

Overhead and Quality Measurements for Multiple Description Coding for Video Services

Frank H.P. Fitzek[§], Bařak Can[§], Ramjee Prasad[§], Marcos Katz^{§§}

[§]Center for TeleInFrastructure (CTIF), Aalborg University

Niels Jernes Vej 12, 9220 Aalborg Øst, Denmark; ph: +45 9635 8678, e-mail: [f]bc|prasad@kom.aau.dk

^{§§} SAMSUNG ELECTRONICS CO., LTD.; e-mail: marcos.katz@samsung.com

Abstract—Multiple description coding (MDC) has received a lot of interest lately, because it allows to split one source of information in multiple entities, where each of the entities is decodable in a stand-alone fashion. The more of these entities are received the more information of the original source can be restored. This opens the door for new applications in many fields of mobile communications. Unfortunately the multiple description coding comes with the price of higher bandwidth requirements. This overhead is caused by the coding process itself and the related network transport. In this work we present a solid investigation of the overhead arising from the splitting process for video coding. The overhead caused by the underlying network protocols is highlight. Furthermore we investigate the quality of the decoded video in absence of specific streams. A main conclusion we draw here is that we have to consider the scenario carefully whether or not multiple description coding should be chosen over single description coding.

Keywords—MDC, overhead measurements, PSNR, H.26L, H.263, CIF, QCIF

I. INTRODUCTION AND RELATED WORK

MDC has emerged as an attractive coding scheme for robust transmission over channels with transient bursty error characteristics such as wireless systems during a deep fade. MDC is a great solution to this problem at the price of redundancy. The packet losses, which is most probable and can be correlated over wireless links, can cause severe degradation in the received video quality.

The application of MDC can be classified into two main groups, namely path diversity and heterogeneous terminals. We give a short introduction of the related work.

a. Path diversity: Various network infrastructures such as multi-cast services in cellular networks, multi-hop networks and wired networks utilize MDC to provide path diversity. The study carried out in [1] combats errors encountered in the wireless channel by coding the highly compressed video data into multiple independently decodable streams sent over different paths. An unbalanced MDC scheme to provide path diversity is proposed in [2]. This scheme creates unbalanced Multiple Description (MD) streams by adjusting the frame rate, quantization or spatial resolution of each stream and can achieve unbalanced rates of up to 2:1. Authors examine the effectiveness of unbalanced MDC and path diversity for transmission of *foreman* and *bus* sequences. The recovered video quality for unbalanced rates of up to 1.75:1 and 1.76:1 are compared with that of SD scheme. Their results show that the proposed unbalanced MDC scheme is more robust to bursty errors and network congestions than SD scheme.

The effectiveness of the combination of MDC with multiple path transport (MPT) for video and image transmission over multi-hop wireless links is studied in [3]. Some applications have high bandwidth and stringent reliability requirements. For supporting these kind of applications, the sender may need to send the data stream over several paths to the destination [3]. The study in [3] addresses this need and the authors compare their proposed MDC-MPT system with a system which uses layered coding and asymmetrical paths for the base and enhancement layers. Apart from the standard Quarter Common Intermediate Format (QCIF) and Common Intermediate Format (CIF) sequences given in [4],

authors present average Picture Signal to Noise Ratio (PSNR) for MDC for *Susie*, *Football* and *Flower Garden* video sequences. In [5], an MDC system for multi-hop networks is proposed. The application and network layers cooperate to provide more robustness against severe network conditions. The authors propose a multi-path selection method that chooses a set of paths to achieve more robust media transmission. The media content is transmitted over these intelligently selected multiple paths instead of shortest path or maximally disjoint paths. The simulation results in this paper show an average PSNR improvement of up to 8.1 dB than conventional MDC scheme.

The application of MDC for multi-cast networks has been addressed in [6]. Multi-cast services addresses to a wide variety of receiver terminals having different bandwidth requirements. The study in [6] explores *layered multiple description codes* where base layer descriptions are transmitted to users having low bandwidth capacities whereas the enhancement layer descriptions are transmitted to users having high bandwidth capacities.

