Parameters optimization for a scalable multiple description coding scheme based on spatial subsampling[*]

Marco Folli and Lorenzo Favalli
Universitá degli studi di Pavia
Via Ferrata 1
27100 Pavia, Italy
marco.folli@unipv.it
lorenzo.favalli@unipv.it

Matteo Lanati
Eucentre, Universitá degli studi di Pavia
Via Ferrata 1
27100 Pavia, Italy
matteo.lanati@eucentre.it

ABSTRACT

In this work, we design a multiple description coding (MDC) system for video streams moving from the concept of spatial MDC and introducing an efficient algorithm to obtain substreams that exploits some form of scalability. We first generate four subsequences by sub-sampling, then, from these four subsequences, by jointly coding two of them, we generate two descriptions. Finally, each description is compressed independently with the recent H.264/SVC video coding standard. In order to achieve some sort of scalability, for each description, we predict one of the original subsequences from the other one via inter layer prediction, thus generating a base layer, containing one subsequence, and an enhancement layer, containing the other one. In this paper, we present some results, varying the fraction of the total rate assigned to the base layer, in order to find the better value that guarantees the optimal performances in case not all the (sub)streams are received.

Keywords

H.264/SVC, Multiple Description Coding, scalability, inter layer prediction

1. INTRODUCTION

Transmission of video sequences over both the Internet and wireless networks poses many challenges in terms of bandwidth variations and packet losses, due to congestion on the Internet or due to fading, interference and mobility of the wireless users. Recently, multiple description coding (MDC) has been studied as an approach for transmission of compressed visual information in these environments.

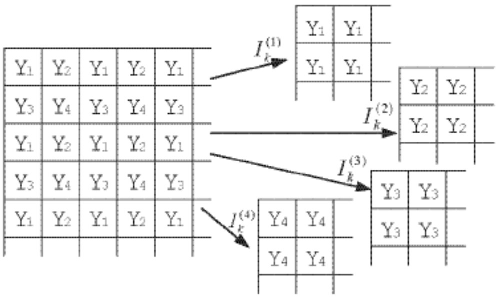

Figure 1: Example of polyphase downsampling system

In a MDC algorithm, several substreams, called descriptors or descriptions, are created from a single source, each at a lower quality than the original, and transmitted over separate channels. Differently from a scalable stream, each of these descriptors is independently decodable, and mutually refinable so that, ideally, receiving all the descriptions allows the full recovery of the single stream coded video. A comprehensive overview of various MDC techniques is provided by Goyal in [1]. A very simple MDC scheme can be obtained by the temporal splitting of the odd and even frames of a video sequence in separate, individually decodable descriptors that can be decoded using standard receivers, as described in [2]. Another simple method for MDC is based on the spatial subsampling of the original video sequence to obtain descriptions by using a polyphase subsampler along rows and columns. This scheme is called Polyphase Spatial Subsampling multiple description (PSS-MD) coding . Vitali *et al.* ([3]) show that in error prone networks such a scheme provides equal or better robustness with respect to other solutions such as Forward Error Correction (FEC), but with a lower system complexity. This is obtained thanks to the inherent redundancy due to the strong correlation among the descriptions. Reducing this redundancy is anyway mandatory to achieve better efficiency. An improvement to the PSS scheme is proposed in [4], where two of the four subsequence are predicted from the others by calculating the absolute value of the difference between "neighboring" subsequences to exploit the high correlation level between the subsampled streams.

The methods described above are very simple and virtually can be applied to every standard video coder, but

[*] Work supported by Italian Ministry of research (MUR) under grants FIRB-RBIN043TKY "Software and Communication Platforms for High-Performance Collaborative Grid" and PRIN-2006099023 PROFILES "Peer to peer beyond file sharing".

