Multiple Description Coding algorithms for H.264 coder

Marco Folli
Electronics Department, Università di Pavia
via Ferrata 1
Pavia, Italy
+39 0382 985923
angelo.marco@ngi.it

Lorenzo Favalli
Electronics Department, Università di Pavia
via Ferrata 1
Pavia, Italy
+39 0382 985923
lorenzo.favalli@unipv.it

ABSTRACT
In this paper we address the problem of video transmission over unreliable networks. The proposed approach moves from the concept of multiple description coding (MDC) introducing some simple algorithms to obtain sub-streams introducing the concepts of either spatial, temporal, SNR adaptation The proposed algorithms have been implemented as pre- post- processing of the basic H.264/SVC codec. Although in principle as many descriptions as desired can be generated, simulation results using only two description show good results both when all descriptions are received and when only one of the two is received. Analysis performed in case of either random or packet losses show the robustness of the algorithms.

Keywords
H.264, multiple description coding, scalability.

1. INTRODUCTION
Transmission of video sequences over networks which don't provide any quality of service guarantee or where transmission errors may occur at any time is a challenging task and has received much attention over the years and in particular in the last years due to the ever increasing availability of efficient video coders and increasing bandwidth availability. Despite the many advances, still transmission of video encounters great difficulties as newer transmission environments enter the scene.

So for example while the concepts of scalable video and multiple description coding have been the subject of a long list of papers all showing the advantages of these approaches in terms of both adaptation capability to different user requirements and network constraints, the emergence of peer-to-peer systems used as efficient (and legal) content distribution networks or mesh/ad-hoc/sensor networks with strong resource and power constraints, pushes once more the research of efficient and reliable video coding algorithms.

Work partially supported by Italian Ministry of University under grant RBIN043TKY "Software and Communication Platforms for High Performance Collaborative Grid"

MOBIMEDIA 2007, August 27-29, Nafpaktos, Greece

DOI 10.4108/ICST.MOBIMEDIA2007.1729

Multiple description coding (MDC) [1], in particular, is the process by which a coder outputs several substreams, all individually decodable at a lower quality than the original. Receiving all the descriptions, on the contrary, ideally allows the full recovery of the single-stream coded video.

The most recent H.264 video coding standard provides a set of coding techniques that efficiently compress the video source signal and allow excellent performances although at the cost of an increase of the computational complexity. Its scalable extension H.264/SVC [2], allows very good performances in terms of flexibility in case of adaptive source/channel coding or variable available bandwidth but is not as robust to packet losses such as those that may be experienced by the transmission over IP based networks.

The starting point of the work presented in this paper is then the joint exploitation of different mixes of scalability and multiple description to take advantage of both schemes. The first method stems from a spatial split among the descriptions described as *polyphase spatial subsampling* [3] (PSS-MDC). In this approach the video sequence to be transmitted is subsampled by row and column thus giving rise to four sub-sequences, which are then independently coded and transmitted. Since each can be used to generate the original sequence by oversampling and interpolation, this technique has good resilience to losses at the expenses of coding efficiency. As a compromise, we propose to predict two of them and then code and transmit only the residual. Another simple way to produce substreams is by temporal subsampling prior then coding thus generating half- (or lower) frame rate streams. Finally an interesting approach has been recently proposed and applied to JPEG2000 [5][6] that mixes MDC and unequal quality coding of different parts of the image. Following the philosophy of these papers, an extension to the H.264 coder is introduced where each description carries the base layer plus a different portion of the enhancement layer.

The proposed algorithms are presented in detail in section 2, while description of their implementation on top of the H.264/SVC coder and simulation results are provided in section 3.

