A COMPARISON OF FOUR VIDEO MULTIPLE DESCRIPTION CODING SCHEMES

Christophe Tillier[1], *Olivier Crave*[1,2], *Béatrice Pesquet-Popescu*[1] *and Christine Guillemot*[2]

[1]GET-Télécom Paris, Signal and Image Processing Dept.
46 rue Barrault 75634 Paris Cedex 13, France
phone: + 33145817192, email: {tillier,crave,pesquet}@tsi.enst.fr
[2]INRIA - IRISA ,TEMICS project
Campus du Beaulieu, Avenue du Général Leclerc, 35042 Rennes, France
email: christine.guillemot@irisa.fr

ABSTRACT

Multimedia communications over best-effort networks often involve situations in which immediate data retransmission is either impossible (e.g., network congestion or broadcast applications) or undesirable (e.g., conversational applications with very low delay requirements). A specific solution developed for increasing the transmission reliability without increasing the delay is known as multiple description coding. *In this paper, we compare four video MDC schemes based on different time splitting patterns and temporal two- or three-band MCTF. Taking into account the temporal distance between frames, we show the respective advantages and drawbacks of these schemes, for central and side decoders.*

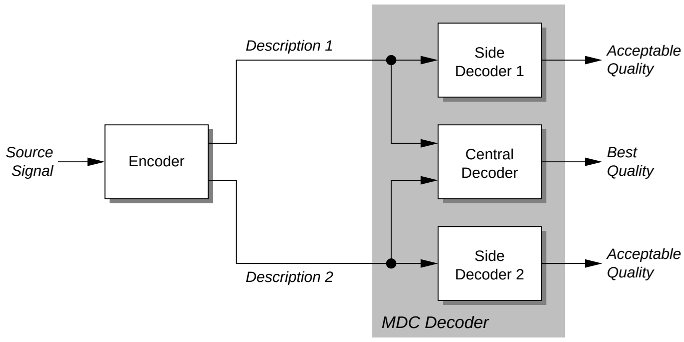

Figure 1: Generic MDC scheme with two descriptions.

1. INTRODUCTION

With increasing use of the Internet and other best-effort networks for multimedia communication, there is a growing need for reliable transmission. Traditional research efforts have concentrated on enhancing existing error-correction techniques; however, recent years have seen an alternative solution emerge and gain increasing attention. This latter solution focuses mainly on the situation in which immediate data retransmission is either impossible (e.g., network congestion or broadcast applications) or undesirable (e.g., conversational applications with very low delay requirements). We are referring to the specific technique known as *multiple description coding* (MDC). The reader is referred to [1] for a comprehensive general review of MDC.

In essence, the MDC technique operates as illustrated in Fig. 1. The MDC encoder produces several correlated—but independently decodable—bitstreams called *descriptions*. The multiple descriptions, each of which preferably has equivalent quality, are sent over as many independent channels to an MDC decoder consisting of a *central decoder* as well as multiple *side decoders*. Each of the side decoders is capable of decoding its corresponding description independently of the other descriptions, producing a representation of the source with some level of minimally acceptable quality. On the other hand, the central decoder can jointly decode multiple descriptions to produce the best-quality reconstruction of the source. In the simplest scenario, the transmission channels are assumed to operate in a binary fashion; that is, if an error occurs in a given channel, that channel is considered damaged, and the entirety of the corresponding bitstream is considered unusable at the receiving end.

The success of an MDC technique hinges on path diversity, which balances network load and reduces the probability of congestion. Typically, some amount of redundancy must be introduced at the source level in order that an acceptable reconstruction can be achieved from any of the descriptions, and such that reconstruction quality is enhanced with every description received. An issue of concern is the amount of redundancy introduced by the MDC representation with respect to a single-description coding, since there exists a trade-off between this redundancy and the resulting distortion. Therefore, a great deal of effort has been spent on analyzing the performance achievable with MDC ever since its beginnings [2, 3] up until recently, e.g., [4].