MDC is also being exploited in wired infrastructures for increased link capacity and reliability. For example, the wired connection to the base stations can have more capacity if MDC is used. MDC divides the data stream into several streams whereby the probability of buffer flow in wired links decreases [7]. Multiple path transport (MPT) scheme is proposed for wired links to provide path diversity. The source coding and traffic splitting is jointly designed for MPT schemes [8]. MDC and layered coding (LC) are the two options for source coding in MPT scheme [8].

b. Heterogeneous terminals: MDC is exploited to support heterogeneous terminals. 4G mobile systems are envisioned to offer wireless services to a wide variety of mobile terminals ranging from cellular phones and Personal Digital Assistants (PDAs) to Laptops [9]. The flexibility of the bandwidth assigned to each descriptor and the number of descriptors assigned to end users makes MDC a very attractive coding scheme for 4G networks. Terminals use the same bandwidth and decide on the number of descriptors that they can receive [10].

In [11], a new MDC technique called Multiple Description Scalable Coding (MDSC) is proposed. MDSC can simply be described as a combination of MDC and scalable coding. This technique addresses receiving device characteristics and bandwidth variations of the channel and also enables tradeoffs between throughput, redundancy and complexity which is not possible with non-scalable MDC schemes. In MDSC, the number and the composition of descriptions are changed dynamically to make the proposed system very robust to changing channel characteristics [11].

After making a literature search, we realized that there is a shortage of performance results of MDC with all of the video sequences given in [4]. We present the performance investigations of MDC for wireless networks using 12 of the video sequences given in [4]. The goal of this paper is to show the drawbacks of MDC in terms of overhead and quality measurements. In literature, authors usually give exact results for the overhead generated by the MDC process. For example, in [12], it was reported that

Tab. 1. YUV QCIF/CIF

video sequence name	video format	number of video frames	info
bridge--close [4]	QCIF and CIF	2000	Charles Bridge (Karlv most) oldest bridge in Prague.
bridge--far [4]	QCIF and CIF	2101	Charles Bridge (Karlv most) oldest bridge in Prague.Far view.
carphone	QCIF	382	Man talking at the phone.
claire [4]	QCIF	494	Female talking to the camera.
container [4]	QCIF	300	Ship leaving the harbor.
foreman [4]	QCIF	400	Man speaking to the camera.
highway [4]	CIF	2000	Driving over highway.
mobile [4]	CIF	300	Train is moving
paris [4]	CIF	1065	Two people talking to each other
salesman [4]	QCIF	449	Salesman in his office
silent [4]	QCIF	300	Woman doing sign language.
tempeste [4]	CIF	260	Moving Cam

the overall data rate of multiple streaming is 11 % higher than single-streaming. However, for better understanding and more solid investigation into the overhead generated by MDC, no exact overhead values should be given as it is strongly dependent on the video content. We are investigating the overhead generated by encoding and network for IPv6. In the following section our measurement setup and the terminology used is introduced.

II. OUR MEASUREMENT SETUP AND TERMINOLOGY

As MDC splits the video stream into multiple descriptors, we have chosen the frame based approach using a splitter entity. The splitter, as given in Figure 1, takes the raw video sequences and splits them into J sub-sequences, ($J > 1$), such that the i -th sub-sequence contains picture i , $J + i$, $2J + i$, and so on. These J subsequences are fed into the video encoder, i.e. H.26L or H.263.

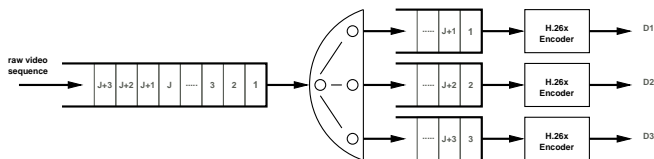


Fig. 1. Construction of multiple substreams and descriptions

A single pixel can be represented in two different color spaces: RGB and YUV. In the RGB color space, the pixel is composed of three colors: Red (R), Green (G) and Blue (B). In the YUV color space, the pixel is represented by its luminance (Y), chrominance (U), and saturation (V) [13]. Our investigations are based on raw video sequences in the YUV format, which are widely used in the research community and available at [4]. QCIF and CIF are two different video formats with 176x144 pixels and 352x288 pixels, respectively [13]. Since wireless links are subject to bandwidth limitations, we concentrate on the QCIF and CIF formats. The video sequences with frame length and a short information are given in Table 1 for the QCIF and CIF formats.