Mobimedia 2008 July 7-9, 2008, Oulu, Finland.
Copyright 2008 ICST ISBN 978-963-9799-25-7/08/07
DOI 10.4108/ICST.MOBIMEDIA2008. 3908

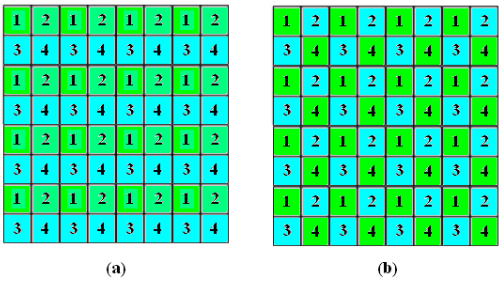

Figure 2: Image subsampling patterns. Left: *"by rows"*. Right: *"quincunx"*

they are only aimed at increasing the robustness by exploiting link diversity. To address other transmission challenges, such as bandwidth variations or device heterogeneity, a scalable approach is required. However, a scalable approach has a dual problem with respect to a multiple descriptions scheme: it lacks robustness of the stream. In order to exploit the advantage of both methods, recently some multiple descriptions scalable coding (MDSC) schemes have been proposed as an efficient hybrid solution. An example of these schemes is the SNR - MDSC algorithm proposed in [4], in which the original sequence is coded with coarse grain scalability (CGS). After that, the enhancement layer is split in two different parts: each description is then formed by sending the base layer plus only one of these parts of the enhancement layer.

The starting point of this paper is to develop an efficient mix of scalability and multiple description to take advantage of both schemes. The application scenario is one in which a scalable multiple description coded stream is distributed over a network organized in a peer-to-peer fashion such as the one described in [5] where multiple multicasting trees are used. In such an environment, a scheme like the one proposed in this work would allow a simple tree management and compatibility among heterogeneous peers.

In order to maintain the compatibility with the H.264/-SVC coder, we developed a method using a pre- and post-processing scheme. In the pre-processing part, we downsample the original sequence by rows and columns generating four sub-sequences that can be independently coded as in the PSS-MD method. To maintain compatibility with the standard, we propose to predict two of them by using some of the tools that guarantee scalability in the H.264/SVC coder. We call the proposed method *Inter Layer Prediction Spatial Multiple Description Scalable Coding* (ILPS-MDSC), because it takes advantage of the inter layer prediction algorithm [6], that is used to generate efficient spatial or even coarse grain scalable streams. By using a scalable approach in order to generate each description, we have focused our work to determine the best ratio between the base layer rate and the full rate of a description, i.e. the ratio that maximize the overall performance of the algorithm in most of the possible situations, such as receiving only one or both description, or any other mix of base and enhancement layers in the two descriptions.

The proposed algorithm is presented in detail in section 3 The original H.264/SVC coder and details of the implementation are provided respectively in sections 2 and 3. Finally, simulation results are provided in section 4.

2. DESCRIPTION OF SCALABLE CODING TOOLS

In this section, we describe the scalable video coding tools used in our proposal to develop the multiple description coding scheme. Aiming at implementing a scalable structure, we selected the layered version of the H.264 coder [7] that can be classified as a layered video coder. In each layer, the basic concepts of motion-compensated prediction and intra prediction are employed as in H.264/AVC. As a first interpretation, the pictures of different layers are coded independently with layer-specific motion information. Nonetheless, the pictures of subsequent layers are often strongly correlated, so that using inter-layer prediction mechanisms may lead to quality improvements. In fact, re-using the base layer motion information in a spatial enhancement layer results in very poor coding efficiency, since motion data are optimized for a low resolution layer. On the other hand, if the motion-residual information of the enhancement layer is optimized, then the quality of the base layer substantially drops. Consequently, every spatial layer needs its own motion-residual information. However, in order to improve the coding efficiency, an inter-layer prediction mechanism that employs the base layer information for enhancement layer coding, seems to be a promising approach. In order to exploit the redundancy between several spatial or SNR layers, additional inter-layer prediction mechanism can be integrated.

We see that the SNR scalability is basically achieved by quantization of the inter-layer prediction residual, while a combination of motion-residual prediction and oversampled pyramid decomposition guarantees the spatial scalability. In particular, the following inter layer tools turned out to provide gains and were included into the scalable video coder:

- prediction of intra-macroblocks using up-sampled base layer intra blocks
- prediction of motion information using up-sampled base layer motion data
- prediction of residual information using up-sampled base layer residual blocks

The same techniques can also be applied when the base layer has the same spatial resolution as the current layer (obtaining, in fact, a coarse grain SNR scalability). In this case, the up-sampling operations are simply discarded. Further details of the scalable extension of H.264/AVC and its application can be found in [8].

