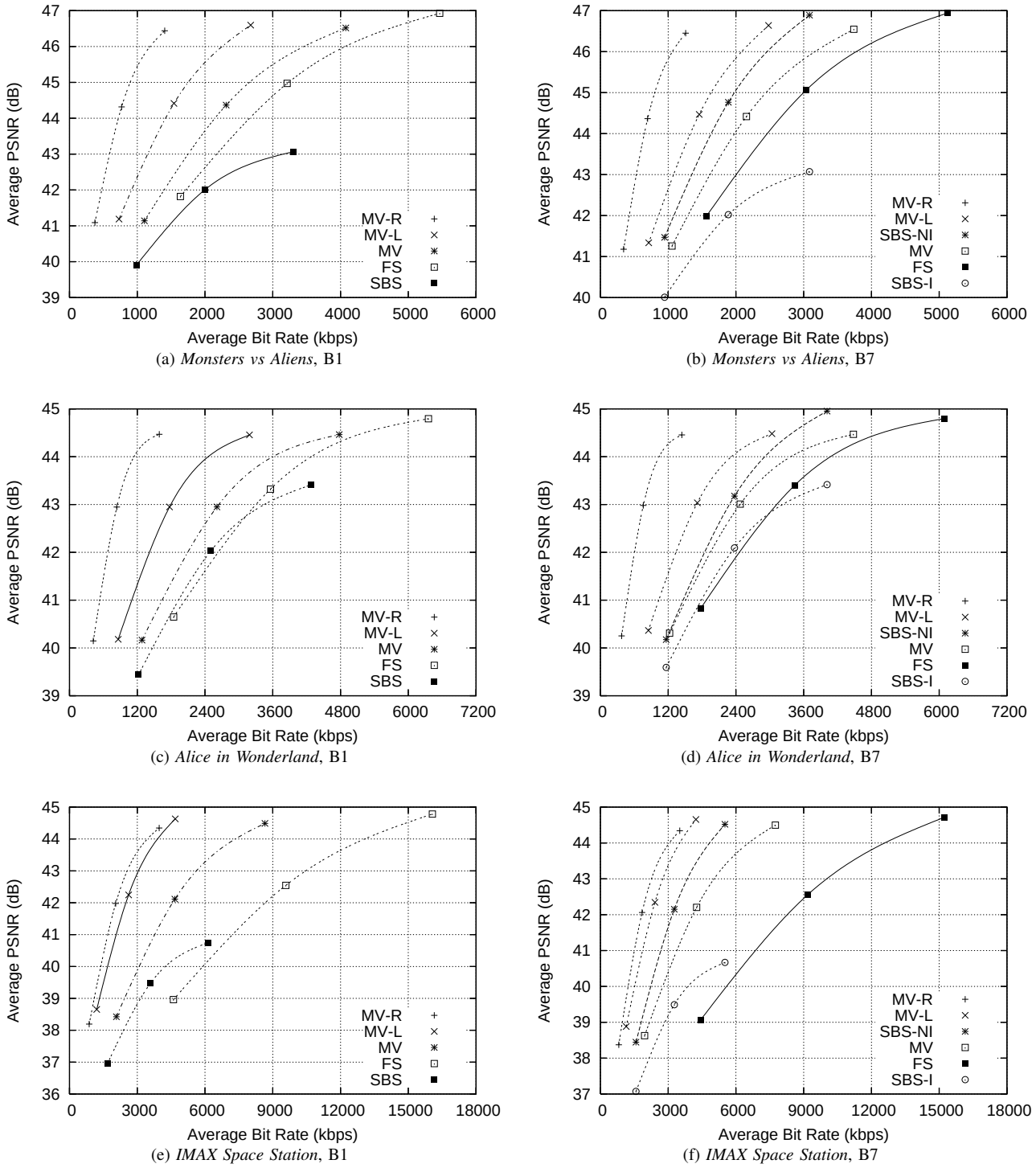
Finally, we observe from Fig. 1 that the MV representation in conjunction with multiview encoding has consistently higher RD efficiency than the FS representation with conventional single-view encoding. The FS representation essentially translates the multi-view encoding problem into a temporally predictive coding problem. That is, the FS representation temporally interleaves the left and right views and then employs state-of-the-art temporal predictive encoding. The results in Fig. 1 indicate that this temporal predictive encoding can not exploit the inter-view redundancies as well as the state-of-the-art multiview encoder.

B. Bitrate variability-distortion (VD) characteristics

In Fig. 2, we plot the VD curves for the examined multiview (MV), frame sequential (FS), and side-by-side (SBS) representation formats; whereby, for MV and FS, we plot both VD curves for sequential (S) merging and aggregation (C). We first observe from Fig. 2 that the MV representation with sequential streaming (MV-S) has the highest traffic variability. This high traffic variability is primarily due to the size differences between successive encoded left and right views. In particular, the left view is encoded independently and is thus typically large. The right view is encoded predictively from the left view and thus typically small. This succession of large and small views (frames), whereby each view is treated as an independent video frame by the transmission system, i.e., is transmitted within half a frame period $1/(2f)$ in the sequential streaming approach, leads to the high traffic variability. Smoothing over one frame period $1/f$ by combing the two views of each frame from the MV encoding significantly reduces the traffic variability. In particular, the MV encoding with aggregation (MV-C) has generally lower traffic variability than the SBS streams.

We further observe from the MV results in Fig. 2 that the left view (MV-L) has significantly higher traffic variability than the right view (MV-R). The large MV-L traffic variabilities are primarily due to the typically large temporal variations in the scene content of the videos, which result in large size variations of the MV-L frames which are encoded with temporal prediction across the frames of the left view. In contrast, the right view is predictively encoded from the left view. Due to the marginal difference between the two perspectives of the scene employed for the two views of 3D video, the content variations between the two views (for a given fixed frame index m) are small relative to the scene content variations occurring over time.

