# Video Coding with Multiple Descriptors and Spatial Scalability for Device Diversity in Wireless Multi-hop Networks

P. Seeling and M. Reisslein

## Abstract

Providing video services to heterogeneous clients in wireless ad hoc networks is particularly challenging as ($i$) the heterogeneous client processing and display capabilities typically prevent clients from processing and displaying the same encoded video information, and ($ii$) wireless connections typically suffer from bandwidth variability and transmission errors. We jointly address these two challenges by introducing a novel video coding strategy which combines multiple description coding, in particular *temporal* descriptors, with layered spatial coding. Our *spatial scalable descriptor* coding strategy enables heterogeneous clients in wireless multi-hop networks with path diversity to receive preencoded video streams over independent paths and to process only that amount of encoded video information that suits their processing and display capabilities. We evaluate our coding strategy through simulation experiments with a highly dynamic video sequence. We find that our coding strategy with two descriptors improves the quality of the received video by approximately 4 dB and cuts the quality variability approximately in half compared to layered coding.

## Index Terms

Ad-hoc networks, wireless networks, multiple description coding, video streaming

## I. Introduction

Enabling video services to mobile clients is challenging, as devices that are capable of receiving video streams typically vary in their connection characteristics, processing capabilities, and display sizes. In addition, the delivery of video over wireless ad hoc networks is complicated by the time-varying and location dependent error characteristics and outages of the various wireless links in the network. In this paper we address these two challenges by introducing a novel spatial scalable descriptor coding strategy. Our strategy combines multiple description coding with hierarchical layered coding in a novel way. We employ spatial scalable layered coding in this paper to fix ideas; the proposed approach can be employed with SNR scalable layered encoding in analogous fashion. In brief, in our approach the frames in the original source video are assigned in round-robin fashion to the descriptor streams, e.g., with two descriptors every other frame is assigned to a given descriptor stream. Each descriptor stream is then independently coded using spatial scalable layered encoding, e.g., with two layer encoding, a base layer and an enhancement layer are created for each descriptor stream.

We evaluate the proposed spatial scalable descriptor coding strategy in the context of wireless ad hoc networks with multiple independent paths between the server and the client, which are commonly available for video streaming, see for instance [1], [2]. We find that the proposed coding strategy results in significantly improved video quality at the client while concurrently enabling flexible support for video services to heterogeneous clients. In particular, clients with low display capabilities can receive and display only the base layers of some or all of the descriptors (depending on their bandwidth) while clients with high display capabilities can receive and display the base layer along with the enhancement layers of some or all of the descriptors. Our simulation results indicate that concurrently with enabling this flexibility in the video services, our coding strategy allows for the effective exploitation of multi-path diversity in ad-hoc networks. In typical wireless transmission scenarios and for highly dynamic video content, our strategy improves the received video quality by approximately 4 dB and cuts the quality variability approximately in half compared to layered encoding, both for clients with low and high display capabilities.

This paper is organized as follows. In the following subsection we review related work. In Section II, we introduce the encoding process for the generation of spatial scalable descriptors. In Section III, we describe the considered transmission system used for the streaming of single and multiple encoded descriptions. In Section IV, we describe the metrics used for the evaluation of the streamed descriptor(s) and provide simulation results in Section V. We conclude in Section VI.

*A. Related Work*

Both multiple description coding (MDC) [3], where multiple independently decodable streams are constructed, and hierarchical layered coding, where one base layer and one or more enhancement layers (which require the lower layers for decoding) are created, have attracted significant interest for providing video services to wireless clients. The performance of layered coding and multiple description coding for wireless video services have been compared in a number of studies, see for instance [1], [4], [5], [6]. In contrast, in this paper we combine multiple description coding and layered coding in an overall coding strategy.