In the first step, we investigate the overhead produced by MDC. We define the overhead as the amount of data of the splitted streams in comparison to a single stream. We distinguish between encoder and network overhead. Then we investigate how sensitive the network overhead is in dependency of the number of sub-streams generated by the splitter. For the investigation of the error-prone link, we present quality measurements versus percentage of successfully received substreams.

For the performance evaluation we refer to the following terminology:

- *Single descriptor stream*: The single descriptor stream is the encoded version with $J = 1$.
- *Multiple descriptor stream*: The multiple descriptor stream is the encoded version with $J > 1$. Thus multiple sub-streams exists, where each of the sub-streams is decodable for itself.

- *Sub-streams*: One encoded entity of the splitted video sequence.
- *Bandwidth*: B_{SD} represents the bandwidth of single description coding. B_i is the bandwidth of each descriptor i .
- *Overhead*: The overhead is defined as the amount of data by which the splitted streams of one video sequence are increased in comparison to the single stream. The overhead is generated by encoding process and network. Unless the network overhead is not stated, we refer to the overhead produced by the encoding process. If network overhead is stated, it includes both the encoding overhead and network overhead. The overhead θ is defined as

$$\theta = \frac{\sum_{i=1}^J B_i}{B_{SD}} - 1 \quad (1)$$

Further overhead such as the network overhead are addressed in Section IV.

- *Video quality*: We use the PSNR. The PSNR represents the objective video quality each video frame by a single number. A video frame is composed by $N \cdot M$ pixels (where M is the length and N the height of the frame). Each pixel is presented by one luminance value and a set of pixels by two chrominance values. Because the human eye is more sensible to the change in luminance we focus only on this parameter. The mean squared error (MSE) and the PSNR in decibels are computed by the following two equations [13]:

$$MSE = \frac{\sum_{i,j} [f(i,j) - F(i,j)]^2}{N \cdot M} \quad (2)$$

$$PSNR = 20 \cdot \log_{10} \left(\frac{255}{\sqrt{MSE}} \right), \quad (3)$$

where $f(i,j)$ represents the original source frame and $F(i,j)$ the reconstructed possibly error-prone frame containing N by M pixels.

- *Quantization Parameter (QP)*: The Discrete Cosine Transform (DCT) is used to convert a block of pixels into a block of transform coefficients. The transform coefficients represent the spatial frequency components of the original block. Then the quantization is achieved as follows [13]:

- 1) Coefficients with an absolute value smaller than the quantizer threshold are set to zero (i.e., they are considered to be in the so-called dead zone).
- 2) Coefficients with an absolute value larger than or equal to the quantizer threshold are divided by twice the QP and rounded to the nearest integer.

The larger the step size, the larger the compression gain, as well as the loss of information. The trade-off between compression and decodable image quality is controlled by setting the QP and quantizer threshold [13].

III. ENCODING OVERHEAD

In this section, results showing the encoding overhead generated by multiple streaming is presented for different video sequences in CIF and QCIF formats. For the overhead calculation of H.26L encoded video streams we used a quantization parameter of 31 as default parameter¹. Later we will investigate even the impact of this parameter by varying these values between 1 and 51. We have agreed on a group of picture (GoP) structure with one I frame and eleven P frames. As for H.263 we distinguish the same two video formats, namely QCIF and CIF. In Figures 2 and 3, the overhead of

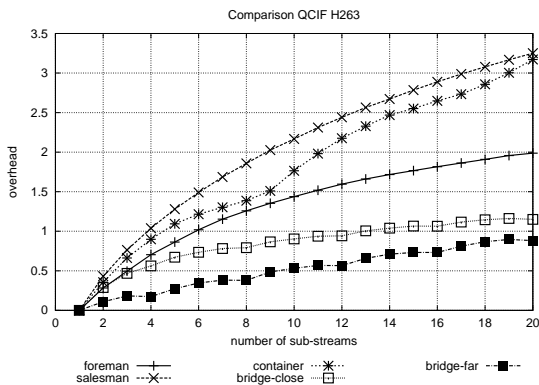


Fig. 2. Overhead of selected H.263 video sequences in the QCIF format for different number of sub-streams.