Turning to the FS representation, we observe that FS with sequential streaming has CoV values near or below the MV representation with aggregation. Similarly to the MV representation, aggregation significantly reduces the traffic variability of the FS representation. In fact, we observe from Fig. 2 that the FS representation with aggregation has consistently the lowest CoV values. The lower traffic variability of the FS representation is consistent with its relatively less RD-efficient



(a) *Monsters vs Aliens*, B1

(b) *Monsters vs Aliens*, B7

(c) *Alice in Wonderland*, B1

(d) *Alice in Wonderland*, B7

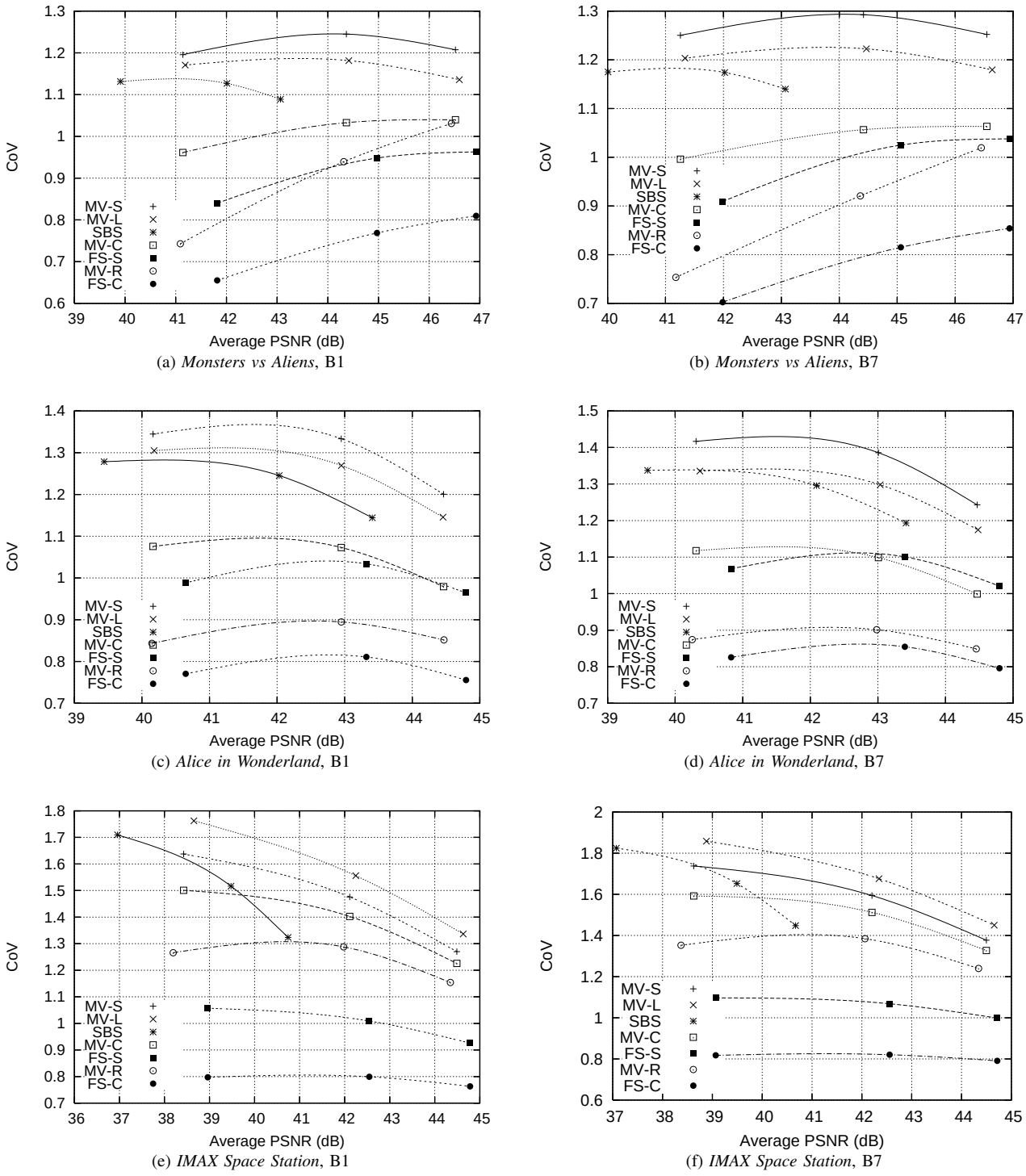
(e) *IMAX Space Station*, B1

(f) *IMAX Space Station*, B7

Fig. 1. RD curves for multiview (MV) representation, frame sequential (FS) representation, and side-by-side (SBS) representation for GoP patterns with one B frame between successive I and P frames (B1) as well as seven B frames between successive I and P frames (B7).

encoding. The MV representation and encoding exploits the inter-view redundancies and thus encodes the two views of each frame more efficiently, leading to relatively larger variations in the encoded frame sizes as the video content and scenes change and present varying levels of inter-view redundancy. The FS representation with single-view encoding, on the other hand, is not able to exploit these varying degrees of inter-view redundancy as well, resulting in less variability in the view and frame sizes, but also larger average frame sizes.

In additional evaluations that are not included here in detail due to space constraints, we found that frame size smoothing over one GoP reduces the traffic variability significantly, especially for the burstier MV representation. For instance,