The combination of layered coding with multiple description coding has received relatively limited attention to date. In [7] the video is coded into multiple descriptions using the priority encoding transmission technique which combines layered video source coding with unequal erasure protection. The individual descriptors are then further split into multiple layers using special unequal erasure protection and overlapping techniques. In contrast, in our approach, the video is directly split into multiple descriptors, which are then individually coded with standard layered coding techniques. In [8] the video is encoded into one base layer and two enhancement layers. The enhancement layers require the base layer for decoding, but are multiple descriptions of the original enhancement layer. As a consequence there is no hierarchy among the enhancement layers as in standard layered encoding where the higher enhancement layers require all the lower enhancement layers for decoding. Our approach fundamentally differs from [8] in that our coding strategy creates two base layers, each with one enhancement layer. The distinct advantage of our approach is that it enables path diversity in ad hoc networks in conjunction with multiple description coding for the base layers, e.g., for video services to clients with low display capabilities.
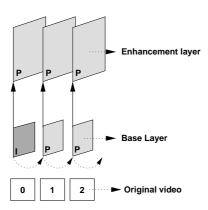
In the wavelet domain, research for combining scalable coding with multiple descriptors has attracted recent research interest, see, e.g., [9] and the references therein. The layering approaches in these previous research efforts are aimed at adaptation of the descriptors to varying channel conditions during transmission. Our approach is different as we consider fixed encodings with predetermined resolutions in the DCT domain, which are suitable for integration into content delivery networks, as shown in [10].

The combination of MDC video coding with streaming over multiple paths was studied in [2] and more recently in [11] for ad hoc wireless networks. The state-based approach in [12] was evaluated for this scenario in [13]. Path diversity is one method of decreasing the probability of simultaneous transmission errors in multiple descriptors and allows additionally to reduce the video traffic load on individual links or hops in the wireless network. The selection of such paths was studied in, e.g., [14].

## II. SPATIAL SCALABLE DESCRIPTOR CODING

For the combination of spatial scalability and multiple descriptors, we employ *temporal* descriptors, which are also referred to as multiple states (and have been studied in isolation in [12]). With the considered temporal descriptors, the original video stream is split into $D$ descriptor streams. Descriptor stream $d,\ d = 1, \ldots, D$, contains the frames $d,\ d + D,\ d + 2D, \ldots$ of the original video frame sequence. To achieve the spatial scalability of the descriptors we encode each individual descriptor stream $d,\ d = 1, \ldots, D$, as follows. We employ the Group of Pictures (GoP) pattern *IPPPPPPPPPPPI* .... With this GoP pattern, every P frame depends on the previous I or P frame only. We employ spatial scalable encoding, to encode a given descriptor stream $d$ into one base layer and one enhancement layer. We encode every enhancement layer frame only with respect to the corresponding base layer frame. We employ the described GoP pattern and dependencies between base and enhancement layer to fix ideas for our illustrations of the proposed spatial scalable descriptor coding strategy and for our experiments. The proposed strategy is general and can be applied with any GoP pattern and form of dependencies between base and enhancement layer. The only requirement is that the different descriptor streams are coded independently and that no dependencies between distinct descriptor streams are introduced in the encoding process.

The conventional spatial scalable layered video coding with the employed GoP pattern and dependencies between base and enhancement layer frames is illustrated in Fig. 1. Note that this conventional scalable encoding is a degenerate from of our spatial scalable descriptor coding with a single descriptor stream, i.e., $D = 1$. We illustrate the (non-degenerate) spatial scalable descriptor coding with $D = 2$ in Fig. 2. Note that the descriptor streams are temporally interleaved, i.e., every second video frame belongs to a given descriptor stream. Also observe that each descriptor stream is independently coded using the employed GoP pattern and dependencies between base and enhancement layer frames. Note however, that there are no dependencies introduced by the coding between the two descriptor streams. Furthermore, note that the frame rate in frames per second (fps) for each individual descriptor is $1/D$ times the frame rate of the original video, i.e., half the frame rate of the original video (or equivalently the single descriptor encoding) for the considered case of $D = 2$ descriptors. Combining the $D$ descriptors at play
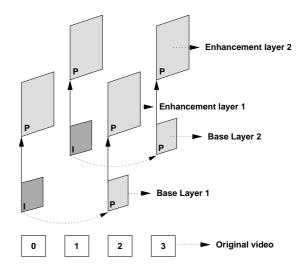
Fig. 1. Conventional spatial scalable video coding, which is a degenerate case of spatial scalable descriptor coding with $D = 1$ descriptor, and which we also refer to as single descriptor (SD) encoding.



