A DISTRIBUTED VIDEO CODING APPROACH FOR MULTIPLE DESCRIPTION VIDEO TRANSMISSION OVER LOSSY CHANNELS

S. Milani and G. Calvagno

Dept. of Information Engineering, University of Padova
via Gradenigo 6/B, 35131 Padova, Italy Phone: +39 049 827 7641, Fax: +39 049 827 7699,
e-mail: {`simone.milani,calvagno`}`@dei.unipd.it`

ABSTRACT

Recent works have proposed the combination of Multiple Description Coding (MDC) schemes with the novel Distributed Video Coding (DVC) paradigm in order to enable robust video transmission over lossy channels. DVC schemes permit mitigating the drifting of channel distortion that characterizes the traditional MDC schemes and prove to be extremely effective at high packet loss rates. The paper presents a Multiple Description Distributed Video Coder (MDDVC) that combines a polyphase-subsampling MDC approach with a novel hybrid Distributed Source Coding technique that processes the video signal both in the pixel and in the transform domain. Tests on simulated lossy channels show that at high loss probabilities the PSNR value of the reconstructed video sequence improves with respect to its traditional MDC counterpart. Moreover, an adaptive approach is also presented combining both the traditional and the DSC-based MDC schemes in order to maximize the quality of the sequence reconstructed at the decoder for all the network conditions.

1. INTRODUCTION

The transmission of video sequences over wireless networks presents several issues due to the presence of delays, packet losses and bandwidth limitations. In order to overcome these problems, different robust video coding strategies have been proposed in the recent literature. Among these, Multiple Description Coding (MDC) schemes [1] have proved to be significantly more efficient with respect to traditional single description coding (SDC) architectures for wireless networks, where channel conditions are time-varying and the bursty nature of packet losses can lead to a significant distortion of the reconstructed sequence. In addition, several video coding architectures based on the information-theoretic concept of Distributed Source Coding (named Distributed Video Coding or DVC schemes) have been proposed during the last years [2]. DVC approaches permit an independent coding of the data at the transmitter, and a robust predictive decoding at the receiver, which can be performed correctly provided that at least one predictor is available from a set of possible candidates.

Recent works have been focusing on combining these two techniques in Multiple Description Distributed Video Coders (MDDVC) since coding the prediction residual using DVC permits reducing the drifting problem in traditional MDC [3]. In [4] Crave *et al.* propose an MDDVC approach based on motion compensation temporal filtering. In this case, the adopted DVC solution is able to mitigate the inefficiency of the original MDC scheme reducing the amount of redundancy introduced in the data stream. The solution proposed in [5] applies a Wyner-Ziv coding technique to a multiple description scalar quantizer, while the approach by Fan *et al.* [6] associates to each description some hash information, which is transmitted with the packet stream of the other description. In addition, the approach presented in [7] by Wang *et al.* adopts a zero-padding of DCT coefficients to generate the different descriptions, which are coded by a SPIHT-based Slepian-Wolf coder. Most of the proposed MDDVC approaches process the video signal in the transform domain, which permits obtaining high compression efficiency, but implies a difficult modeling of the correlation between coefficients belonging to different blocks.