3. PROPOSED SCHEME

We now introduce the MDC variant and try to delineate its main features. In order to preserve the standard coder, a pre- and post- processing scheme is implemented. In the pre-processing part, we downsample the original sequence by rows and columns thus generating four different sub-frames, similarly to what is done in PSS-MD (fig. 1). Then, descriptions are formed by coupling two different subsequences that are sent together to the same standard scalable coder. Two different schemes described in figure 2 can be applied in order to group the subsequences obtained after the polyphase spatial subsampling. In the first one, called *by rows*, we group the subsequences that form even or odd rows. In the second one, called *quincunx*, we group the subsequences so that the pixels forms a quincunx lattice. I.e., if we number the subsequences from one to four and from up to bottom and from left to right, in the *by rows* scheme, we form the first description with subsequences one and two, and the other one with subsequences three and four. Instead, in the *quincunx* scheme, we group the subsequences one and four to form the first description, and the subsequences two and three for the other one. To exploit the correlation between the subsequences, we configure the coder to generate a coarse grain scalable enhancement layer of one of the two. In practice we let the coder think that the two subsequences represent two different layers.

This is obtained using the inter-layer prediction described in 2 to explicit the redundancy between the subsequences by assigning one of them to the base layer, and the other one to the enhancement layer. By doing so, most of the correlation is eliminated by the prediction algorithms thus giving a better representation of the original subsequence. Then, the coarse grain scalable coded descriptor is transmitted. At the decoder side, we reconstruct the "enhanced subsequence" by first decoding the base layer plus enhancement layer stream, then we extract the base layer in order to decode also the other subsequence. The coder structure used in this algorithm is the one represented in figure 3. In the post-processing part, the original sequence is obtained by merging the descriptors. In case of a lost descriptor, the missing pixels are reconstructed by interpolation from the received one.

4. RESULTS

The software used in our experiments in H.264/SVC version 8.1. The different options provided by the coder have been set as follows

- 1/4 pixel accuracy for motion estimation
- a single reference frame
- GOP size 8
- I frame only at the beginning
- 16x16, 16x8, 8x16, 8x8 inter-prediction blocks with SAD metric
- CABAC
- CIF sequence with 30 fps

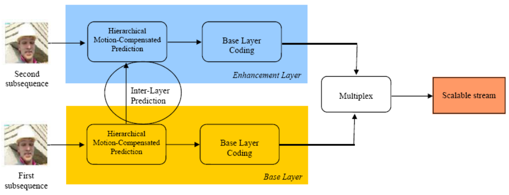

Figure 3: Coder structure needed to perform ILPS-MDSC with inter layer prediction structure highlighted

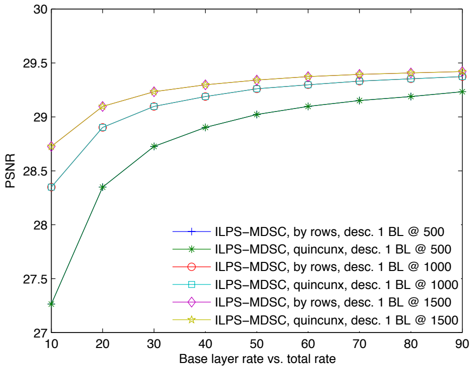

Figure 4: Foreman, performance when receiving only one subsequence

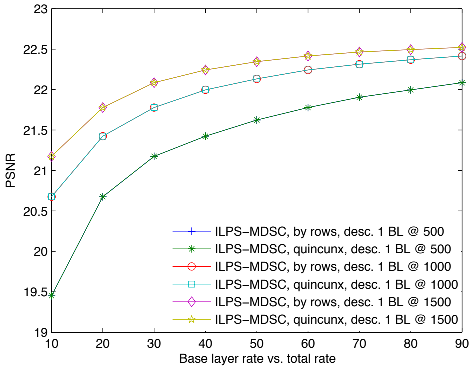

Figure 5: Mobile, performance when receiving only one subsequence

Results are reported using the sequences *foreman* (video calling environment), *football* (high motion sequence), *tempete* (medium complexity sequence with small objects) and *mobile* (high complexity sequence), in order to find the optimal ratio between the base layer rate and the full description rate when receiving only one (i.e. the base layer of one descriptor), two (one complete descriptor or base layers of both descriptors), three (a full descriptor and the base layer of the other one) and four subsequence (both full descriptors). The results are shown as the average PSNR of the sequence at different ratios, varying from 10 to 90 percent. The total bitrate of each descriptor is chosen to be respectively 500, 1000, and 1500 kbit/s in order to exploit the optimal value at different rates.