Fig. 2. Spatial scalable descriptor video coding with $D = 2$ temporal descriptors, which we also refer to as multiple descriptor (MD) encoding.

out time gives the original frame rate. With this encoding approach, clients can selectively process descriptors and layers so as to suit their heterogeneous processing and display capabilities.

The proposed spatial scalable descriptor coding strategy comes at a cost (overhead). This encoding overhead is due to the increased temporal spacing between the consecutive frames in a given descriptor stream. As a result of the increasing temporal inter-frame distance, the temporal correlations in a given descriptor stream are smaller compared to the correlations between consecutive frames in the original video frame sequence. Hence, the motion estimation and compensation in the interframe coding of a given descriptor stream result in larger motion vectors and encoded prediction errors, which in turn results in larger frame sizes.

## III. TRANSMISSION SYSTEM

The general setup of the video transmission system under consideration is illustrated in Fig. 3. We simulate one
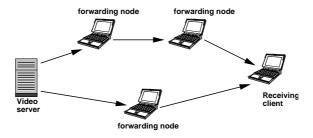


Fig. 3. Video transmission system with two independent paths.

path $p$ for each descriptor $D$ ($p = D$), whereby we assume the individual path characteristics to be independent of each other for more than a single path, as is common, see e.g., [1], [2].

Let $H_p$ denote the number of hops along each path $p$. The number of hops on each path $H_p$ is initially drawn randomly (uniform) from the interval $[1, 5]$. Let $h_p$ denote an individual hop of the previously drawn hops $H_p$ on path $p$, i.e., $h_p = 1, \ldots, H_p$. Let $b_D(h_p)$ denote the bandwidth of hop $h$ on path $p$ for the streaming of $D$ descriptors overall (or $p$ paths in total). The bandwidth of each hop for a single descriptor is initially drawn randomly (uniform) from $b_{D=1}(h_p) \in [1, 2, 5.5, 11]$ Mbps. To ensure a fair comparison, we set the *available bandwidth* for the transmission of a video descriptor on a given hop to the bandwidth divided by the number of descriptors, i.e.,

$$b_D(h_p) = \frac{1}{D} b_{D=1}(h_p). \tag{1}$$

For two descriptors the available bandwidth for each descriptor is thus half the bandwidth for that hop or $b_2(h_p) = \frac{1}{2} b_1(h_p)$.

During transmission time, bandwidth changes on each hop can occur at the boundaries of time slots of length one second to model different node mobility levels. This bandwidth change is modeled as random walk among the aforementioned bandwidths and a bandwidth of 0. Let $P(\text{up})$ and $P(\text{down})$ denote the probability for an increase or decrease in available bandwidth for the next second. The *bandwidth change probability* is assumed to be identical for an increase and decrease in bandwidth, i.e., $P(\text{up}) = P(\text{down})$. Clearly, the probability of a hop bandwidth to remain constant is then calculated as $P(\text{remain}) = 1 - P(\text{up}) - P(\text{down})$. (Note that we do not differentiate between these probabilities for individual hops.) If the bandwidth of one hop along a path $p$ is 0, the path collapses. We define the time needed for finding and setting up a new path as *retransmission timeout* denoted by $r$. After the retransmission timeout, the described algorithm is repeated.

During the transmission, we employ the Gilbert-Elliot (GE) channel model to simulate bit errors introduced on each hop of the path as given in [15]. At path setup time, the initial GE states are drawn randomly for each hop as either *good* or *bad*. A bit error on a packet triggers a retransmission from the preceding node on the path, which we assume to be triggered by IEEE802.11b data link layer acknowledgment frames, assuming data link layer retransmissions. Each frame is thus sent in a send-and-wait fashion from one intermediate node to the next. We only consider transmission delays in our simulation.