(a) *Monsters vs Aliens*, B1 (b) *Monsters vs Aliens*, B7

(c) *Alice in Wonderland*, B1 (d) *Alice in Wonderland*, B7

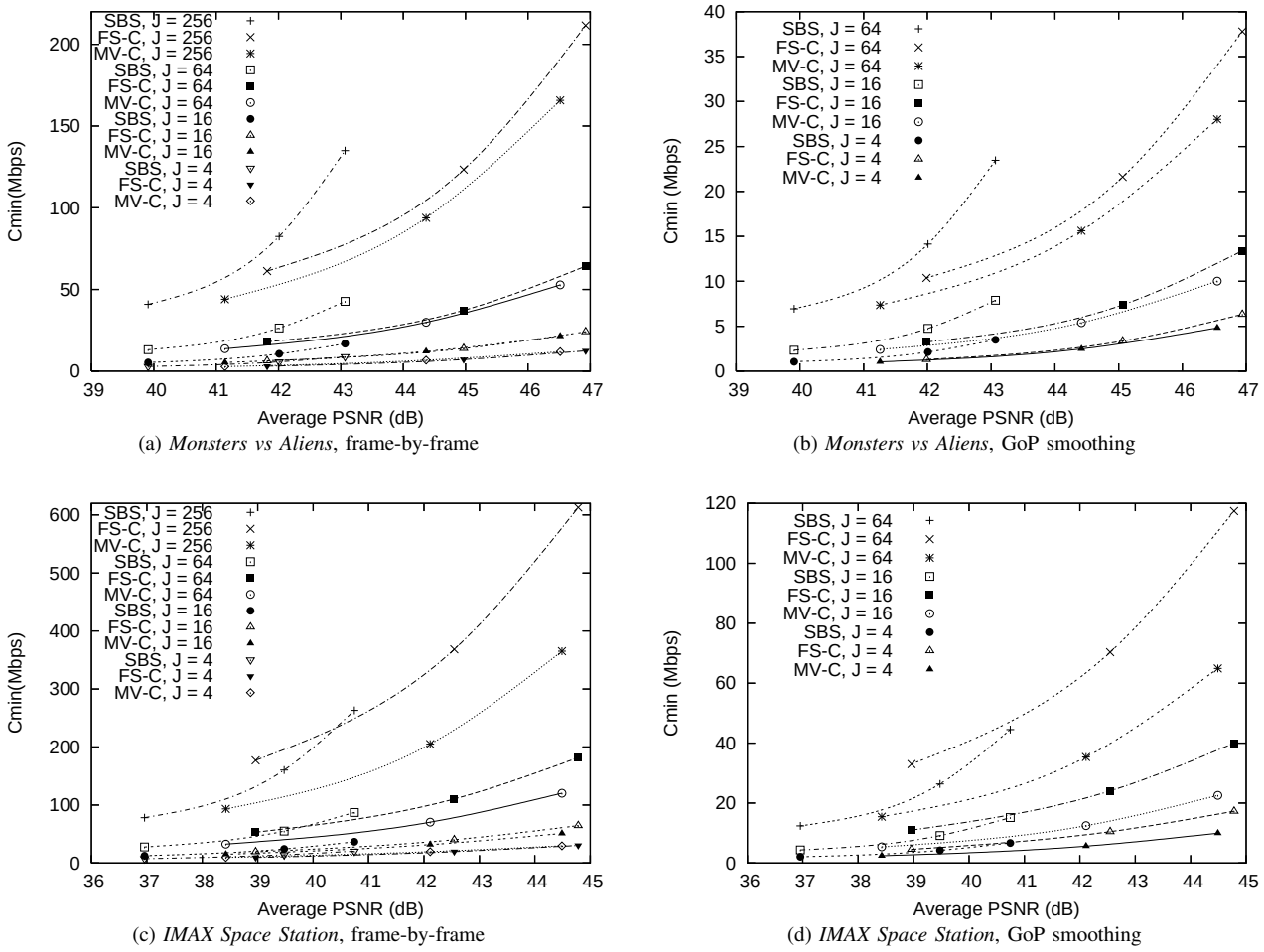
(e) *IMAX Space Station*, B1 (f) *IMAX Space Station*, B7

Fig. 2. VD curves for different representation formats and streaming approaches for B1 and B7 GoP patterns.

for *Monsters vs Aliens*, the CoV value of 1.05 for MV-C for the middle point in Fig. 2(a) is reduced to 0.65 with GoP smoothing. Similarly, the corresponding CoV value of 1.51 for *IMAX Space Station* (Fig. 2(c)) is reduced to 0.77. The CoV reductions are less pronounced for the FS representation: the middle CoV value of 0.81 for FS-C in Fig. 2(a) is reduced to 0.58, while the corresponding CoV value of 0.82 in Fig. 2(c) is reduced to 0.70.

V. Statistical Multiplexing Evaluations

In this section, we conduct statistical multiplexing simulations to examine the impact of the 3D video representations on the bandwidth requirements for streaming with minuscule loss probabilities [44]. For the MV and FS representations, we



(a) *Monsters vs Aliens*, frame-by-frame

(b) *Monsters vs Aliens*, GoP smoothing

(c) *IMAX Space Station*, frame-by-frame

(d) *IMAX Space Station*, GoP smoothing

Fig. 3. Required minimum link transmission bit rate $C_{\min}$ to transmit J streams with an information loss probability $P_{\text{loss}}^{\text{info}} \leq \epsilon = 10^{-5}$. GoP structure B1 with one B frame between I and P frames.