Before showing the results, we make a consideration about the interpolation schemes. We use different interpolation methods accordingly to the different number and type of subsequences received. If we receive only one subsequence, then we recover the missing information by the mean of the nearest pixels. If we receive two subsequence, recovery can be performed in two different ways depending on the exact combination of received substreams (either base+enhancement of the same description, or the base layers of the two descriptions): if the received pixels are in a *row* fashion, then we recover the missing information considering the mean of

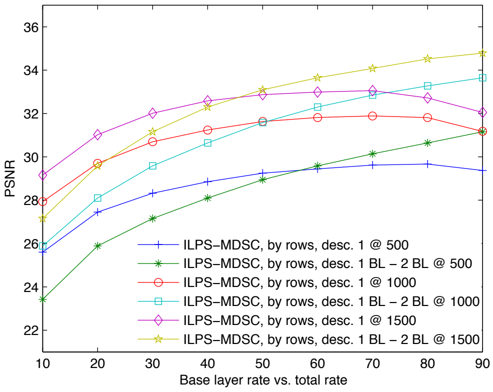

Figure 6: Football, performance when receiving two subsequences, method *by rows*

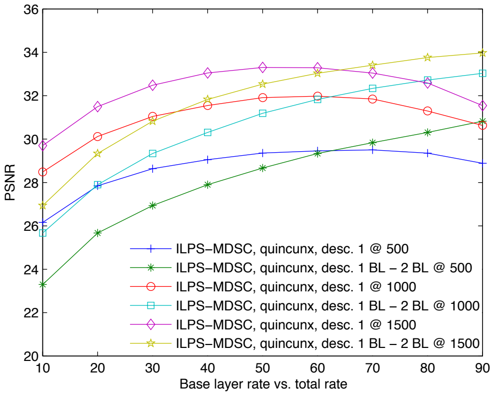

Figure 7: Football, performance when receiving two subsequences, method *quincunx*

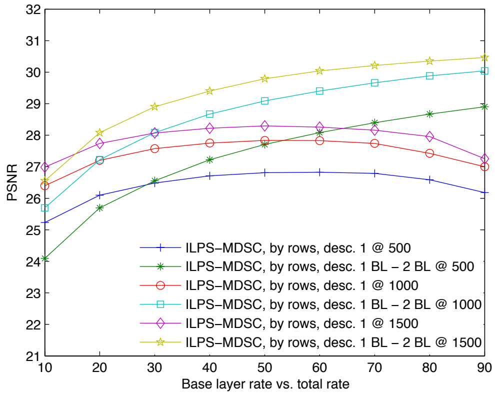

Figure 8: Tempete, performance when receiving two subsequences, method *by rows*

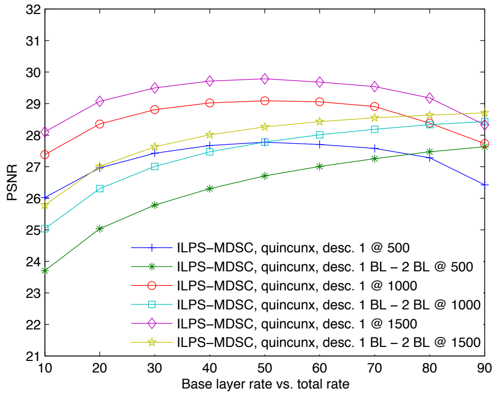

Figure 9: Tempete, performance when receiving two subsequences, method *quincunx*

the two nearest pixels, otherwise, if the received pixels are in *quincunx* fashion, then we recover the missed information as the mean of the four nearest pixels. Finally, when we received three of the four subsequence, the missing information is obtained by interpolation of the eight nearest pixels. From now on, we show some results of the simulation made: although they represent a subset of the possible ones, the missing results are not reported because they lead to similar considerations.