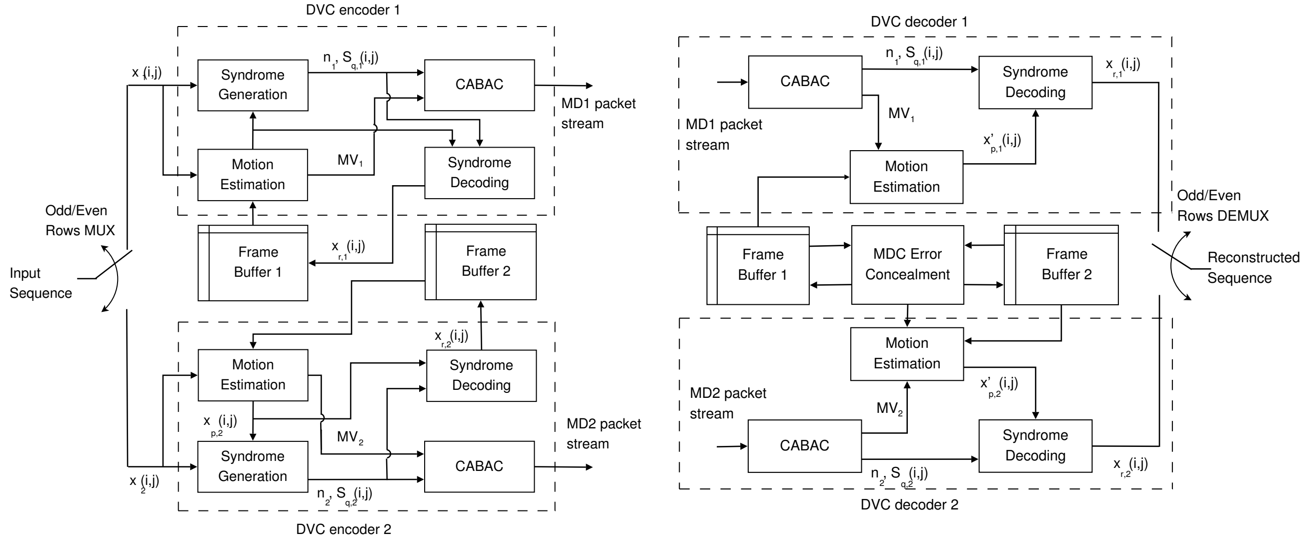
The paper proposes a novel MDDVC scheme that adopts a hybrid DVC unit that processes the video signal both in the spatial and in the temporal domain. More precisely, the correlation between blocks is estimated in the spatial domain, though the generated syndromes are compressed via a transform coding algorithm. The scheme generates a set of "distorted" syndromes, which can be efficiently compressed and permit reconstructing the coded signal with a satisfying visual quality. Experimental results will show that the proposed approach permits improving the PSNR value of the reconstructed sequence of 1 dB at most with respect to the original MDC scheme.

In the following, Section 2 provides an overview of the implemented codec, while Section 3 describes the Distributed Source Coding (DSC) scheme that have been adopted to characterize the prediction residual signal. Section 4 presents an enhanced adaptive solution that combines both traditional MDC and MDDVC techniques in order to obtain an enhanced MDC strategy that adapts to the channel characteristics. Experimental results in Section 5 show how the proposed DSC approaches improves the quality of the coded sequences at high loss rates. Conclusions are drawn in Section 6.

2. THE PROPOSED MDC ARCHITECTURE

Many MDDVC schemes that have been presented in literature are derived from previous traditional MDC counterparts replacing the traditional temporally-predictive video coder with a Wyner-Ziv Video coding scheme (see [3] as an example). Similarly, the proposed coder relies on a traditional polyphase-subsampling MDC architecture and adopts a novel DVC video coding paradigm to code each description separately.

Figure 1 shows the block diagrams of the employed encoder and decoder. The input sequence is split into two descriptions sampling odd and even rows of pixels, which are sent to two separate H.264/AVC-based DVC coders. Each coder has been designed reusing the block structure of H.264/AVC, where the residual coding adopts a DSC-based



(a) Encoder (b) Decoder

Figure 1: Block diagrams for the proposed encoder and decoder.

algorithm in place of the traditional solution based on the Displaced Frame Difference (DFD).

Each 4×4 block of each description is approximated by the motion estimation unit generating a predictor block identified by a motion vector. The residual signal is then coded using a Distributed Source Coding paradigm (described in Section 3) creating two packet streams (associated to the two descriptions) that are independently sent to the receiver. Each packet stream is decoded using an independent DVC decoder, which is helped by a multiple description error concealment whenever some information is lost. In case one description is missing, it is possible to estimate the lost rows interpolating the pixel rows of the other description. As a result, the corresponding decoder can use a degraded version of the missing information, which nevertheless permits a correct decoding of the following frames in case it is sufficiently correlated with the lost data (more details will be given in Section 3).