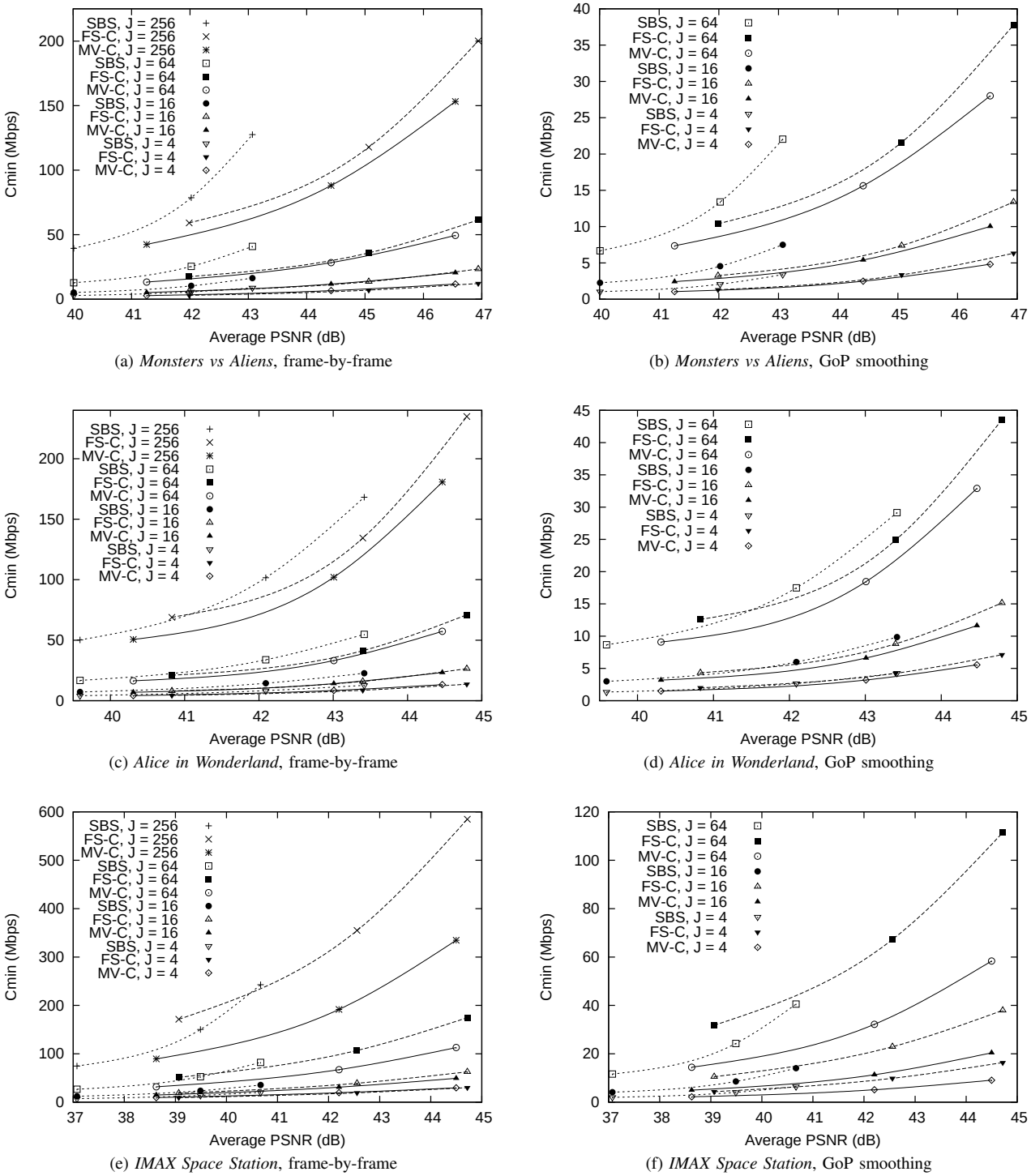
consider the combined (C) streaming approach where the pair of frames for each frame index m is aggregated and the GoP smoothing approach.

1) Simulation Setup: We consider a single "bufferless" statistical multiplexer [36], [44], [45] which reveals the fundamental statistical multiplexing behaviors without introducing arbitrary parameters, such as buffer sizes, cross traffic, or multi-hop routing paths. Specifically, we consider a link of transmission bitrate C [bit/s], preceded by a buffer of size C/f [bit], i.e., the buffer holds as many bits as can be transmitted in one frame period of duration $1/f$. We let J denote the number of 3D video streams fed into the buffer. Each of the J streams in a given simulation is derived from the same encoded 3D video sequence; whereby, each stream has its own starting frame that is drawn independently, uniformly, and randomly from the set $[1, 2, \ldots, M]$. Starting from the selected starting frame, each of the J videos places one encoded frame of the SBS representation (multiview frame of the MV-C or FS-C representation) into the multiplexer buffer in each frame period. If the number of bits placed in the buffer in a frame period exceeds C/f, then there is loss. We count the number of lost bits to evaluate the information loss probability $P_{\text{loss}}^{\text{info}}$ [44] as the proportion of the number of lost bits to the number of bits placed in the multiplexer buffer. We conduct many independent replications of the stream multiplexing, each replication simulates the transmission of M frames (with "wrap-around" to the first frame when the end of the video is reached) for each stream, and each replication has a new independent set of random starting frames for the J streams.

2) Evaluation Results: We conducted two types of evaluations. First, we determined the maximum number of streams $J_{\max}$ that can be transmitted over the link with prescribed transmission bit rate $C = 10,\ 20$, and 40 Mb/s such that $P_{\text{loss}}^{\text{info}}$ is less than a prescribed small $\epsilon = 10^{-5}$. We terminated a simulation when the confidence interval for $P_{\text{loss}}^{\text{info}}$ was less than 10 % of the corresponding sample mean.

Second, we estimated the minimum link capacity $C_{\min}$ that accommodates a prescribed number of streams J while keeping $P_{\text{loss}}^{\text{info}} \leq \epsilon = 10^{-5}$. For each $C_{\min}$ estimate, we performed 500 runs, each consisting of 1000 independent video streaming simulations. We discuss in detail the representative results from this evaluation of $C_{\min}$ for a given number of streams J. The results for the evaluation of $J_{\max}$ given a fixed link capacity C indicate the same tendencies.

We observe from Figs. 4(a), (c), and (e) that for small numbers of multiplexed streams, namely $J = 4$ and 16 streams