2. PROPOSED SCHEMES
We now introduce the MDC variants and try to delineate their main features. For the sake of comparison, a rough MDC scheme based on temporal subsampling is also considered. In this approach, the video sequence is first splitted in two subsequences by alternatively picking even and odd frames. The two streams are independently coded and no side information or motion vector

processing is performed so that this scheme is conceptually very simple

2.1 Spatial MDC scheme

As first step, each frame is subsampled by a factor of four simply picking alternatively pixels from each row and column. We thus obtain four images I_k (k=1,2,3,4) named in simple sequential manner from the top left. Due to their high correlation and having as a target to transmit only two descriptions, we consider only two subframes as independent while the others are predicted: coded and predicted are then considered as a single descriptions. We now need to select which of the frames to code and which to predict. After some experiments, the best results have been on average obtained considering as independent the top left and low right positions, i.e. those numbered 1 and 4 while those numbered 2 and 3 are then predicted. We will then have

$$B_2(i,j) = \frac{1}{4}\left[I_1(i,j) + I_1(i+1,j) + I_1(i,j-1) + I_1(i+1,j-1)\right]$$

$$B_3(i,j) = \frac{1}{4}\left[I_4(i,j) + I_4(i-1,j) + I_4(i,j-1) + I_4(i-1,j-1)\right]$$

where $I_k(i,j)$ represents the value of pixel (i,j) in sub-frame k. The residuals are then given by

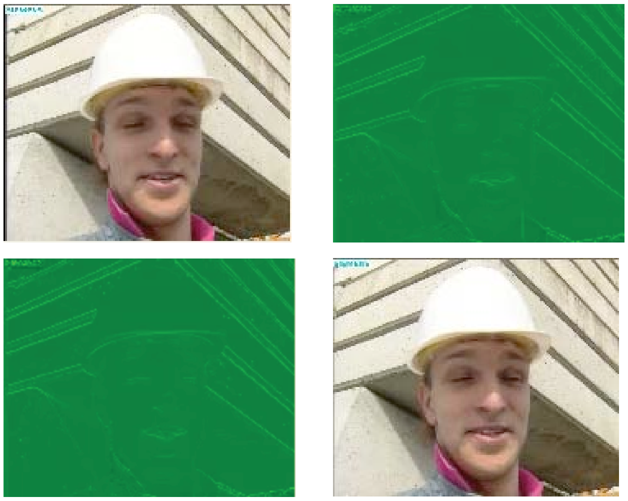
$$D_2 = |I_2 - B_2|$$

$$D_3 = |I_3 - B_3|$$

a sign bit is added for every pixel, in order to reconstruct correctly the original frame.

Figure 1 shows the four subframe before coding. Each description is then formed by a coded frame and the correspondent predicted frame (e.g 1 and 2 forms the first description, 3 and 4 forms the second one).

At the receiver, for each description, we decode both subimages and, after that, we reconstruct the predicted one from the other. If only one description is received, a linear interpolation by rows is applied for the missing pixels.



Figur1 1. Subsequences predicted and coded in the spatial multiple description algorithm

2.2 SNR MDC scheme

Differently from the previous algorithms, this scheme doesn't perform any subsampling of the sequence but cannot be performed "on the fly" as it requires observation of the sequence over a given, although arbitrary, time span to equalize the actual bit rate between the two descriptions.

Let's now consider a generic scalable approach and define R_1 the rate of the complete sequence and R_2 the rate of the base layer. Then each description carries the base layer while the enhancement layer is assigned evenly to the different descriptions according to different possible schemes. The easiest approach, which is presented in this paper, is to use temporal subsampling of the enhancement layer only which is assigned in turn to each description. Other schemes may employ the optimization of the quantization in the enhancement layer(s). For our purpose now, given an interval M, the average bit rate for each of the N possible descriptions is given by

$$\frac{R1 + (N-1)R2}{N}$$

For two descritions this naturally reduces to the average. The enhancement layer may be generated in a way similar to that of SNR scalable coders: in the present work we have used the fine grai salable (FGS coder) as it is very versatile. At the receiver, if both descriptions are received, we select the description that carries the enhancement layer for every time interval and then we send it to the decoder in order to obtain the reconstructed sequece with the best quality as possible. The redundancy introduced by this scheme is very high as the base layer is transmitted "as is" in all descriptions. Indeed, this scheme may be exploited in highly unreliable channels to reduce the channel coding overhead exploiting diversity itself as a mean to recover the original sequence or at least a good part of it also exploiting the inherent robustness of the FGS coding. It may also be used jointly with other MDC schemes subsampling the base layer and dividing it in the different substreams.