Figures 4 and 5 show the performance when only one subsequence is received. Obviously, as we decode only the base layer of one description, the quality of the reconstructed sequence improves as far as we increase the ratio. However, when we assign to the base layer a more than 60% of the total rate, the improvement lowers as we reach the asymptotic value of the reconstruction algorithm.

Figures 6, 7, 8 and 9 show the performance when receiving two subsequence, either representing a complete description or the base layers of both descriptions. In this figures, we have two different behaviors: the full descriptions seem to have the maximum value with a ratio between 50 and 60%. This indicates that it's better to generate balanced descriptors assigning the overall rate to the base and the enhancement layer in an almost ”fair” way. Plus, we can say that, due to the exploited redundancy of our scheme, the ratio of 60% seems to give the better performance in most cases. On the other hand, the reconstructed sequence obtained only by the base layers of both descriptions gives the same trends of the ”one subsequence” figures. However, above the 70% the gain in very low compared with the loss in case of receiving only one full descriptor. Figures 10, 11, 12 and 13 show the performance for three received subsequences. When this happens, we always have one full descriptor and both base layers. The results seem to be influenced to a greater extent by the rate of the enhancement layer rather than by the gain of having another base layer, so the maximum performance seem to be achieved with a ratio of about 60-70%, nearer to the previous case of sequence reconstructed from a full descriptor than to the case when only both base layers are received. Finally, figures 14 and 15 show the performance when all the subsequences are received. As we can see, the maximum performance is reached with ratio of 60%, thus confirming all the given above consideration. An interesting result that can be viewed from our simulations is that at lower bitrates the algorithm seems to achieve better performance if the ratio is increased of about 10%. This seems to

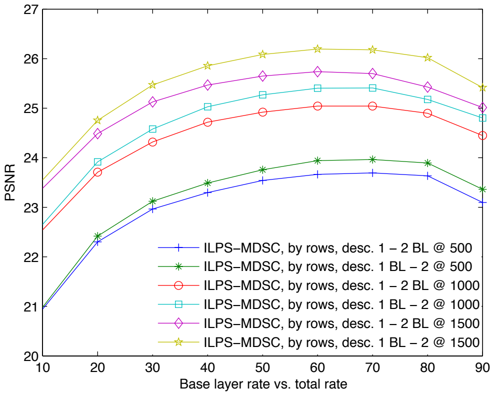

Figure 10: Mobile, performance when receiving three subsequences, method *by rows*

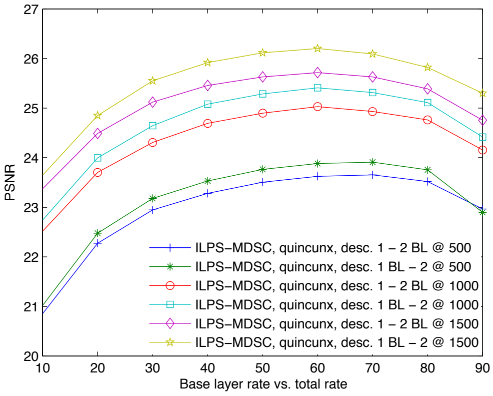

Figure 11: Mobile, performance when receiving three subsequences, method *quincunx*

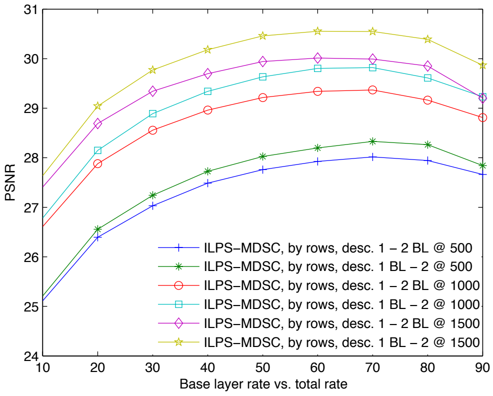

Figure 12: Tempete, performance when receiving three subsequences, method *by rows*