(a) *Monsters vs Aliens*, frame-by-frame

(b) *Monsters vs Aliens*, GoP smoothing

(c) *Alice in Wonderland*, frame-by-frame

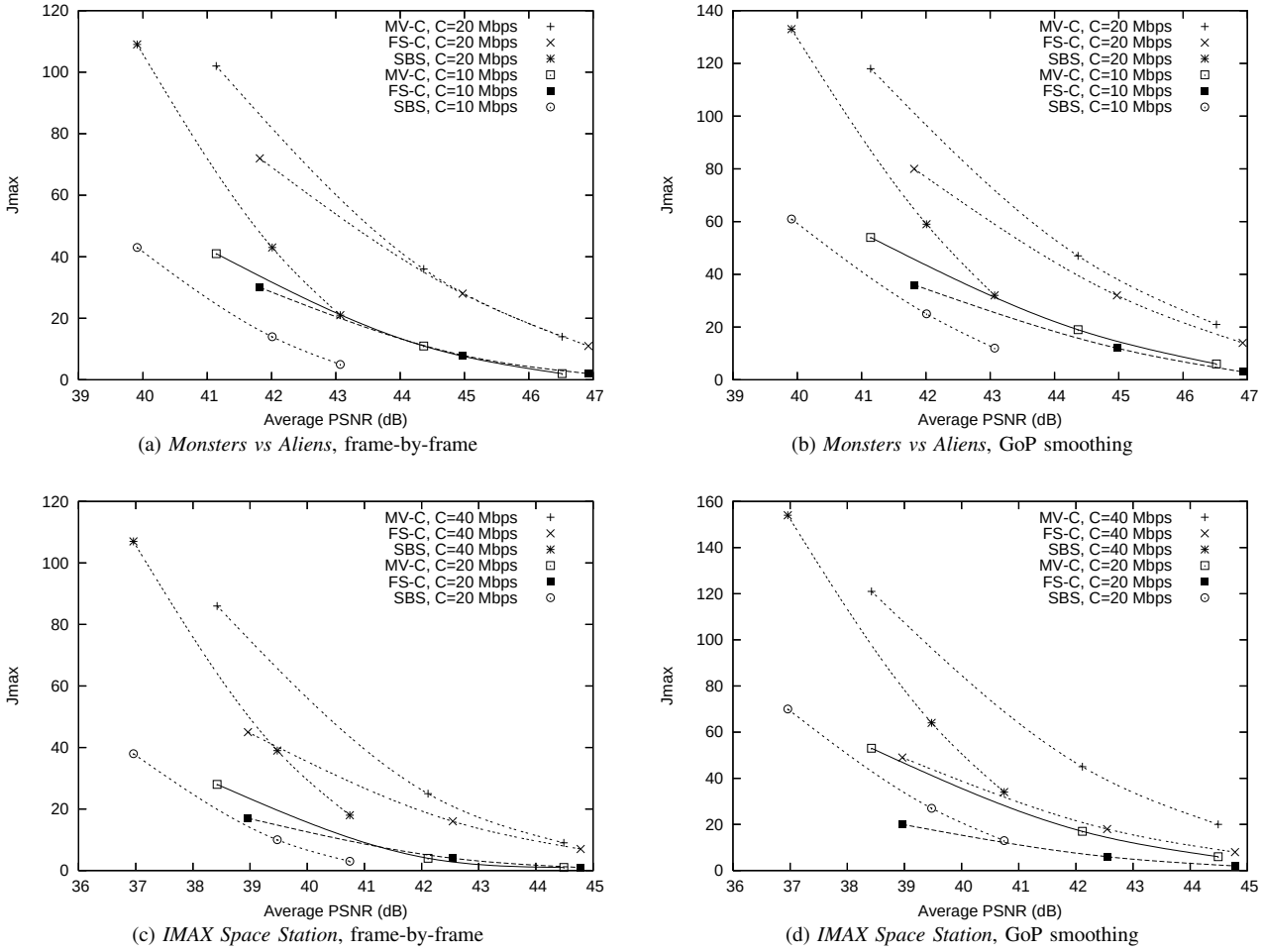
(d) *Alice in Wonderland*, GoP smoothing

(e) *IMAX Space Station*, frame-by-frame

(f) *IMAX Space Station*, GoP smoothing

Fig. 4. Required minimum link transmission bit rate $C_{\min}$ to transmit J streams with an information loss probability $P_{\text{loss}}^{\text{info}} \leq \epsilon = 10^{-5}$. GoP structure B7 with seven B frames between I and P frames.

for *Monsters vs Aliens* and *Alice in Wonderland*, as well as $J = 4$ streams for *IMAX Space Station*, the MV and FS representations require essentially the same transmission bitrate. Even though the MV representation and encoding has higher RD efficiency, i.e., lower average bit rate for a given average PSNR video quality, the higher MV traffic variability makes statistical multiplexing more challenging, requiring the same transmission bit rate as the less RD efficient FS representation (which has lower traffic variability). We further observe from Figs. 4(a), (c) and (e) that increasing the statistical multiplexing effect by multiplexing more streams, reduces the effect of the traffic variability, and, as a result, reduces the required transmission bit rate $C_{\min}$ for MV-C relative to FS-C.



(a) *Monsters vs Aliens*, frame-by-frame

(b) *Monsters vs Aliens*, GoP smoothing

(c) *IMAX Space Station*, frame-by-frame

(d) *IMAX Space Station*, GoP smoothing

Fig. 5. Maximum number of supported streams with an information loss probability $P_{\text{loss}}^{\text{info}} \leq \epsilon = 10^{-5}$ for given link transmission bit rate C. GoP structure B1 with one B frame between I and P frames.

We observe from Figs. 4(b), (d), and (f) that GoP smoothing effectively reduces the MV traffic variability such that already for small numbers of multiplexed streams, i.e., a weak statistical multiplexing effect, the required transmission bitrate for MV is less than that for FS.