As an example of MDC, consider a wireless network in which a mobile receiver can benefit from multiple descriptions if they arrive independently, for example, on two neighboring access points. In this case, when moving between these two access points, the receiver might capture one or the other access point, and, in some cases, both. Another way to take advantage of MDC in a wireless environment is by using two frequency bands for transmitting the two descriptions. For example, a laptop may be equipped with two wireless cards (e.g., 802.11a and g) with each wireless card receiving a different description. Depending on the dynamic changes in the number of clients in each network, one wireless card may become overloaded, and the corresponding description may not be transmitted. In wired networks, different descriptions can be routed to a receiver through different paths by incorporating this information into the packet header [5]. In this situation, the initial scenario of binary “on/off” channels might no longer be of interest. For example, in a typical CIF-format video sequence, one frame might be encoded into several packets. In such cases, the system should be designed to take into consideration individual or bursty packet losses rather than a whole description. Several directions have been investigated for video using MDC. In [6,7,8,9], the proposed schemes are largely deployed in the spatial domain within hybrid video coders such as MPEG and H.264/AVC; a thorough survey on MDC for such hybrid coders can be found in [10].

On the other hand, only a few works investigate MDC schemes that introduce source redundancy in the temporal domain, although this approach has shown some promise. In [11], a balanced interframe MDC was proposed starting from the popular DPCM technique. In [12], the reported MDC scheme consists of temporal subsampling of the coded error samples by a factor of 2 so as to obtain two threads at the encoder which are further independently encoded using prediction loops that mimic the decoders (i.e., two side prediction loops and a central prediction loop). MDC has also been applied to MCTF-based video coding: existing work for $t+2D$ video codecs with temporal redundancy addresses 3-band filter banks [13, 14]. Another direction for wavelet-based MDC video uses the polyphase approach in the temporal or spatiotemporal domain of coefficients [15, 16, 17].

In this paper, we compare four video MDC schemes based on different time splitting patterns and temporal two- or three-band MCTF. Taking into account the temporal distance between frames, we show the respective advantages and drawbacks of these schemes, for central and side decoders.

The remaining of the paper is organized as follows: in Section 2 we describe the four schemes which will be compared. In Section 3 are shown the experimental results and we conclude in Section 4, where perspectives for future work are also proposed.

2. TEMPORAL MULTIPLE DESCRIPTION CODING SCHEMES

In the first scheme, illustrated in Fig. 2, odd and even frames are splitted between the two descriptions. One level of a motion-compensated Haar decomposition is then applied on the frames of each description. The temporal detail frames are encoded, while the passage from one level to the next one is done by interleaving the approximation frames from both descriptions. This new sequence will be subsequently distributed again among the two descriptions.

The second scheme (see Fig. 3) starts by splitting groups of two consecutive frames between the descriptions. Again, one level of a Haar MCTF is applied to these couples of frames, and the details are encoded in their respective descriptions. As before, the passage from the first level to the next one is done by interleaving the approximation frames from the two descriptions. Next, the scheme continues as the first one, by encoding with Haar MCTF odd and even frames in different descriptions. One can remark it is not possible to have the same gathering as at the first level in groups of two frames, since we would perform temporal filtering on approximation frames coming from different descriptions, so in case one of them is lost, it will not be possible to reconstruct any of them. Another remark is that longer temporal filters would also be difficult to use in this framework, since for all the MDC schemes presented here the temporal distance between frames in the same description is higher than one, and the longer the filter, the smaller the correlation between the frames. Therefore, we restrict ourselves to Haar MCTF, even though the coding performance of 5/3 MCTF is known to be better in absence of losses.

In this second scheme, since the encoding is performed on couples of successive frames, one can already expect a better performance of the central decoder of this scheme compared with the first one, where one over two frames are considered in each description. However, when only one description is received, in the first scheme the side decoder will have to reconstruct one over two frames. The temporal distance between missing frames being only one, this task is not very difficult, and visual and objective performance may be expected to be good. On the other hand side, for the second scheme the temporal distance between missing frames from the lost description is of two, so their interpolation could be more tricky.