3. RESULTS

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

- ¼ pixel accuracy for motion estimation
- a single reference frame
- GOP of size 8
- I frame only at the beginning
- 16x16, 16x8,8x16,8x8 inter-prediction blocks with SAD metric
- CABAC
- CIF sequences with 30 fps

Results are reported using the sequences *foreman*, *mobile* and *tempete* considering the *no-losses* case and a case with a 10% random packet loss.

In Figure 2, the spatial-multiple description approach of section 2.1 is compared with a PSS-MDC and with a single description

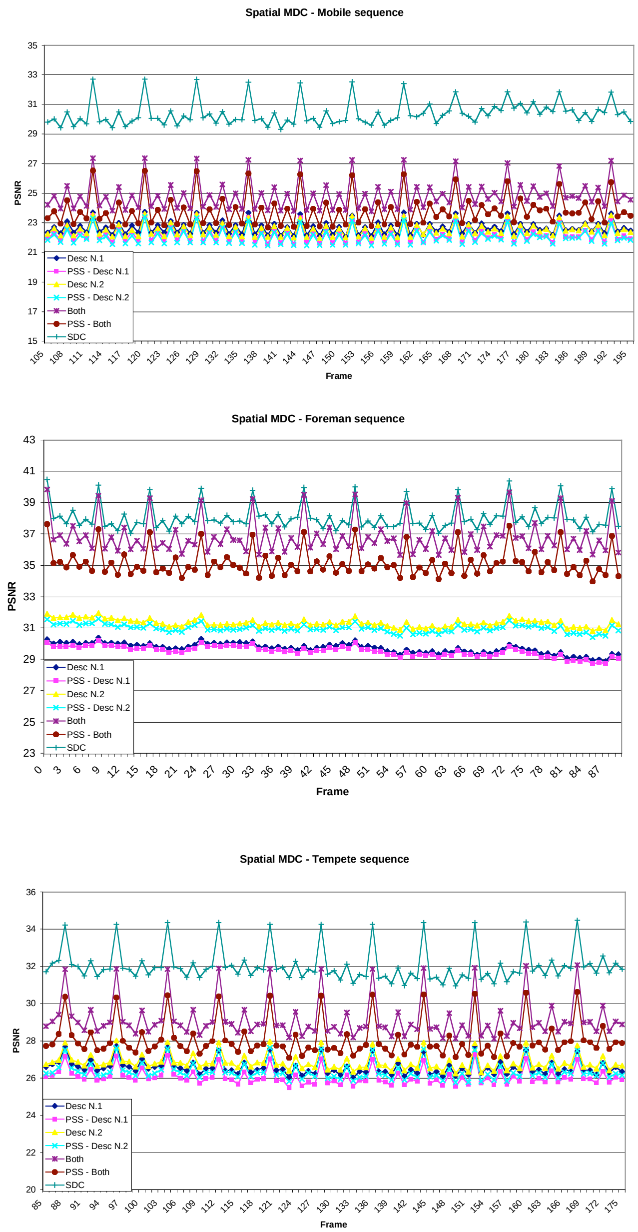


Figure 2. Performance of the spatial MDC algorithm no-losses.

coded (SDC) sequence. The target bit rate for any description is 384 kb/s, obtained by assigning at the subframe coded the rate of 240 kbit/s and at the subframe predicted the rate of 144 kbit/s. For the PSS-MDC, we assign the same rates as the spatial multiple description approach, only without prediction. It is possible to note that the proposed scheme not only performs better than the "base" PSS-MDC but almost performs as good as SDC. If we introduce losses of up to 10% of transmitted packets (Figure 3), still our algorithm perfoms better that PSS-MDC while SDC rapidly degrades. Table 1 provides the average PSNR for the three MD approaches in the case without losses.