selected H.263 and H.26L video sequences in the QCIF format for different number of sub-streams is given. As it can be observed in these figures, the encoding overhead of the H.263 encoder is larger than the encoding overhead of H.26L. H.26L encoder has been developed much later than the development of H.263 [13]. Because of different encoder settings this comparison may be unfair, but it is not the main goal to compare the two encoders, but to look into the overhead. In Figure 2, while the `container` sequence has an overhead three times larger than single encoded video in case $J \geq 18$, the overhead of the `bridge--far` sequence is below 100% for all $J \leq 20$. For all video sequences the overhead is less than 50% if only two sub-streams are used.

In Figure 4 the overhead of selected H.263 video sequence in the CIF format for different number of sub-streams is given. From these figures, we report two main insights: i.) Increasing J results in a larger overhead and ii.) The overhead itself depends dramatically on the video content chosen. In contrast to the H.263 measurements given in Figures 2 and 4, the curve in Figure 3 is not that smooth. This effect can be explained by the used GoP structure. While the encoded video stream with only one descriptor consists out of exactly 1/12 I frames and 11/12 P frames, the encoded stream with multiple descriptors has a much higher ratio of I frames. The percentage of I frames, λ , in one descriptor is calculated as given in Equation 4

$$\lambda = \left\lceil \frac{N}{GoP * J} \right\rceil, \quad (4)$$

where N is the number of frames of the original video sequence, GoP is the number of frames per GoP, and J is the number of descriptors of the video sequence. The operator $\lceil x \rceil$ returns the smallest integer that is not less than its argument. For a video length of 300 frames (as for `container` and `mobile`) the percentage

¹The given QP refers to the I frame, the P frames are encoded than with QP-1, thus resulting in higher quality. This decision is based on the default settings of the decoder.

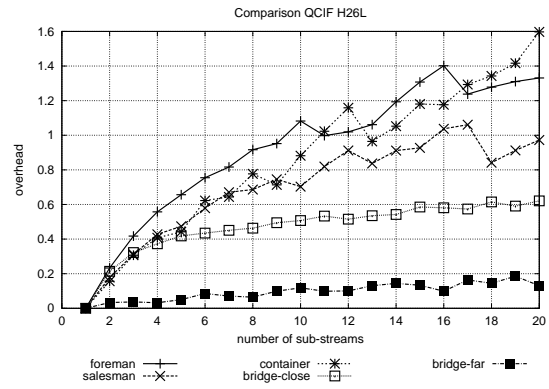


Fig. 3. Overhead of selected H.26L video sequences in the QCIF format for different number of sub-streams.

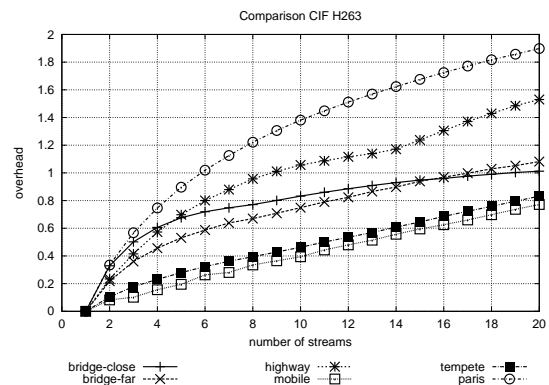


Fig. 4. Overhead of selected H.263 video sequences in the CIF format for different number of sub-streams.

of I frames increases by 44% for 12 descriptors. This explains the gaps in Figure 3. The effect vanishes with larger video sequences and lower quantization values (see investigation of the quantization parameters).

In Figures 5 and 6, the effect of QP on the generated overhead is highlighted. Obviously the data rate will decrease with larger values for QP. As QP increases from one to 21, the video compression gain increases and the quality of the video as well as the required bandwidth decreases [13]. As the QP increases above 21 up to 51, the generated overhead starts to decrease down to the level where QP equals one. It is important to note, that QP values about 31 will be used in wireless environments as it combines small data rates and high video quality. This means that the overhead by multiple streaming is very high.