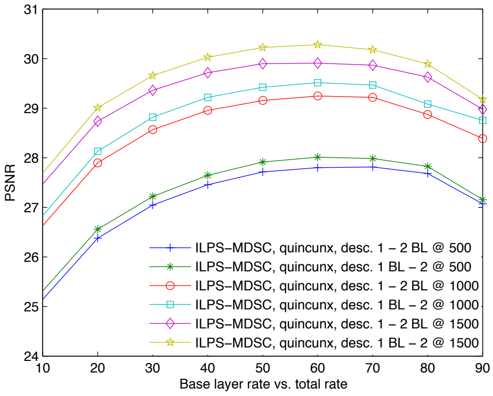

Figure 13: Tempete, performance when receiving three subsequences, method *quincunx*

be caused by the greater efficiency of the inter-layer prediction scheme at lower bitrate than at higher bitrate.

Two considerations may be made comparing the performance of the *by rows* and *quincunx* schemes. In the *by rows* scheme, the pixels that form each description are closer, and this results in a higher correlation among the subsequences that will form the descriptor. In this case, the ILPS-MDSC scheme can code the enhancement layer more efficiently, i.e with a lower bitrate. This can be seen in our experiments with a better overall performance obtained by this scheme when the ratio is about 10% larger than the same simulation with a *quincunx* scheme. The other consideration is about the overall performance of both scheme when receiving different subsequences. As we can see from the figures, due to the the greater or smaller correlation between the pixels of the subsequences that form the descriptor, the *by rows* and *quincunx* scheme gives different performances when receiving an even number of subsequences. In case of receiving only two subsequences, when they are in a *by row* fashion, the algorithm always gives worse performance compared to the *quincunx* scheme, this is due to the reduced information (the pixels are more correlated) that the reconstruction algorithm can use to recover the original sequence. On the contrary, in case all the subsequences are received, due to the higher efficiency of the ILPS-MDSC scheme, the descriptors coded in a *by rows* scheme give better performance than the same *quincunx* scheme. Then, when receiving an odd number of subseqeunces, the performances are almost the same. Finally, the performances of different spatial multiple description schemes (PSS, the spatial scheme proposed in [4], called DPSS, and the ILPS) are shown in tables 1 and 2 for the *foreman* sequence, and in tables 3 and 4 for the *football* sequence. For the ILPS scheme, we have chosen as ratio between the base layer and the whole stream the one that gives the better results (60-70%). We can see that the proposed method gives better performance with respect to the considered spatial multiple description schemes, in particular when the number of description received is low (one or two descriptions). This is due to the better bitrate allocation of the proposed method in comparison with the other methods in which the bitrate is simply equally divided between the subsequences of the descriptors. When the number of description is high (three or four description), the ILPS algorithm performs better than the PSS and gives almost the same performance ad the DPSS.

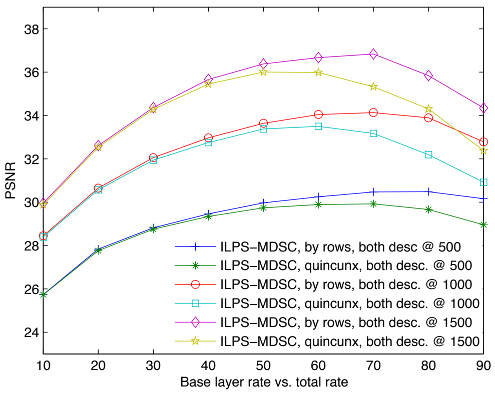

Figure 14: Football, performance when receiving all subsequences

5. CONCLUSIONS

In this paper we have introduced a novel algorithm to generate multiple descriptions in a H.264/SVC coder and shown its performance by varying the ratio between the base layer rate and the full description rate. The experiments show that the better overall performance can be obtained by using a ratio of about 60-70% in almost every considered sequences. Work is in progress to introduce it in ”real” network scenarios to exploit the optimal value of the ratio in case of missing packets and to evaluate the algorithms in a flexible content-distribution framework considering either peer to peer and wireless environments. Also, it is under way the introduction of Fine Granular Scalability in each layer in order to find some rate distortion analytic functions that fits well the rate distortion curve obtained.