Figures 4 and 5 depict the same scenarios for the temporal MDC. We note that when receiving both descriptions, reconstruction is

Figure 3. Performance of the spatial MDC algorithm with 10% losses

"ideal" while losses degrade the quality of the stream due the reduced frame rate.

Table 1 Average PSNR for spatial MD without losses

Sequence	Desc N.1	Desc N.2	Both	SDC
Foreman	29,7	31,3	36,8	38
Mobile	22,8	22,7	25	30,5
Tempete	26,59	26,82	29,25	32,05

Finally, figures 6 and 7 allow evaluating the FGS-based MDC of section 2.2. In this case, the base layer has rate 144kbit/s while the enhancement has a rate of 624 Kbit/s thus giving an overall bit rate of 624 kbit/sec that may be divided into 2x384 kbit/sec substreams, comparable with the previously defined sequence. We may notice that the performance in case of reception of both substreams is again comparable with a single description coder. Note also the switch between the two substreams given the simple scheme introduced in this work.

Table 2 Average PSNR for temporal MD without losses

Sequence	Desc N.1	Desc N.2	Both	SDC
Foreman	33,2	33,2	39,5	38
Mobile	26,4	26,4	32,1	30,5
Tempete	29,4	29,4	33,5	32,05

Figure 4. Performance of the temporal MDC algorithm without losses

. Table 3 Average PSNR for SNR MD without losses

Sequence	Desc N.1	Desc N.2	Both	SDC
Foreman	35,8	35,4	37,3	38
Mobile	28,8	29,2	30,7	30,5
Tempete	31,33	31,16	32,65	32,05

Finally, figure 8 gives a sample of reconstructed images for the two proposed approaches, a temporal MDC and a single description coder. It can be seen that the visual quality is very similar among all different approaches. Figure 9 gives on the contrary the visual results when only one description is received.

Figure 5. Performance of the temporal MDC algorithm with 10%losses

4. CONCLUSIONS

In this paper we have introduced some possible novel algorithms to generate multiple descriptions in an H.264 coder and shown their performance in comparison with other approaches and with single derscription coding. Work is in progress to improve these algorithms and to introduce them in “real” network scenarios to exploit their adaptability and robustness features. Numerical results in Tables 1-3 show the average PSNR value for the proposed approaches, a temporal MDC and a single description coder. We can note that SNR and temporal multiple description scheme give similar performance as the SDC, while the spatial scheme gives a slightly lower performance

5. REFERENCES

[1] V.K. Goyal, “Multiple Description Coding: Compression meets the network,” *IEEE Signal Proc. Mag., pp. 74-93,* Sept. 2001

[2] 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*.

[3] R. Bernardini, M. Durigon, R. Rinaldo, L. Celetto, and A. Vitali, "Polyphase spatial subsampling multiple description coding of video streams with H.264," *Image Processing, 2004. ICIP '04, pp. 3213-3216 Vol. 5, Oct. 2004*.

Figure 6 Performance of the SNR MDC algorithm without losses.

Figure 7 Performance of the SNR MDC algorithm with 10% losses.

[4] Zhe Wei, Canhui Cai, and Kai-Kuang Ma, “H.264-based Multiple Description Coding and Its DSP Implementation,” *IEEE International Conference on Image Processing (ICIP), Atlanta, GA, USA, pp. 3253-3256, Oct. 2006*.

[5] T. Tillo, and G. Olmo, “A novel multiple description coding scheme compatible with the JPEG 2000 decoder,” *IEEE Signal Processing Letter, vol.11, no.11, pp. 908-911, Nov 2004*.