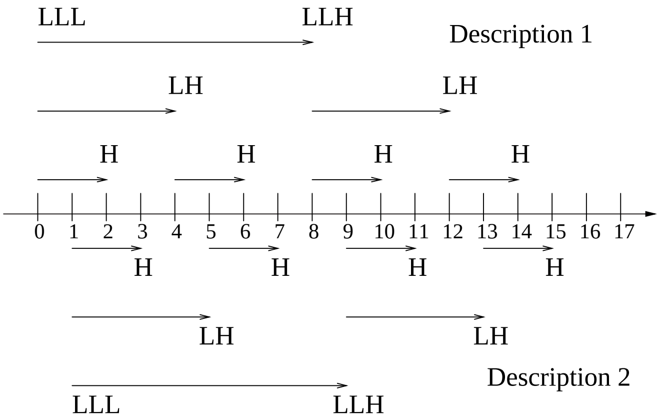

Figure 2: *First MDC scheme: odd/even temporal splitting and two-band Haar MCTF.*

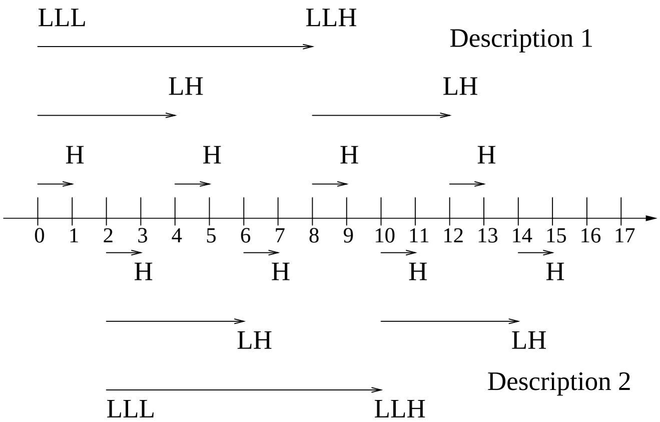

Figure 3: *Second MDC scheme: frames go two by two to descriptions and then a two-band Haar MCTF is applied in each one.*

The third scheme is similar to the second one, except that groups of three consecutive frames are separated in each description (see Fig. 4). A Haar three-band MCTF [18] is this time applied on triplets and approximation frames are interleaved to form the new sequence at the second decomposition level.

The last MDC scheme, illustrated in Fig. 5 involves a temporal splitting of the input frames in odd and even ones,

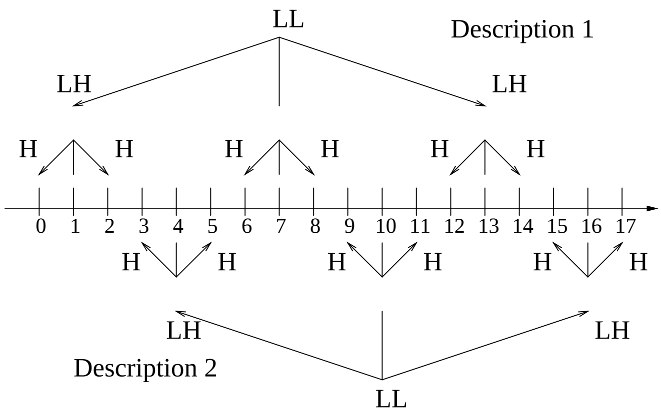

Figure 4: *Third MDC scheme: a three-band MCTF is applied to groups of three frames of each description.*

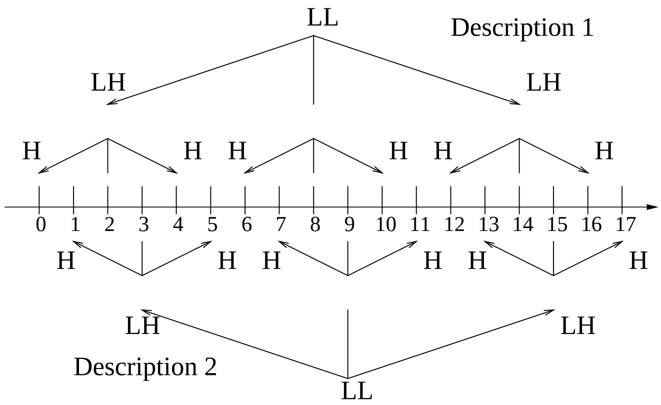

Figure 5: *Fourth MDC scheme: odd and even frames are separated and a three-band MCTF is then applied in each description.*