IV. NETWORK OVERHEAD

In the IP world, the transmission of multiple descriptors to the receiver is mostly done by using the real time protocol (RTP) together with the real time control protocol (RTCP). RTP and RTCP protocols transport their packet via the user datagram protocol (UDP) to the underlying Internet protocol (IP) layer. As given in Figure 7, the multiple description coded streams are sent to the RTP layer by the application layer. RTP layer packetizes each data segment into an RTP packet. Each layer adds an additional overhead to each data stream that is transmitted. Using RTP/UDP/IP an overhead of 40 byte per stream has to be taken into account.

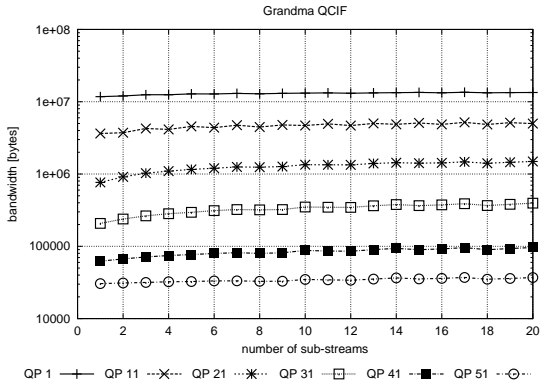


Fig. 5. Bandwidth requirements for the grandma video sequence in the QCIF format for different quantization values.

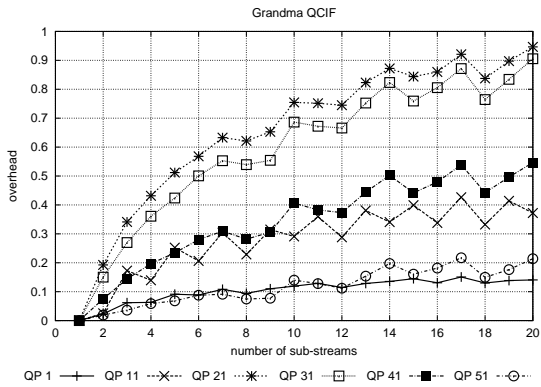


Fig. 6. Overhead for the grandma video sequence in the QCIF format for different quantization values.

As the descriptors are transmitted over the IP network, in Figure 8, the network overhead in addition to the encoding overhead for the foreman video sequence encoded at three different QP values (1, 31, and 51) is given for IP version 6. Splitting the video sequence into 20 sub-streams leads to an overhead of 1.3 introduced by the encoder. If the network overhead of IPv6 is also taken under consideration, the overhead increases up to 9.2. Thus in some scenarios the network overhead becomes dramatically larger than the overhead produced by the encoding process.

V. QUALITY MEASUREMENTS

After the overhead measurements presented in Sections III and IV, we investigate the quality of the video in dependency of successfully received sub-streams. In the following subsection, the methodology that is used for quality measurements is given.

A. Measurement Methodology

As metric for the video quality we use PSNR. If a descriptor is lost, we are assuming that all its frames are lost completely. This is a realistic scenario for the support of heterogeneous terminals and for multi-hop networks. Heterogeneous terminals listen to a subset of descriptors which is due to their bandwidth limitations. In case of multi-hop networks, if one of the hop in the transmission is completely down, than the descriptors transmitted via that hop are completely lost.

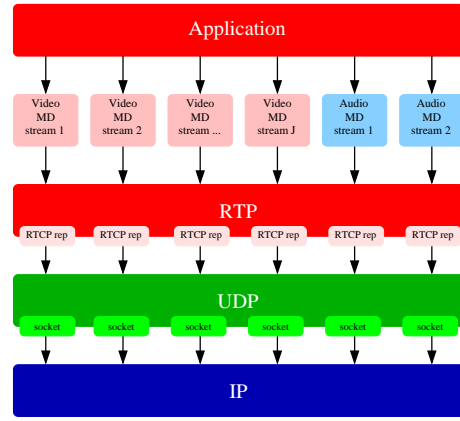


Fig. 7. Protocol stack in case MDC is used

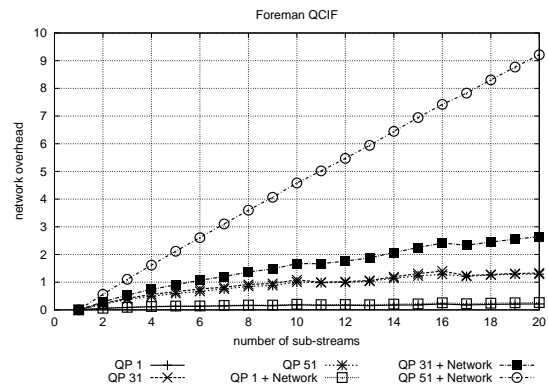


Fig. 8. The network overhead (RTP/UDP/IPv6) for the foreman video sequence and three different quantization values (namely 1, 31, and 51) for IP version 6