The frames that could not reach the receiving client are discarded at the second boundary, i.e., we assume that the source node(s) maintain(s) a frame (or packet) buffer for one second and discard(s) the unsent content every second. The same discarding scheme is replicated for each forwarding node along the path. This ensures that the forwarding nodes do not have to have large buffer or processing requirements. This scheme is therefore applicable to any device class that can constitute the multi-hop network under consideration.

In our experiments, we employ the standard spatial scalable MPEG-4 codec with settings such that decoding a base layer frame gives a down-sampled version of the video in the QCIF format ($176{\times}144$ pixels). If the enhancement layer frame is added, the full size CIF format ($352{\times}288$ pixels) is available.

## IV. QUALITY EVALUATION METRICS

We evaluate the received video stream by decoding the low resolution QCIF and the high resolution CIF versions of the received video. If the base layer data for a frame was not successfully received, we display again the previous correctly received base layer frame. If the enhancement layer data of a frame is not available, we upsample and display the corresponding base layer frame. The resulting objective quality is calculated as peak signal-to-noise ratio (PSNR) for the luminance (Y) component, as the human eye is most sensitive to this component. Let $q$ denote the quantization scale setting for the encoding of the video and $N$ denote the total number of frames. We denote an individual pixel value in the $n$th encoded (and subsequently decoded) video frame by $F_n^q(x,y)$ and its encoded, *transmitted*, and subsequently decoded (and up-sampled in case of lost enhancement layer frames) counterpart by $f_n^q(x,y)$. Let $X$ and $Y$ denote the resolution in pixels of the source video. Let $\Delta$ denote the distance between the currently displayed frame of the (recombined) video and the last successfully received frame as illustrated in Fig. 4 for the base and enhancement layer case. We then compare $F_n(x,y)$ to $f_{(n-\Delta)}^q(x,y)$ and calculate the mean squared
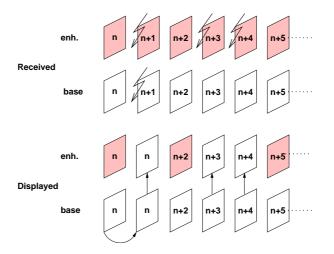


Fig. 4.  Video frame replication and upsampling for base and enhancement layer video frames.

transmission error (MSE) for an individual frame as

$$M_n^q = \frac{1}{X \cdot Y} \sum_{x=0}^{X-1} \sum_{y=0}^{Y-1} [F_n^q(x,y) - f_{(n-\Delta)}^q(x,y)]^2. \tag{2}$$

The average of the mean squared transmission errors is calculated as

$$\bar{M}_q = \frac{1}{N} \sum_{n=0}^{N-1} M_n. \tag{3}$$

From the average of the mean squared transmission errors we calculate the average transmission quality (average quality of the received video sequence) using the standard relationship between the MSE and the PSNR as

$$\bar{Q}_q = 10 \log_{10} \frac{255}{\bar{M}}. \tag{4}$$

Generally, the perceived video quality is maximized if the quality of the individual frames is maximized (i.e., the average quality is maximized) and the quality variation of the frames is minimized. To capture the quality variations we calculate the standard deviation of the received video quality as

$$\sigma_M^2(q) = \frac{1}{N-1} \sum_{n=0}^{N-1} (M_n^q - \bar{M}_q)^2 \tag{5}$$

and

$$\sigma_Q^q = 10 \log_{10} \sqrt{\sigma_M^2(q)}. \tag{6}$$

## V. RESULTS

In this section we evaluate results obtained for the video transmission of our proposed scalable descriptor coding strategy. We compare a single descriptor ($D = 1$, SD) to two descriptors ($D = 2$, MD$_1$ and MD$_2$), whereby each descriptor is streamed over an independent path. We order the frames prior the transmission so that for any given time interval of length one second, we first send all base layer frames and use the remaining bandwidth to send the enhancement layer frames. We use a 120 second sequence from the movie *The Transporter* ($N = 2880$) which shows a highly dynamic car chase with multiple scene changes. The source video is encoded at 24 frames per second for the single descriptor and 12 frames per second for each of the two considered descriptors. The video is furthermore encoded with three different quantization scale settings ($q = 5, 7, 10$). In Table I the bandwidth requirements averaged over the 120 seconds for each layer and descriptor are given. In addition we present the encoding overhead for the combined size of the two descriptors versus the size of the single descriptor. For the