	One Desc.,	Two Desc.,	Three Desc.,	Four Desc.
PSS	29	29.6	33.4	35.6
DPSS	29	29.6	33.6	36.3
ILPS	29.1	29.6	33.7	36.3

Table 1: Performance of different spatial MD methods for the *Foreman* sequence, *by rows* @ 500 kbit/s

	One Desc.,	Two Desc.,	Three Desc.,	Four Desc.
PSS	29	33.1	33.5	35.6
DPSS	29	29.6	33.6	36.3
ILPS	29.1	33.1	33.7	36

Table 2: Performance of different spatial MD methods for the *Foreman* sequence, *quincunx* @ 500 kbit/s

	One Desc.,	Two Desc.,	Three Desc.,	Four Desc.
PSS	28	28.5	28.6	29
DPSS	28.1	29.2	29.2	30
ILPS	28.7	29.6	29.7	30.25

Table 3: Performance of different spatial MD methods for the *Football* sequence, *by rows* @ 500 kbit/s

	One Desc.,	Two Desc.,	Three Desc.,	Four Desc.
PSS	28	28.7	28.7	29
DPSS	28.1	29.1	29.1	30.1
ILPS	28.7	29.5	29.5	29.9

Table 4: Performance of different spatial MD methods for the *Football* sequence, *quincunx* @ 500 kbit/s

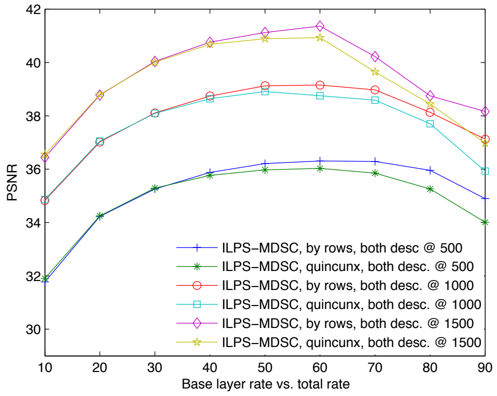

Figure 15: Foreman, performance when receiving all subsequences

6. REFERENCES

[1] V.K. Goyal, “Multiple Description Coding: Compression meets the network,” *Int’l Conf. on Image Processing (ICIP)*, vol. 18, no. 5, pp. 74 - 93, Sept. 2001.

[2] Y. Wang and S. Lin, “Error resilient video coding using multiple description motion compensation,” *IEEE Trans. CSVT*, vol. 12, no. 6, pp. 438 - 452, June 2002.

[3] A. Vitali, F. Rovati, R. Rinaldo, R. Bernardini and M. Durigon, “Low-Complexity Standard-Compatible Robust and Scalable Video Streaming over Lossy/Variable Bandiwith Networks,” Proc. of *IEEE Int’l Conf. on Cons. Elec.*, Jan. 2005, Las Vegas, USA, pp. 1022-1025.

[4] M. Folli, L. Favalli, “Multiple Description Coding algorithms for H.264 coder,” *Mobimedia ’07*, Nafpaktos, Greece, Aug. 2007.

[5] J. D. Mol, D. H. Epema, H. J. Sips, “The Orchard algorithm: building multicast trees for P2P video multicasting without free riding,”, *IEEE Trans. on Multimedia*, vol. 9, no. 8, Dec. 2007.

[6] H. Schwarz, D. Marpe, and T. Wiegand, “Basic concepts for supporting spatial and SNR scalability in the scalable H.264/MPEG4-AVC extension,” *Proc. of IWSSIP 2005*, Chalkida, Greece, Sept. 2005.

[7] H. Schwarz, T. Hinz, H. Kirchhoffer, D. Marpe, and T. Wiegand, “Technical description of the HHI proposal for SVC CE1,” *ISO(IEC JTC1/SC29/WG11*, Doc. m11244, Palma de Mallorca, Spain , Oct. 2004.

[8] R. Schafer, H. Schwarz, D. Marpe, and T. Wiegand, “MCTF and Scalability Extension of H.264/AVC and its applications to video transmission, storage and surveillance,” *Visual Comm. and Image Proc.*, July 2005.