The differences between two pictures are taken into account pixel-wise. To calculate the PSNR values in dependency of the percentage of successfully received descriptors we select a subset of descriptors randomly and feed this to the merger. In case descriptors are lost, some frames will be missing in the video sequence. As a very simple error concealment, we are *freezing* the last successfully received frame until we have a new one. By freezing we mean that we make a copy of the last received frame for the next frame until an update frame is received. This procedure is shown in Figure 9 in case descriptor 2 is missing. In this case frame 1, J, \dots will be displayed two times.

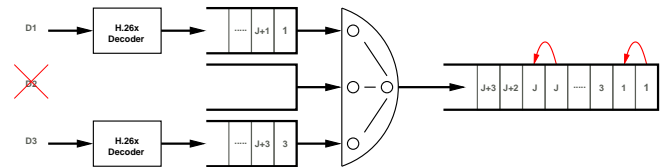


Fig. 9. Merger for $J=3$, where one descriptor is missing.

For the quality measurements we focus only the H.26L encoder. The calculation of PSNR values is done as given in [13] using the videometer tool [14]. The videometer tool calculates the PSNR values and is able to freeze video frames in case the following frames are lost. This is important to have some sort of error resilience. We have repeated the measurements multiple times with a confidence interval of 99% for the PSNR value.

B. Results

In Figure 10, the PSNR measurements using the foreman video sequence in the QCIF format versus percentage of successfully received sub-streams is given. In this figure, we are considering only one video sequence, namely foreman video sequence which includes high motion. For 33.3% of successfully received descriptors, the PSNR value is higher in the case where we are losing 1 out of 3 descriptors for $J = 3$ rather than 5 out of 15 for $J = 15$. Consequently, we observe that, the picture quality degrades for a larger value of J at the same percentage of successfully received sub-streams.

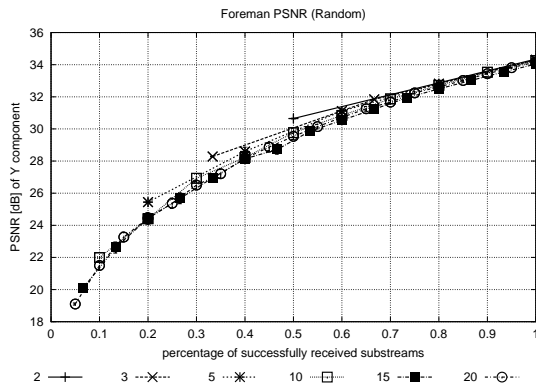


Fig. 10. PSNR measurements using the foreman video sequences in the QCIF format versus percentage of successfully received sub-streams (random approach).

In Figure 11, PSNR measurements versus percentage of successfully received sub-streams for a total of $J = 15$ descriptors are given. The results are obtained using 6 different video sequences. The foreman video sequence includes high motion. Whereas clarie and container video sequences include relatively lower motion than foreman. As we can observe in the figure, if the motion in a given video sequence is high, the slope of the PSNR degradation curve is also high. For example, for a loss of 60 % out of $J = 15$ descriptors, the PSNR degradation is 6 dB for foreman, 2 dB for claire and 0.5 dB for container video sequence. Thus, we conclude that the more the motion in a given video sequence, the more important is to receive as many descriptors as possible for a given J .

VI. CONCLUSION AND OUTLOOK

MDC is being used in many applications to provide path diversity and offers new possibilities for 4G networks. The advantages of the MDC coding mentioned above comes with the price of increased overhead. Our results presented in this paper show that the amount of overhead depends on the video content, the encoding process, number of descriptors generated and the network.