3. A DISTRIBUTED SOURCE CODING APPROACH FOR CODING THE PREDICTION RESIDUALS

In the technical literature DVC approaches can be divided into two main groups: schemes that process the video signal in the pixel domain (see [8] as an example) and schemes that operate in the transform domain (like the coder by Aaron *et al.* [9] and the PRISM coder by Puri *et al.* [10]). The former require a lower computational complexity and prove to be effective when no feedback channel is available, but they have a limited compression efficiency. The latter provide higher compression ratios, but the needed computational effort is considerably higher, since the motion search in the decoding process is performed in the transform domain. In our work, we adopted a hybrid pixel-transform domain DVC scheme that permits reducing the decoding complexity and improves the robustness of the video stream to packet losses with high compression gains.

Given the current 4×4 block $\mathbf{x}_m$ of pixels for description MDm ($m = 1, 2$) and its predictor $\mathbf{x}_{p,m}$, for each pixel $x_m(i,j)$ of block $\mathbf{x}_m$ at position (i,j), $i,j = 0,\dots,3$, we compute the number of bits $n_m(i,j)$ as

$$n_m(i,j) = \begin{cases} \lfloor \log_2 (d_m(i,j)) \rfloor + 2 & \text{if } d_m(i,j) < \delta \\ 0 & \text{otherwise,} \end{cases} \qquad (1)$$

where $d_m(i,j) = |x_m(i,j) - x_{p,m}(i,j)|$ and δ is a threshold value depending on the Quantization Parameter (QP) chosen for the current block (in our setting, we have set $\delta = \Delta/12$ where Δ is the quantization step associated to the current QP). Then, the coding unit computes the maximum value

$$n_m = \max_{i,j=0,\dots,3} n_m(i,j) \qquad (2)$$

within the current block and, in case n_m is higher than 0, it generates a block of syndromes $s_m(i,j)$ via the following equation

$$s_m(i,j) = x_m(i,j) \& (2^{n_m} - 1) \qquad (3)$$

where the symbol & denotes a bitwise AND operator. In this way, the n_m least significant bits of each pixel are selected to generate the block $\mathbf{s}_m$. The value $s_m(i,j)$ is called "syndrome" (as in [10]), while the parameter n_m is the number of syndrome bits. The block $\mathbf{s}_m$ is then transformed via the 4×4 H.264/AVC integer DCT into the block $\mathbf{S}_m$, which is quantized into the block $\mathbf{S}_{q,m}$, dequantized, and inversely-transformed into the block $\mathbf{s}_{r,m}$. In our implementation, we adopt the same set of quantization steps adopted in the H.264/AVC standard [11]. The syndrome $s_{r,m}(i,j)$ is a lossy version of the original syndrome $s_m(i,j)$ such that $s_{r,m}(i,j) = s_m(i,j) + e_m(i,j)$, where $e_m(i,j)$ is the distortion introduced by the quantization in the transform domain. Each lossy syndrome $s_{r,m}(i,j)$[1] identifies a different quantizer $Q_{s_{r,m}(i,j)}$ with quantization step 2^{n_m} and offset $s_{r,m}(i,j)$ such that the reconstruction levels for the adopted quantizer can be written as $s_{r,m}(i,j) + k\, 2^{n_m}$, where $k \in \mathbb{Z}$ is the associated index of the quantized value.

---

[1] Here the term *"lossy"* in relation to the syndromes refers to the lossy coding scheme adopted to characterize them.