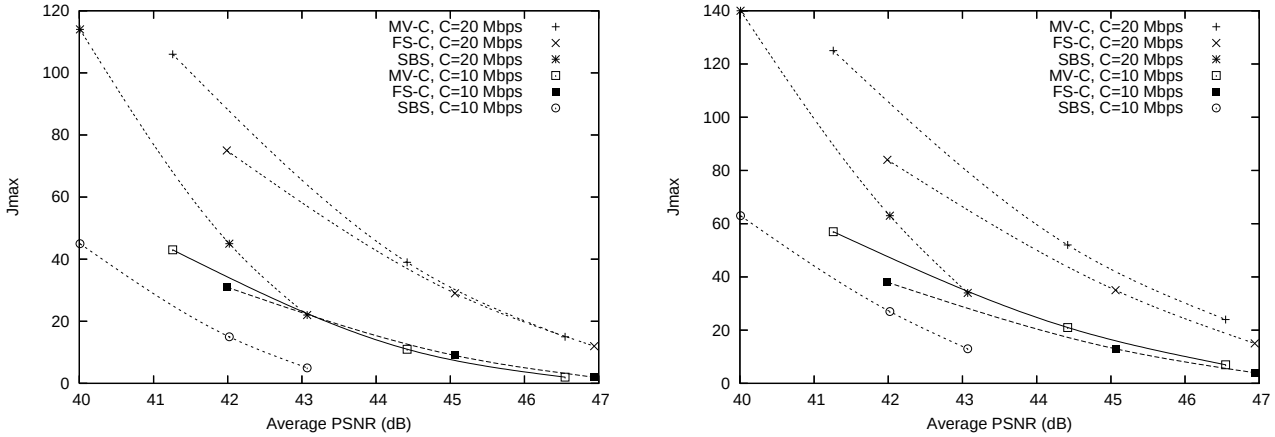
VI. Conclusion and Future Work

We have compared the traffic characteristics and fundamental statistical multiplexing behaviors of state-of-the-art multiview (MV) 3D video representation and encoding with the frame sequential (FS) and side-by-side (SBS) representations encoded with state-of-the-art single-view video encoding. We found that the SBS representation, which permits transmission of two-view video with the existing single-view infrastructure, incurs significant PSNR quality degradations compared to the MV and FS representations due to the sub-sampling and interpolation involved with the SBS representation. We found that when transmitting small numbers of streams without traffic smoothing, the higher traffic variability of the MV encoding leads to the same transmission bitrate requirements as the less RD efficient FS representation with single-view coding. We further found that to reap the benefit of the more RD efficient MV representation and coding for network transport, traffic smoothing or the multiplexing of many streams in large transmission systems is required.

There are many important directions for future research on the traffic characterization and efficient network transport of encoded 3D video, and generally multiview video. One direction is to develop and evaluate smoothing and scheduling mechanisms that consider a wider set of network and receiver constraints, such as limited receiver buffer or varying wireless link bitrates, or collaborate across several ongoing streams [46], [47]. Another avenue is to exploit network and client resources, such as caches or cooperating peer clients for efficient delivery of multiview video services. Broadly speaking, these effective transmission strategies are especially critical when relatively few video streams are multiplexed as, for instance, in access networks, e.g., [48]–[50], and metro networks, e.g., [51]–[54]. Moreover, the challenges are especially pronounced in networking scenarios in support of applications with tight real-time constraints, such as gaming [55]–[57] and real-time conferencing and tele-immersion [58]–[60].

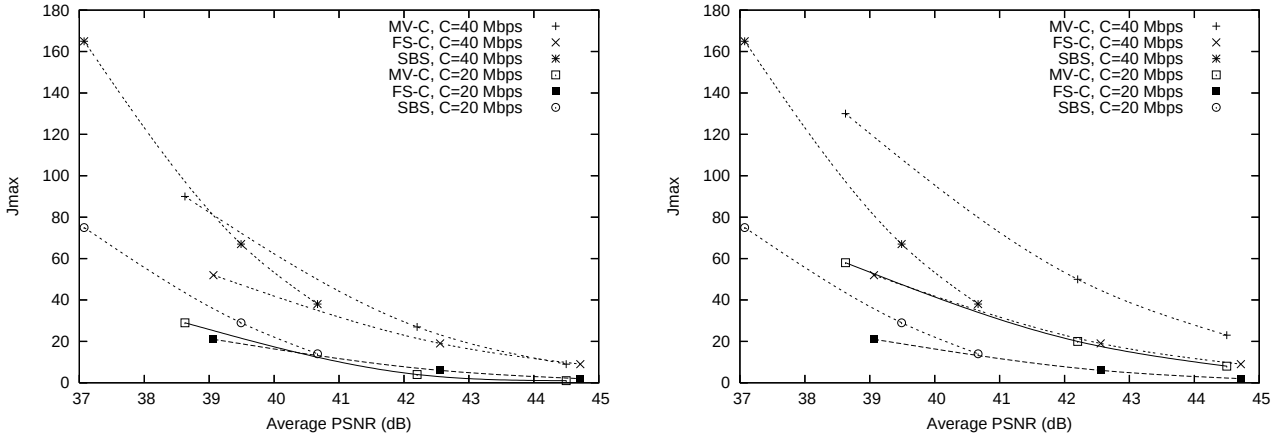
References

[1] A. Pulipaka, P. Seeling, M. Reisslein, and L. Karam, "Traffic and statistical multiplexing characterization of 3-D video representation formats," *IEEE Trans. Broadcasting*, vol. 59, no. 2, pp. 382–389, 2013.



(a) *Monsters vs Aliens*, frame-by-frame

(b) *Monsters vs Aliens*, GoP smoothing



(c) *IMAX Space Station*, frame-by-frame

(d) *IMAX Space Station*, GoP smoothing

Fig. 6. Maximum number of supported streams with an information loss probability $P_{\text{loss}}^{\text{info}} \leq \epsilon = 10^{-5}$ for given link transmission bit rate C. GoP structure B7 with seven B frames between I and P frames.

[2] P. Merkle, K. Muller, and T. Wiegand, "3D video: acquisition, coding, and display," *IEEE Trans. Consum. Electron.*, vol. 56, no. 2, pp. 946–950, May 2010.

[3] J. Morgade, A. Usandizaga, P. Angueira, D. de la Vega, A. Arrinda, M. Velez, and J. Ordiales, "3DTV roll-out scenarios: A DVB-T2 approach," *IEEE Trans. Broadcast.*, vol. 57, no. 2, pp. 582–592, Jun. 2011.

[4] A. Vetro, A. M. Tourapis, K. Muller, and T. Chen, "3D-TV content storage and transmission," *IEEE Transactions on Broadcasting*, vol. 57, no. 2, Part 2, pp. 384–394, 2011.

[5] A. Vetro, W. Matusik, H. Pfister, and J. Xin, "Coding approaches for end-to-end 3D TV systems," in *Proc. Picture Coding Symp.*, 2004.

[6] K. Willner, K. Ugur, M. Salmimaa, A. Hallapuro, and J. Lainema, "Mobile 3D video using MVC and N800 internet tablet," in *Proc. of 3DTV Conference*, May 2008, pp. 69–72.

[7] Y. Chen, Y.-K. Wang, K. Ugur, M. M. Hannuksela, J. Lainema, and M. Gabbouj, "The emerging MVC standard for 3D video services," *EURASIP Journal on Advances in Signal Processing*, 2009.