Another conclusion drawn in this paper is that the quality of a received video sequence for a given number of successfully received descriptors depends on the video content. As the motion in a video sequence increases, the effect of lost descriptors on received video quality increases. A solid and comprehensive overview of MDC is offered in this paper that was not done in any other paper before. Further results are available in our technical report [15].

ACKNOWLEDGMENT

This paper is based upon work supported by SAMSUNG, Korea.

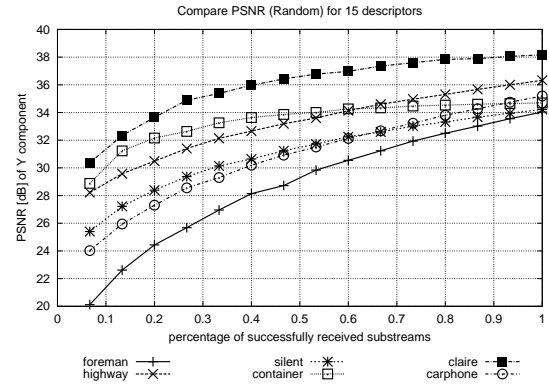


Fig. 11. Comparison of PSNR degradation for different video sequences for $J = 15$.

REFERENCES

- [1] J. G. Apostolopoulos, "Error Resilient Video Compression Through The Use of Multiple States," in *IEEE Int'l Conf. on Image Processing*, vol. 3, Vancouver, BC, Canada, September 2000, pp. 352–355.
- [2] J. G. Apostolopoulos and S. J. Wee.
- [3] 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. 777–834, September 2002.
- [4] "Video traces for network performance evaluation: YUV 4:2:0 Video Sequences," <http://trace.eas.asu.edu/yuv/yuv.html>.
- [5] A. C. Beğen, Y. Altunbaşak, and Ö. Ergün, "Multi-path Selection for Multiple Description Encoded Video Streaming," in *IEEE Int'l Conf. on Communications*, vol. 3, Anchorage, Alaska, USA, May 2003, pp. 1583–1589.
- [6] P. A. Chou, H. J. Wang, and V. N. Padmanabhan, "Layered Multiple Description Coding," in *In Proc. Packet Video Workshop*, April 2003.
- [7] Vivek K Goyal, "Multiple Description Coding: Compression Meets the Network," *IEEE SIGNAL PROCESSING MAGAZINE*, vol. 01, pp. 74–93, September 2001.
- [8] Yao Wang, Shivendra Panwar, Shunan Lin, and Shiwen Mao, "Wireless video transport using path diversity: multiple description vs layered coding," *Image Processing Proceedings*, vol. 1, pp. 1–21 – 1–24, September 2002.
- [9] R. Prasad and S. Hara, *Multicarrier Techniques for 4G Mobile Communications*, ser. Universal Personal Communications. Artech House, 2003.
- [10] Frank H.P. Fitzek, Başak Can, Ramjee Prasad, DS Park, and Youngkwon Cho, "Application of Multiple Description Coding in 4G Wireless Communication Systems," *WWRF 8 bis*, February 2004.
- [11] M. Schaar and D. S. Turaga, "Multiple Description Scalable Coding Using Wavelet-Based Motion Compensated Temporal Filtering," in *IEEE Int'l Conf. on Image Processing*, vol. 3, Barcelona, Spain, September 2003, pp. 489–492.
- [12] Y. J. Liang, E. G. Steinbach, and B. Girod, "Multi-stream Voice over IP Using Packet Path Diversity," in *IEEE 4th Workshop on Multimedia Signal Processing*, 2001.
- [13] Frank Fitzek, Patrick Seeling, and Martin Reisslein, *Wireless Internet*, March 2004, no. ISBN: 0849316316, Electrical Engineering & Applied Signal Processing Series, ch. Video Streaming in Wireless Internet.
- [14] P. Seeling, F. Fitzek, and M. Reisslein, "Videometer," *IEEE Network*, vol. 17, no. 1, p. 5, jan 2003.
- [15] F. H. P. Fitzek, "Application and Network Measurements for Multiple Description Coded Video Sequences," Aalborg University, Niels Jernes Vej 12, room A5 204 DK-9220 Aalborg Denmark, Tech. Rep. ISBN 87-90834-46-1, ISSN 0908-1224, June 2004.