TABLE I

AVERAGED BANDWIDTH REQUIREMENTS IN KBPS FOR ENCODED VIDEO STREAMS FROM *The Transporter*.

| Layer | Descr. | Quantization scale | | |
|---|---|---|---|---|
| | | $q = 5$ | $q = 7$ | $q = 10$ |
| base | SD | 355.85 | 243.82 | 163.27 |
| | MD$_1$ | 226.08 | 156.13 | 106.52 |
| | MD$_2$ | 222.72 | 156.04 | 106.48 |
| | Overhead | 0.26 | 0.28 | 0.30 |
| enh. | SD | 604.49 | 410.63 | 264.51 |
| | MD$_1$ | 305.23 | 208.73 | 136.60 |
| | MD$_2$ | 303.80 | 207.98 | 136.18 |
| | Overhead | 0.01 | 0.01 | 0.03 |

base layer we observe that the encoding overhead increases with the quantization scale $q$, which is explained as follows. The temporal differences between consecutive frames determine the encoded residues from motion estimation algorithms or intra coded information. As the overall frame size generally decreases sharply for larger quantization scales $q$, the remaining errors from the temporal offset as well as MPEG header information account for a larger fraction of the total encoded information with larger quantization scales. As result the overhead increases for multiple descriptors. For the enhancement layer, we do not observe such behavior of the encoding overhead. As described in Section II, the enhancement layer frames are only encoded with respect to the corresponding base layer frame and are therefore not affected by the encoding overhead caused by the temporal offset. The only remaining overhead in

the enhancement layer is caused by MPEG header information and very small encoding differences compared to the single descriptor.

We additionally observe from Table I that the combined size of base and enhancement layer is lower than the minimum bandwidth available for streaming (1 Mbps) for the single descriptor. In absence of errors, delays, and protocol overheads, we could thus always transmit all base and enhancement layer frames. For two descriptors and a quantization scale of $q = 5$, the encoding overhead of the base layer raises the combined bit rate of base and enhancement layer over the minimum transmission bandwidth (0.5 Mbps) for each descriptor. Even without the additional effects for sending the encoded video, we would thus not be able to stream all the enhancement layer frames completely with our transmission order. For the quantization scale settings $q = 7$ and $q = 10$, the combined bandwidths of base and enhancement layers is always lower than the minimum transmission bandwidth for each of the two descriptors.

For our following evaluation of a client that aims at receiving the base layer and a client that aims at receiving base and enhancement layer, we assume that a reconnection timeout of $r = 2$ seconds and a node mobility (modeled as bandwidth change probability) of 0.25 are a basic scenario and present averages of 50 simulation runs. We then extend our evaluation for different retransmission timeout values and different node mobilities for the quantization scale $q = 7$.

### A. Base Layer Receiving Client

In this section we consider a client that aims at reception of the base layer stream only. The results for a single descriptor (SD), i.e., $D = 1$, and two descriptors (MD), i.e., $D = 2$, are given in Table II. We observe that the

TABLE II

AVERAGED BASE LAYER RESULTS FOR DIFFERENT QUANTIZATION SCALES $q$ FOR *The Transporter*, RETRANSMISSION TIMEOUT $r = 2$ AND BANDWIDTH CHANGE PROB = 0.25, FIXED.