[8] A. Vetro, T. Wiegand, and G. J. Sullivan, "Overview of the stereo and multiview video coding extensions of the H.264/MPEG-4 AVC standard," *Proc. of the IEEE*, vol. 99, no. 4, pp. 626–642, Apr. 2011.

[9] G. Akar, A. Tekalp, C. Fehn, and M. Civanlar, "Transport methods in 3DTV—a survey," *IEEE Trans. Circuits and Sys. for Video Techn.*, vol. 17, no. 11, pp. 1622–1630, Nov. 2007.

[10] 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, Apr. 2011.

[11] "DIOMEDES. [Online]." http://www.diomedes-project.eu.

[12] T. Schierl and S. Narasimhan, "Transport and storage systems for 3-D video using MPEG-2 systems, RTP, and ISO file format," *Proceedings of the IEEE*, vol. 99, no. 4, pp. 671–683, Apr. 2011.

[13] Y. Zikria, S. Malik, H. Ahmed, S. Nosheen, N. Azeemi, and S. Khan, "Video Transport over Heterogeneous Networks Using SCTP and DCCP," in *Wireless Networks, Information Processing and Systems*, ser. Communications in Computer and Information Science. Springer Berlin Heidelberg, 2009, vol. 20, pp. 180–190.

[14] 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 *Proc. of 3DTV Conference*, May 2009, pp. 1–4.

[15] E. Kurutepe, M. Civanlar, and A. Tekalp, "Client-driven selective streaming of multiview video for interactive 3DTV," *IEEE Trans. Circuits Sys. Video Techn.*, vol. 17, no. 11, pp. 1558–1565, Nov. 2007.

[16] D. Wu, L. Sun, and S. Yang, "A selective transport framework for delivery MVC video over MPEG-2 TS," in *Proc. of IEEE Int. Symp. on Broadband Multimedia Systems and Broadcasting*, Jun. 2011, pp. 1–6.

[17] P. Seeling and M. Reisslein, "Video transport evaluation with H.264 video traces," *IEEE Communications Surveys and Tutorials*, vol. 14, no. 4, pp. 1142–1165, Fourth Quarter 2012.

[18] A. Alheraish, S. Alshebeili, and T. Alamri, "A GACS modeling approach for MPEG broadcast video," *IEEE Transactions on Broadcasting*, vol. 50, no. 2, pp. 132–141, Jun. 2004.

[19] N. Ansari, H. Liu, Y. Q. Shi, and H. Zhao, "On modeling MPEG video traffics," *IEEE Trans. Broadcast.*, vol. 48, no. 4, pp. 337–347, Dec. 2002.

[20] D. Fiems, B. Steyaert, and H. Bruneel, "A genetic approach to Markovian characterisation of H.264/SVC scalable video," *Multimedia Tools and Applications*, vol. 58, no. 1, pp. 125–146, May 2012.

[21] A. Lazaris and P. Koutsakis, "Modeling multiplexed traffic from H.264/AVC videoconference streams," *Computer Communications*, vol. 33, no. 10, pp. 1235–1242, Jun. 2010.

[22] K. Shuaib, F. Sallabi, and L. Zhang, "Smoothing and modeling of video transmission rates over a QoS network with limited bandwidth connections," *Int. Journal of Computer Networks and Communications*, vol. 3, no. 3, pp. 148–162, May 2011.

[23] L. Rossi, J. Chakareski, P. Frossard, and S. Colonnese, "A non-stationary Hidden Markov Model of multiview video traffic," in *Proceedings of IEEE Int. Conf. on Image Processing (ICIP)*, 2010, pp. 2921–2924.

[24] H. Alshaer, R. Shubair, and M. Alyafei, "A framework for resource dimensioning in GPON access networks," *Int. Journal of Network Management*, vol. 22, no. 3, pp. 199–215, May/June 2012.

[25] J.-W. Ding, C.-T. Lin, and S.-Y. Lan, "A unified approach to heterogeneous video-on-demand broadcasting," *IEEE Transactions on Broadcasting*, vol. 54, no. 1, pp. 14–23, Mar. 2008.

[26] L. Qiao and P. Koutsakis, "Adaptive bandwidth reservation and scheduling for efficient wireless telemedicine traffic transmission," *IEEE Trans. Vehicular Technology*, vol. 60, no. 2, pp. 632–643, Feb. 2011.

[27] T. H. Szymanski and D. Gilbert, "Internet multicasting of IPTV with essentially-zero delay jitter," *IEEE Transactions on Broadcasting*, vol. 55, no. 1, pp. 20–30, Mar. 2009.

[28] J. Cosmas, J. Loo, A. Aggoun, and E. Tsekleves, "Matlab traffic and network flow model for planning impact of 3D applications on networks," in *Proceedings of IEEE Int. Symp. on Broadband Multimedia Systems and Broadcasting (BMSB)*, Mar. 2010.

[29] N. Manap, G. DiCaterina, and J. Soraghan, "Low cost multi-view video system for wireless channel," in *Proceedings of IEEE 3DTV Conference*, May 2009.

[30] "JVC 3D Techn." http://www.jvc.eu/3d_monitor/technology/video.html.

[31] D. Broberg, "Infrastructures for home delivery, interfacing, captioning, and viewing of 3-D content," *Proceedings of the IEEE*, vol. 99, no. 4, pp. 684–693, Apr. 2011.

[32] E.-K. Lee, Y.-K. Jung, and Y.-S. Ho, "3D video generation using foreground separation and disocclusion detection," in *Proceedings of IEEE 3DTV Conference*, Tampere, Finnland, Jun. 2010.

[33] M. Lukacs, "Predictive coding of multi-viewpoint image sets," in *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing*, Tokyo, Japan, 1986, pp. 521–524.

[34] "JSVM encoder." http://ip.hhi.de/imagecom_G1/savce/downloads/SVC-Reference-Software.htm.

[35] G. Saygili, C. G. Gurler, and A. Tekalp, "Evaluation of asymmetric stereo video coding and rate scaling for adaptive 3D video streaming," *IEEE Trans. Broadcasting*, vol. 57, no. 2, pp. 593–601, Jun. 2011.