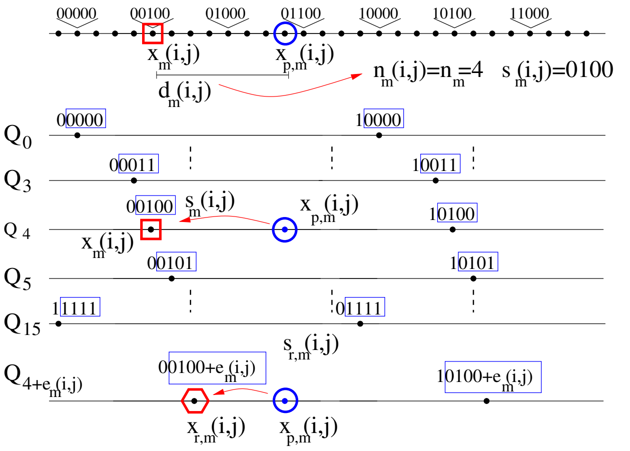


Figure 2: Example of syndrome generation and decoding. Note that using the original syndrome $s_m(i,j)$ leads to a lossless reconstruction of $x_m(i,j)$. The proposed scheme transforms and quantizes $s_m(i,j)$, and as a consequence, the final quantizer is $Q_4 + e_m(i,j)$ in place of Q_4.

Given the predictor block $\mathbf{x}_{p,m}$, it is possible to reconstruct the coded pixel $x_{r,m}(i,j) = x_m(i,j) + e_m(i,j)$ by quantizing $x_{p,m}(i,j)$ using the quantizer characteristics associated to $s_{r,m}(i,j)$. As an example, Figure 2 reports the syndrome generation and reconstruction processes for the pixel $x_m(i,j)$ of description MDm. According to the distance $d = 7$ from its predictor $x_{m,p}(i,j)$, the syndrome $s_m(i,j) = 0100$ (associated to the shifted quantization characteristics Q_4) is made of the $n_m = 4$ least significant bits of $x_m(i,j)$. After coding $s_m(i,j)$ into the syndrome $s_{r,m}(i,j)$ (associated to the characteristics $Q_4 + e_m(i,j)$), it is possible to reconstruct a distorted version of $x(i,j)$ by quantizing $x_{m,p}(i,j)$ with $Q_4 + e_m(i,j)$.

Note that a correct decoding is also possible using a different predictor $x'_{p,m}(i,j) \neq x_{p,m}(i,j)$ provided that the correlation between $\mathbf{x}_m$ and $\mathbf{x}'_{p,m}$ is the same or higher (i.e. the difference $d'_m(i,j) = |x_m(i,j) - x'_{p,m}(i,j)|$ in eq. (1) leads to a value $n'_m(i,j) \leq n_m(i,j)$). The higher is n_m, the more robust is the stream since a larger number of possible candidate predictors can be used in a successful decoding at the expense of increasing the coded bit rate.[2]

The decoding scheme of the proposed DVC approach is similar to the scheme adopted in the PRISM coder (see [10]), made exception for the fact that it codes syndromes with a lossy technique and operates in the pixel domain. However, this difference permits reducing the computational complexity at the decoder since the motion search is performed in the pixel domain and the inverse quantization and transform are not strictly related to syndrome decoding.

Moreover, the variances of coefficient distributions significantly change depending on the spatial frequencies, and therefore, the robustness and the compression efficiency of the scheme can be seriously degraded by a wrong modelization. As an example, the PRISM-like coder in [12] specifies different syndrome bits $n_m(i,j)$ for each coefficient and requires modifying the context structure of the arithmetic coder since the statistics of the coded data is completely changed.

In the proposed coder, we simplify the characterization of the correlation between $\mathbf{x}_m$ and $\mathbf{x}_{p,m}$ since n_m does not significantly vary depending on (i,j). Moreover we are able to easily reuse the coding contexts and several blocks of the H.264/AVC coder [11] since the coded signal $S_{q,m}(i,j)$ presents a statistics similar to the Intra coefficients.

The adoption of a DSC coding technique in place of a traditional residual coding scheme, like that of H.264/AVC, reduces the compression efficiency (as syndromes need more bits to be coded), but improves the robustness of the coding scheme at high loss rates (see results in Section 5), which correspond to the channel conditions where MDC provides a better video quality to the end-user with respect to traditional FEC techniques, like the adoption of channel codes [13]. However, for sequences with high vertical correlation the traditional MDC scheme provides a better performance with respect to the DSC-based scheme since the lost information can be estimated quite well by the error concealment algorithm. The same outcome can be noticed at low loss rates since the introduced distortion is quite small. In these cases, the error drifting is limited, and the lower source coding distortion introduced by traditional MDC schemes permits a better estimate of the lost parts. As a consequence, it is possible to design an effective adaptive strategy that employs both the DSC unit and the traditional residual coding methods.