|  | Descr. | Quantization scale | | |
|---|---|---|---|---|
|  |  | $q = 5$ | $q = 7$ | $q = 10$ |
| Correct frames | SD | 0.85 | 0.84 | 0.85 |
| (relative) | MD | 0.83 | 0.81 | 0.84 |
| Transmission quality | SD | 22.25 | 22.25 | 22.48 |
| $\bar{Q}_q$ | MD | 26.45 | 26.56 | 27.33 |
| Stddev. quality | SD | 5.93 | 5.80 | 5.81 |
| $\sigma_Q^q$ | MD | 2.69 | 2.90 | 2.28 |

transmission quality as defined in Eq. (4) is at least 4dB higher for the proposed spatial scalable descriptor coding scheme than for an individual descriptor. The difference to the single descriptor increases with the quantization scale. We also observe that the standard deviation of the received video quality is less than half for the proposed scheme in comparison to the single descriptor. We furthermore observe that the proportion of correctly received frames is slightly smaller for the MD approach. These effects are explained as follows. The smaller proportion of correctly delivered frames with the MD approach is due to each descriptor having only half the bandwidth available on its path compared to the bandwidth of the single descriptor. (We do this to give a video stream the same amount of total bandwidth with each of the compared approaches.) The increased transmission quality and reduced quality variation despite the slightly smaller proportion of correctly received frames with the MD approach are due to the combined effects of multiple description coding and multi-path streaming. In case a transmission path collapses in the MD approach, the loss in frame quality is relatively small as only every second frame of the original stream is lost. In contrast, if a path collapses in the SD approach then several consecutive frames are lost until the path is re-established after the retransmission timeout $r$ (which is 2 sec corresponding to 48 frames in the considered scenario). Due to the correlations between consecutive frames, the MSE between the original frames and the frames redisplayed for a single frame period (as is the case with the MD approach) is lower than for an extended period of redisplayed frames as in the single descriptor case. Also, note that the probability of both paths collapsing simultaneously in the MD approach is small compared to the probability of a single path collapse.

We next examine the influence of different retransmission timeouts and node mobilities for a quantization scale of $q = 7$ as illustrated with a 90% confidence interval in Figs. 5 and 6. We observe that the transmission qualities for the proposed scheme are always significantly higher than for the single descriptor. The transmission qualities for single and two descriptors decline roughly linear with the increasing timeout and node mobility level, with a tendency to sub-linearity in the case of two descriptors. This corroborates earlier findings showing that multiple descriptors perform better than layered encodings especially in highly error-prone environments. For an increase in
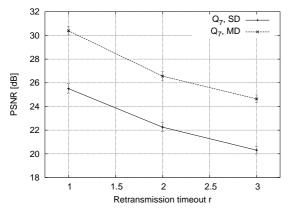
Fig. 5. Transmission quality for *The Transporter* base layer (QP=7) with different reconnection timeout durations.
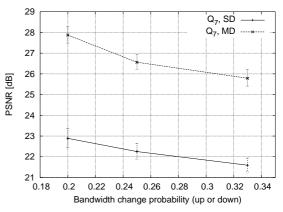


Fig. 6. Transmission quality for *The Transporter* base layer (QP=7) with different node mobilities modeled by the bandwidth change probability.

retransmission timeout $r$ from 1 to 3 sec we observe a stronger effect on the transmission quality than for the increase in node mobility from 0.2 to 0.33. This is explained by the direct influence of the retransmission timeout value on the number of lost frames, whereas increased node mobility not always causes path collapses and retransmission timeouts.

We thus conclude that the client aiming at the reception of the base layer benefits from the proposed scheme by achieving a significantly higher average transmission quality and smaller quality variation than would be possible by the single descriptor approach.

### B. Base and Enhancement Layer Receiving Client

For a client with a high display capability, the reception of enhancement layer frames in addition to the corresponding base layer frames is most desirable. Recall that the available bandwidth is shared between the base and enhancement layer for each descriptor so that the base layer frames are always transmitted before the enhancement layer frames and that the base layer frame is always upsampled if no enhancement layer frame is available. The results for the enhancement layer are given in Table III. We observe that similar to the base layer only case the

TABLE III

AVERAGED ENHANCEMENT LAYER RESULTS FOR DIFFERENT ENCODING QUALITIES.