[36] P. Seeling and M. Reisslein, "The rate variability-distortion (vd) curve of encoded video and its impact on statistical multiplexing," *IEEE Transactions on Broadcasting*, vol. 51, no. 4, pp. 473–492, Dec. 2005.

[37] M. Reisslein, J. Lassetter, S. Ratnam, O. Lotfallah, F. Fitzek, and S. Panchanathan, "Traffic and quality characterization of scalable encoded video: a large-scale trace-based study, Part 1: overview and definitions," Arizona State Univ., Tech. Rep., 2002.

[38] G. Van der Auwera and M. Reisslein, "Implications of smoothing on statistical multiplexing of H. 264/AVC and SVC video streams," *IEEE Trans. on Broadcasting*, vol. 55, no. 3, pp. 541–558, Sep. 2009.

[39] "Microsoft Research 3D Test sequences." http://research.microsoft.com/en-us/um/people/sbkang/3dvideodownload/.

[40] G. Van der Auwera, P. David, and M. Reisslein, "Traffic and quality characterization of single-layer video streams encoded with the H.264/MPEG-4 advanced video coding standard and scalable video coding extension," *IEEE Trans. Broadcast.*, vol. 54, no. 3, pp. 698–718, Sep. 2008.

[41] S. Chikkerur, V. Sundaram, M. Reisslein, and L. Karam, "Objective video quality assessment methods: A classification, review, and performance comparison," *IEEE Trans. Broadcast.*, vol. 57, no. 2, pp. 165–182, Jun. 2011.

[42] K. Seshadrinathan, R. Soundararajan, A. Bovik, and L. Cormack, "Study of subjective and objective quality assessment of video," *IEEE Trans. Image Processing*, vol. 19, no. 6, pp. 1427–1441, Jun. 2010.

[43] N. Ozbek, A. Tekalp, and E. Tunali, "Rate allocation between views in scalable stereo video coding using an objective stereo video quality measure," in *Proc. of IEEE ICASSP*, 2007, pp. I–1045–I–1048.

[44] J. Roberts and U. Mocci and J. Virtamo, *Broadband Network Traffic: Performance Evaluation and Design of Broadband Multiservice Networks, Final Report of Action COST 242, Lecture Notes in Computer Science Vol. 1155*. Springer Verlag, 1996.

[45] S. Racz, T. Jakabfy, J. Farkas, and C. Antal, "Connection admission control for flow level QoS in bufferless models," in *Proceedings of IEEE INFOCOM*, Mar. 2005, pp. 1273–1282.

[46] S. Bakiras and V. Li, "Maximizing the number of users in an interactive video-on-demand system," *IEEE Trans. on Broadcasting*, vol. 48, no. 4, pp. 281–292, Dec. 2002.

[47] M. Reisslein and K. Ross, "High-performance prefetching protocols for VBR prerecorded video," *IEEE Network*, vol. 12, no. 6, pp. 46–55, Nov./Dec. 1998.

[48] F. Aurzada, M. Scheutzow, M. Reisslein, N. Ghazisaidi, and M. Maier, "Capacity and delay analysis of next-generation passive optical networks (NG-PONs)," *IEEE Trans. Commun.*, vol. 59, no. 5, pp. 1378–1388, May 2011.

[49] H. Song, B. W. Kim, and B. Mukherjee, "Long-reach optical access networks: A survey of research challenges, demonstrations, and bandwidth assignment mechanisms," *IEEE Communications Surveys and Tutorials*, vol. 12, no. 1, pp. 112–123, 1st Quarter 2010.

[50] J. Zheng and H. Mouftah, "A survey of dynamic bandwidth allocation algorithms for Ethernet Passive Optical Networks," *Optical Switching and Networking*, vol. 6, no. 3, pp. 151–162, Jul. 2009.

[51] A. Bianco, T. Bonald, D. Cuda, and R.-M. Indre, "Cost, power consumption and performance evaluation of metro networks," *IEEE/OSA J. Opt. Comm. Netw.*, vol. 5, no. 1, pp. 81–91, Jan. 2013.

[52] M. Maier and M. Reisslein, "AWG-based metro WDM networking," *IEEE Commun. Mag.*, vol. 42, no. 11, pp. S19–S26, Nov. 2004.

[53] M. Scheutzow, M. Maier, M. Reisslein, and A. Wolisz, "Wavelength reuse for efficient packet-switched transport in an AWG-based metro WDM network," *IEEE/OSA Journal of Lightwave Technology*, vol. 21, no. 6, pp. 1435–1455, Jun. 2003.

[54] M. Yuang, I.-F. Chao, and B. Lo, "HOPSMAN: An experimental optical packet-switched metro WDM ring network with high-performance medium access control," *IEEE/OSA Journal of Optical Communications and Networking*, vol. 2, no. 2, pp. 91–101, Feb. 2010.

[55] M. Bredel and M. Fidler, "A measurement study regarding quality of service and its impact on multiplayer online games," in *Proc. NetGames*, 2010.

[56] F. Fitzek, G. Schulte, and M. Reisslein, "System architecture for billing of multi-player games in a wireless environment using GSM/UMTS and WLAN services," in *Proc. ACM NetGames*, 2002, pp. 58–64.

[57] C. Schaefer, T. Enderes, H. Ritter, and M. Zitterbart, "Subjective quality assessment for multiplayer real-time games," in *Proc. ACM NetGames*, 2002, pp. 74–78.

[58] G. Kurillo and R. Bajcsy, "3D teleimmersion for collaboration and interaction of geographically distributed users," *Virtual Reality*, vol. 17, no. 1, pp. 29–43, 2013.

[59] M. Pallot, P. Daras, S. Richir, and E. Loup-Escande, "3D-live: live interactions through 3D visual environments," in *Proc. Virtual Reality Int. Conf.*, 2012.

[60] R. Vasudevan, Z. Zhou, G. Kurillo, E. Lobaton, R. Bajcsy, and K. Nahrstedt, "Real-time stereo-vision system for 3D teleimmersive collaboration," in *Proc. IEEE ICME*, 2010, pp. 1208–1213.