for the two descriptions, followed by a Haar three-band MCTF on each flow. As in the case of two-band schemes, for this decomposition, compared with the previous one, one can expect lower performance for the central decoder. At the side decoders, due to the smaller temporal distance between frames used for interpolating missing ones, one may expect an improvement compared to the third scheme. Indeed, for the last proposed scheme the temporal distance between missing frames is only one, while for the third scheme the side decoders will have to interpolate from frames being spaced of three frames to fill in gaps resulting from the loss of one description. On the other hand, there is a loss in performance related to the fact that the original encoding is done on frames spaced by one, instead of groups of consecutive frames. These two antagonist trends will be studied in the next section in an experimental framework.

3. EXPERIMENTAL RESULTS

We have implemented the four proposed MDC video coding schemes using the MC-EZBC software [19]. Three temporal levels of decomposition are performed for the two-band MCTF schemes (i.e., the first and the second one) and two levels for the three-band MCTF schemes (i.e., the third and the fourth one). The temporal interpolation and error concealment are achieved here in the simplest way, by making a copy of the closest decoded frame (which may be a past or a future one), if this one is at a temporal distance of one from the missing frame. When this distance is more than one, as is the case for the side decoders of the third scheme (see Fig. 4), the two missing frames adjacent to existing frames are reconstructed by a simple copy, and then averaged to give the reconstruction of the frame in the middle of the lost group of three. Simulations have been conducted on several test sequences, and results are presented for Mobile and Foreman, in QCIF format at 30 fps.

The first two schemes, involving two-band MCTF, are compared in Figs. 6 and 8. As expected, the central decoder of the second scheme performs better than that of the first one. However, one can remark even the side decoder of this second scheme slightly overperforms the one of the first scheme. This can be explained by the very simple interpolator used in our simulations. Indeed, when one receives only the first description, one can see from Fig. 3 that the frame 2 is reconstructed by copying the frame 1 and frame 3 by copying frame 4. This is not worse than copying frame 1 from 0 and frame 3 from 2 in Fig. 2. The reconstruction of missing frames is therefore equivalent in the two schemes and we find, as for the central decoders, that the coding efficiency of the second scheme, where frames are encoded by couples of consecutive frames, is better than for the first one.

One can also remark that even though the two schemes only differ at the first temporal level of decomposition, the difference in coding performance is quite important, of almost 1 dB for the central decoders and almost 0.5 dB between the side decoders.

The performance comparison of the third and fourth schemes, based on three-band MCTF, is illustrated in Fig. 7 and 9. As in the case of two-band MCTF schemes, grouping consecutive frames before filtering and encoding them in different descriptions leads as expected to better results for the central decoder of the third scheme (almost 2 dB). The side decoder also takes advantage of this good performance at the encoding stage, and the fact that we need to interpolate three missing frames when only one description is received does not seem to balance the loss in coding performance of the fourth scheme (almost 1dB), due to the frame separation at the beginning.

4. CONCLUSION AND FUTURE WORK

In this paper, we have considered a video MDC scheme based on temporal splitting of the frames in a sequence, followed by a MCTF. We generalized it to splitting groups of frames and to three-band MCTF. Experimental results have shown that grouping consecutive frames before filtering and encoding them in different descriptions provides better results than directly separating the frames, and this for the central decoders as well as for the side decoders. This effect seems to be even more important for groups of three frames and subsequent three-band MCTF than for groups of two.