4. A HYBRID DFD/DSC APPROACH FOR MULTIPLE DESCRIPTION VIDEO CODING

The previous section has presented a novel DVC scheme that significantly improves the quality of the reconstructed sequence at high packet loss rates and for sequences with a medium or low vertical correlation. As a consequence, it is possible to design an MDC coding strategy that adapts itself to the channel conditions and the characteristics of the video sequence in order to maximize the video quality of the signal reconstructed at the decoder.

For each 32×16 block B of the input sequence (corresponding to one macroblock per description), the adaptive strategy evaluates the vertical gradient measure

$$G_v = \frac{1}{256} \sum_{i,j=0}^{15} |B(2\,i, j) - B(2\,i+1, j)|. \tag{4}$$

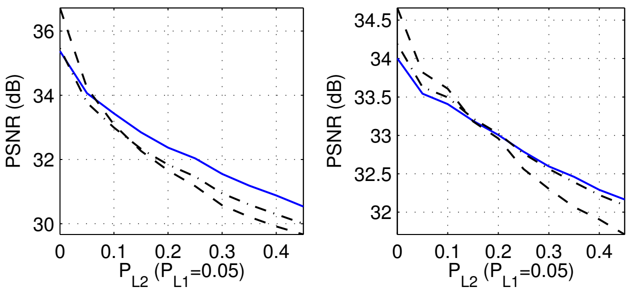
Depending on the value G_v and on the estimated packet loss probability P_L, the coding strategy adopts either the DSC based or the DFD-based coding approach for the current macroblock according to the following rule

```
**if** $(G_v < T_0$ OR $P_L < P_{Th1})$ OR $(G_v < T_1$ AND $P_L < P_{Th2})$ **then**
    use DFD-based coder;
**else**
    use DSC-based coder;
**end if**.
```

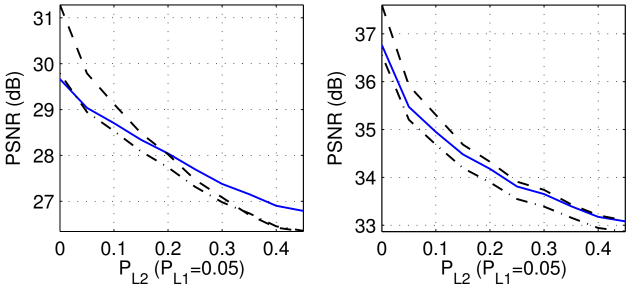
On the basis of an extensive set of simulations, we have set the threshold T_0 equal to 3.5 and T_1 equal to 7. As for the loss probability, an extensive set of simulations have lead to set the threshold P_{Th1} to 0.1 and the threshold P_{Th2} to 0.12. In this way, the coding mode adapts itself to the changing characteristics of the signal and of the channel. The loss percentage P_L can be estimated from the control packets of the transmission protocol (e.g it is possible to use the data from

---

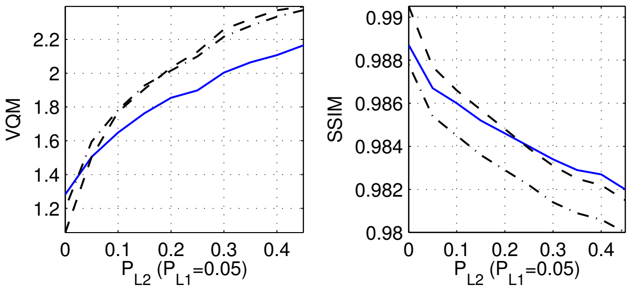
[2] Note that for $n_m = 8$ we have Intra coding for uncompressed video signals sampled with 8 bits/pixel.