[6] E. Baccaglini, T. Tillo, and G. Olmo, “Network Adaptive Multiple Description Coding For JPEG2000,” *IEEE International Conference on Image Processing, Genova, pp. 936-939, September 2005*

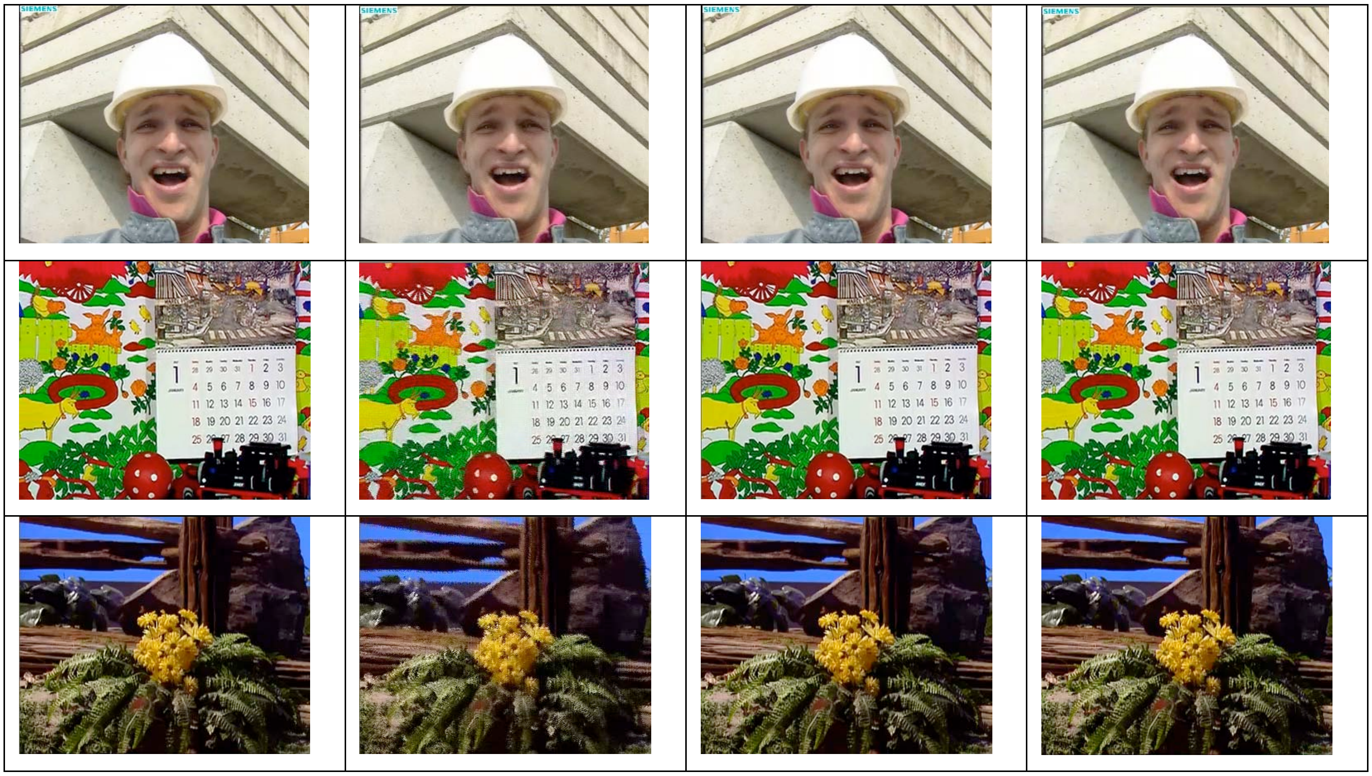


Figure 8. comparison of visual results. Left to right: SDC, Spatial MD, temporal MD, SNR MD

Figure 9. Single description received: top row spatial MD, lower row SNR MD.