Moreover, the comparison of the proposed schemes leads to very similar results on sequences with very different content, both from the motion and texture point of view, and for a large range of bitrates. This gives a strong indication in

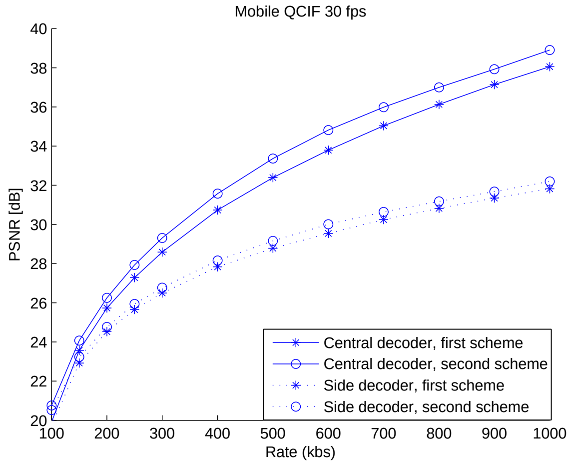

Figure 8: Performance comparison of the first and second MDC schemes (Mobile, QCIF 30 fps).

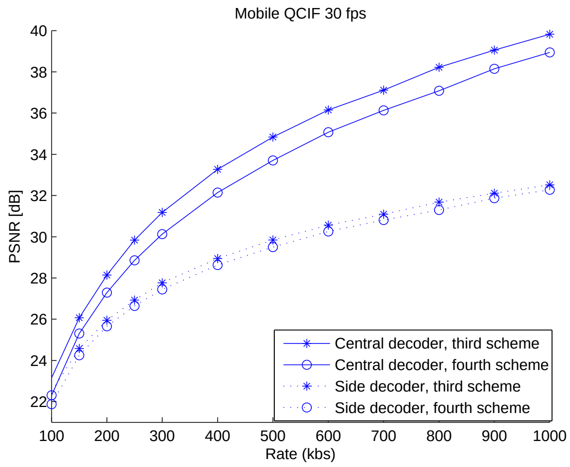

Figure 9: Performance comparison of the third and fourth MDC schemes (Mobile, QCIF 30 fps).

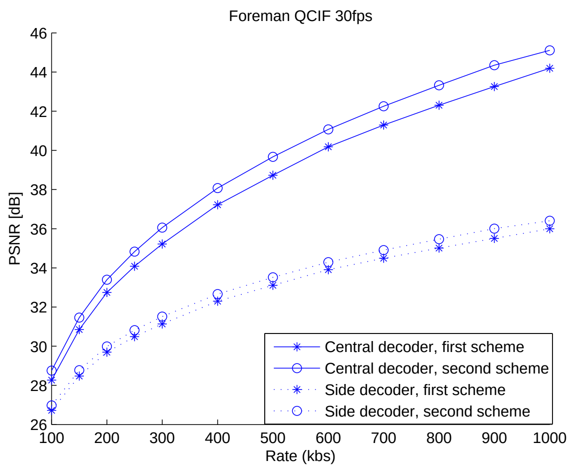

Figure 6: Performance comparison of the first and second MDC schemes (Foreman, QCIF 30 fps).

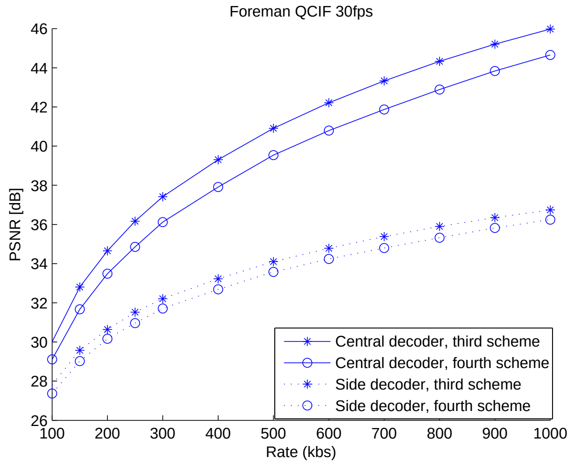

Figure 7: Performance comparison of the third and fourth MDC schemes (Foreman, QCIF 30 fps).

favor of the second and third schemes.

Future work will focus on studying the role of a more complex interpolator in improving the performance of side decoders. It would also be interesting to check if similar conclusions are obtained if instead of the MCTF encoding a more classical hybrid video codec is used.