|  | Descr. | Quantization scale | | |
|---|---|---|---|---|
|  |  | $q = 5$ | $q = 7$ | $q = 10$ |
| Correct frames | SD | 0.74 | 0.79 | 0.84 |
| (relative) | MD | 0.7 | 0.78 | 0.82 |
| Transmission quality | SD | 22.18 | 22.20 | 22.46 |
| $\bar{Q}_q$ | MD | 26.19 | 26.23 | 27.20 |
| Stddev. quality | SD | 5.92 | 5.80 | 5.82 |
| $\sigma_Q^q$ | MD | 2.70 | 2.42 | 2.30 |

proposed scheme outperforms the single descriptor by at least 4dB in transmission quality. With an increase in the quantization scale, the difference in transmission quality increases as well. The effect of increased quantization scales on the transmission quality is lower than observed for the base layer in Table II. For the quality variation, the quality variability, measured as standard deviation, is approximately cut approximately in half when two instead of one descriptors are used. With increased quantization scale we observe especially for the MD approach a decrease in quality variability. We furthermore observe that the proportion of correctly received frames is lower than in the base layer case for SD and MD approaches and increasing with an increase in the quantization scale. For the difference of the proportion of correctly received frames between SD and MD and approach we observe a generally diminishing trend. These effects are explained as follows. The reduction in the proportion of correctly received frames is caused by the larger size of the enhancement layer frames compared to the base layer frames. The limited bandwidth of the MD approach compared to the SD approach thus limits the successful transmission of all enhancement layer frames. This is similar to the base layer only case, but the effect of limited bandwidth is stronger on the enhancement layer case, as we always transmit the base layer before then enhancement layer. For the higher transmission quality as

well as the reduced transmission quality variability, we refer to the reasoning for the base layer only case. We note, however, that for the displayed CIF-sized video we consider here, the upsampled base layer frames are used and not the base layer frames themselves, see Fig. 4.

We next examine the results for different reconnection timeouts $r$ and bandwidth change probabilities for a quantization scale of $q = 7$ as illustrated with a 90% confidence interval in Figs. and . We observe that the
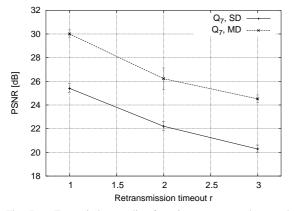


Fig. 7. Transmission quality for *The Transporter* base and enhancement layer (QP=7) with different reconnection timeout durations.
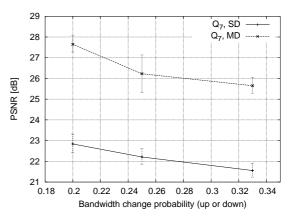


Fig. 8. Transmission quality for *The Transporter* base and enhancement layer (QP=7) with different node mobilities modeled by the bandwidth change probability.

transmission qualities are always significantly higher for two descriptors compared to the single descriptor. The decline in transmission qualities with increased retransmission timeouts $r$ and node mobilities exhibit the same behavior we previously observed in the base layer only case. We again observe the stronger influence of the retransmission timeout $r$ in comparison to the influence of the node mobility on the transmission quality. The explanation for the effect observed here is similar to the base layer only case. In extension to the base layer, however, here the bandwidth changes have an additional effect, as especially for larger numbers of necessary retransmissions of the base layer frames, the minimum available bandwidth along a path determines the number of enhancement layer frames that can be transferred.

For the client aiming at the reception of the base and enhancement layer, we thus conclude that the benefits from the proposed scheme compared to the single descriptor approach allow a transmission quality similar to that obtained for the base layer case.

## VI. Conclusion and Outlook

We have introduced a novel spatial scalable descriptor coding strategy which combines layered spatial video coding with *temporal* descriptors for providing video services to heterogeneous devices in wireless networks. Our simulations using a difficult source video sequence and path diversity typical for the wireless multi-hop network scenarios indicate that our spatial scalable descriptor coding provides significantly higher perceptual video quality than the comparable single layered approach. Our strategy enables the effective streaming of different preencoded video resolutions to heterogeneous wireless client devices without adjustments at the time of streaming. By allocating only a fraction of the available bandwidth to each of the descriptors, our approach additionally contributes to load balancing in wireless networks.