(a) `foreman` $R_b^i = 800$ kbit/s (b) `football` $R_b^i = 800$ kbit/s



(c) `bus` $R_b^i = 1000$ kbit/s (d) `news` $R_b^i = 500$ kbit/s



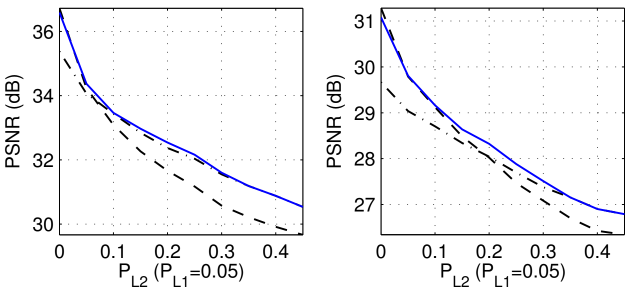
(e) `foreman` $R_b^i = 800$ kbit/s (f) `foreman` $R_b^i = 800$ kbit/s

Figure 3: Experimental results for different MDC algorithms with different residual coding units. The graphs report the average values of PSNR, VQM and SSIM metrics for the reconstructed sequences vs. the packet loss probability P_{L2} for description MD2 (the loss probability P_{L1} for description MD1 is set to 0.05). The solid lines denote the data for the DSC-based coder, the dashed line those for traditional DFD-based coder, and the dash-dotted line those for the PRISM-like coder.

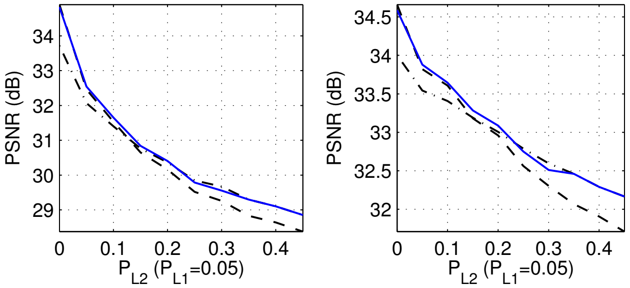
RTCP packets in case we are using RTP for video transmission or we can rely on the IEEE 802.11k protocol that reports the number of lost frame from which it is possible to estimate the percentage of lost packets). Experimental results in the following section show how this adaptation significantly improves the performance of the scheme permitting the encoder to find out the optimal setting.

5. EXPERIMENTAL RESULTS

The proposed DSC and hybrid DFD/DSC coders have been tested simulating the transmission of different video sequences in a scenario where packet streams are affected by losses. More precisely, each description is transmitted on an independent channel, which is simulated using a Gilbert two-state model with burst length $L_B = 4$. The loss probability P_{L2} for description MD2 varies within the range $[0.05, 0.45]$ with steps of 0.05, while the loss probability P_{L1} for description MD1 is fixed to 0.05. In this way, it is possible to grant a certain diversity between loss patterns that improves the performance of MDC schemes. In our tests we coded different CIF sequences at different bit rates R_b^i, $i = 1, 2$, with GOP structure IPPP, slices of 22 macroblocks, and CABAC entropy coding. The adopted rate-distortion optimization strategy and the rate control algorithms are those defined within the JVT for the H.264/AVC coder.



(a) `foreman` $R_b^i = 800$ kbit/s (b) `bus` $R_b^i = 1000$ kbit/s



(c) `table` $R_b^i = 800$ kbit/s (d) `football` $R_b^i = 1000$ kbit/s

Figure 4: Experimental results for different MDC algorithms with different residual coding units. The graphs report the average PSNR values of the reconstructed sequences vs. the packet loss probability P_{L2} for description MD2 (the loss probability P_{L1} for description MD1 is set to 0.05). The dash-dotted lines denote the data for the DSC-based coder, the dashed line those for traditional DFD-based coder, and the solid line those for the Hybrid DFD/DSC coder.