REFERENCES

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

[2] L. Ozarow, "On a source-coding problem with two channels and three receivers," *Bell System Technical Journal*, vol. 59, no. 10, pp. 1909–1921, December 1980.

[3] A. A. El Gamal and T. M. Cover, "Achievable rates for multiple descriptions," *IEEE Transactions on Information Theory*, vol. 28, no. 6, pp. 851–857, November 1982.

[4] R. Venkataramani, G. Kramer, and V. K. Goyal, "Multiple description coding with many channels," *IEEE Transactions on Information Theory*, vol. 49, no. 9, pp. 2106–2114, September 2003.

[5] J. G. Apostolopoulos, "Reliable video communication over lossy packet networks using multiple state encoding and path diversity," in *Visual Communications and Image Processing*, B. Girod, C. A. Bouman, and E. G. Steinbach, Eds. San Jose, CA: Proc. SPIE 4310, January 2001, pp. 392–409.

[6] W. S. Lee, M. R. Pickering, M. R. Frater, and J. F. Arnold, "A robust codec for transmission of very low bit-rate video over channels with bursty errors," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 10, no. 8, pp. 1403–1412, December 2000.

[7] A. R. Reibman, H. Jafarkhani, Y. Wang, M. T. Orchard, and R. Puri, "Multiple-description video coding using motion-compensated temporal prediction," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 12, no. 3, pp. 192–204, March 2002.

[8] I. V. Bajić and J. W. Woods, "Domain-based multiple description coding of images and video," *IEEE Transactions on Image Processing*, vol. 12, no. 10, pp. 1211–1225, October 2003.

[9] N. Franchi, M. Fumagalli, R. Lancini, and S. Tubaro, "Multiple description video coding for scalable and robust transmission over IP," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 15, no. 3, pp. 321–334, March 2005.

[10] Y. Wang, A. R. Reibman, and S. Lin, "Multiple description coding for video delivery," *Proceedings of the IEEE*, vol. 93, no. 1, pp. 57–70, January 2005.

[11] V. A. Vaishampayan and S. John, "Balanced interframe multiple description video compression," in *Proceedings of the International Conference on Image Processing*, vol. 3, Kobe, Japan, October 1999, pp. 812–816.

[12] Y. Wang and S. Lin, "Error-resilient video coding using multiple description motion compensation," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 12, no. 6, pp. 438–452, June 2002.

[13] M. van der Schaar and D. S. Turaga, "Multiple description scalable coding using wavelet-based motion compensated temporal filtering," in *Proceedings of the International Conference on Image Processing*, vol. 3, Barcelona, Spain, September 2003, pp. 489–492.

[14] C. Tillier, B. Pesquet-Popescu, and M. van der Schaar, "Multiple descriptions scalable video coding," in *Proceedings of the European Signal Processing Conference*, Vienna, Austria, September 2004.

[15] J. Kim, R. M. Mersereau, and Y. Altunbasak, "Network-adaptive video streaming using multiple description coding and path diversity," in *Proceedings of the IEEE International Conference on Multimedia and Expo*, vol. 2, Baltimore, MD, July 2003, pp. 653–656.

[16] N. Franchi, M. Fumagalli, G. Gatti, and R. Lancini, "A novel error-resilience scheme for a 3-D multiple description video coder," in *Proceedings of the Picture Coding Symposium*, San Francisco, CA, December 2004.

[17] S. Cho and W. A. Pearlman, "Error resilient compression and transmission of scalable video," in *Applications of Digital Image Processing XXIII*, A. G. Tescher, Ed. San Diego, CA: Proc. SPIE 4115, August 2000, pp. 396–405.

[18] C. Tillier and B. Pesquet-Popescu, "3D, 3-band, 3-tap temporal lifting for scalable video coding," in *Proceedings of the International Conference on Image Processing*, Barcelona, Spain, 2003.

[19] P. Chen and J. W. Woods, "Bidirectional MC-EZBC with lifting implementation," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 14, no. 10, pp. 1183–1194, October 2004.