Further enhancements to the decoding part of our strategy are possible by implementation of state-based recovery (similar to the recovery technique developed in [12]) and by post-processing of the up-sampled base layer frames. In addition the sending order of frames has an impact on the perceived quality given different content dynamics and an impact on content provider benefits given different client economics and is currently under evaluation.

## Acknowledgment

## REFERENCES

[1] S. Lin, Y. Wang, S. Mao, and S. Panwar, "Video transport over ad-hoc networks using multiple paths," in *Proc. of IEEE International Symposium on Circuits and Systems*, Phoenix, AZ, May 2002, pp. 57–60.

[2] N. Gogate, D.-M. Chung, S. S. Panwar, and Y. Wang, "Supporting image and video applications in a multihop radio environment using path diversity and multiple description coding," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 12, no. 9, pp. 777–792, 2002.

[3] V. Goyal, "Multiple description coding: Compression meets the network," *IEEE Signal Processing Magazine*, vol. 18, no. 5, pp. 74–93, Sept. 2001.

[4] A. R. Reibman, Y. Wang, X. Qui, Z. Jiang, and K. Chawla, "Transmission of multiple description and layered video over an EGPRS wireless network," in *Proc. of IEEE International Conference on Image Processing (ICIP)*, vol. 2, Vancouver, BC, Canada, Sept. 2000, pp. 139–139.

[5] Y. Wang, S. Panwar, S. Lin, and S. Mao, "Wireless video transport using path diversity: Multiple description vs layered coding," in *Proc. of IEEE International Conference on Image Processing (ICIP)*, Rochester, NY, Sept. 2002, pp. 21–24.

[6] Y.-C. Lee, J. Kim, Y. Altunbasak, and R. M. Mersereau, "Layered coded vs. multiple description coded video over error-prone networks," *Signal Processing: Image Communication*, vol. 18, no. 5, pp. 337–356, May 2003.

[7] P. A. Chou, H. J. Wang, and V. N. Padmanabhan, "Layered multiple description coding," in *Proc. Packet Video Workshop*, Nantes, France, May 2003.

[8] L. P. Kondi, "A rate-distortion optimal hybrid scalable/multiple-description video codec," in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Montreal, Canada, May 2004.

[9] M. van der Schaar and D. S. Turaga, "Multiple description scalable coding using wavelet-based motion compensated temporal filtering," in *Pro. IEEE Inernational Conference on Image Processing (ICIP)*, Barcelona, Spain, Sept. 2003, pp. 489–492.

[10] J. Apostolopoulos, T. Wong, W.-T. Tan, and S. Wee, "On multiple description streaming with content delivery networks," in *Proc. of IEEE INFOCOM 2002*, vol. 3, New York, NY, June 2002, pp. 1736–1745.

[11] S. Mao, S. Lin, S. S. Panwar, Y. Wang, and E. Celebi, "Video transport over ad hoc networks: Multistream coding with multipath transport," *IEEE Journal on Selected Areas in Communications*, vol. 21, no. 10, pp. 1721–1737, Dec. 2003.

[12] J. G. Apostolopoulos, "Error resilient video compression through the use of multiple states," in *Proc. of IEEE International Conference on Image Processing (ICIP)*, vol. 3, Vancouver, BC, Canada, Sept. 2000, pp. 352–355.

[13] ——, "Reliable video communication over lossy packet networks using multiple state encoding and path diversity," in *SPIE Visual Communications and Image Processing (VCIP)*, San Jose, CA, Jan. 2001, pp. 392–409.

[14] A. C. Begen, Y. Altunbasak, and O. Ergun, "Multi-path selection for multiple description encoded video streaming," in *Proc. of IEEE International Conference on Communications (ICC)*, vol. 3, Anchorage, AK, May 2003, pp. 1583–1589.

[15] J.-P. Ebert and A. Willig, "A Gilbert-Elliot bit error model and the efficient use in packet level simulation," Technical University of Berlin, Berlin, Germany, Tech. Rep., 1999.