The plots in Figure 3 report the values of the PSNR, VQM [14] and SSIM[15] metrics averaged over 10 channel realizations for a given P_{L2}. In each figure we compare the proposed MDDVC algorithm (referenced as DSC-based) with the traditional MDC scheme (adopted in [13]) based on DFD (referenced as DFD-based) and the MDDVC algorithm obtained replacing the syndrome generation unit in the scheme of Fig. 1(a) with syndrome generation strategy of the DSC coding scheme in [12] (referenced as PRISM-like). It is possible to notice that for most of the sequences the DSC-based scheme improves the quality of the reconstructed sequences whenever the loss probability becomes significant. As an example, Fig. 3(a) shows that for $P_{L2} > 0.1$ the PSNR value for the DFD-based scheme is always lower with respect to the proposed coder (the difference is approximately 1 dB for $P_{L2} = 0.4$). The same improvement can be noticed using the other quality metrics VQM and SSIM (see Fig. 3(e) and 3(f)). Note that the performance of the PRISM-like coder is either the same or worse than the performance of the DFD-based approach. This fact is partly due to the adopted rate-distortion and rate-control algorithms, whose parameters are optimized and tuned for the H.264/AVC coder, and partly due to the difficulties in estimating and characterizing the correlation of transform coefficients (the DSC scheme we are considering do not have a feedback channel).

The same behavior can be noticed for other sequences too (see Fig. 3(b) and 3(c)), despite the fact that the crossing point between the DFD and the DSC plots depends on the characteristics of the video sequences. As a matter of fact, for the sequence `bus`, which presents a higher vertical correlation, the DSC-based approach becomes competitive for a higher P_{L2} value with respect to the sequence `foreman`. This fact is utterly evidenced by the results for the sequence `news` (see Fig. 3(d)), where the high vertical correlation and the low amount of motion allow the error concealment to perform quite well, and therefore, the PSNR values of the traditional DFD scheme are always higher or equal to those of the DSC scheme. Therefore, the adaptive algorithm presented in Section 4 proves to be an optimal solution since it is able to switch from the DSC setting to the DFD setting depending on the correlation measure G_v and on the estimated channel loss probability. Figure 4 reports the average PSNR values vs. P_{L2} for the DSC-based, the DFD-based and the hybrid algorithms. It is possible to notice that for all the sequences the algorithm is able to always choose the best configuration permitting also a small improvement for the sequence `bus` near the crossing point at $P_{L2} = 0.17$ (see Fig. 4(b)). This fact is possible since the adaptivity of the approach makes possible to find the optimal configuration according to the characteristics of the channel and of the sequence at different time intervals.

6. CONCLUSION

The paper proposed an MDDVC scheme that generates two descriptions using a polyphase subsampling of odd and even rows and codes them using a DSC coder based on lossy syndromes. The proposed scheme outperforms its traditional MDC counterpart (up to 1 dB in PSNR) and other DSC solutions at high loss rates and whenever the vertical correlation of the input sequence is not too high. An adaptive approach, which switches from the DSC coding scheme to the traditional one, compensates this drawback maximizing the quality of the reconstructed sequence at the decoder in all the cases. Future work will be devoted to extend the designed approach to more effective MDC schemes and to a scalable coding of the input sequence in terms of temporal resolution and visual quality.

REFERENCES

[1] V. K. Goyal, “Multiple Description Coding: Compression Meets The Network,” *IEEE Signal Processing Mag.*, vol. 8, no. 5, pp. 74–93, Sept. 2001.

[2] B. Girod, A. Aaron, S. Rane, and D. Rebollo-Monedero, “Distributed Video Coding,” *Proc. of the IEEE, Special Issue on Video Coding and Delivery*, vol. 93, no. 1, pp. 71–83, Jan. 2005, Invited Paper.

[3] J. Wang, X. Wu, S. Yu, and J. Sun, “Multiple Descriptions in the Wyner-Ziv Setting,” in *Proc. of 2006 IEEE International Symposium on Information Theory (ISIT 2006)*, Seattle, WA, USA, June 9 – 14, 2006, pp. 1584 – 1588.

[4] O. Crave, C. Guillemot, B. Pesquet-Popescu, and C. Tillier, “Robust Video Transmission Based on Distributed Multiple Description Coding,” *EURASIP Journal on Wireless Communications and Networks: Special issue on Multimedia over Wireless Network*, no. 1, Jan. 2008.

[5] M. Wu, A. Vetro, and C. W. Chen, “Multiple Description Image Coding with Distributed Source Coding and Side Information,” in *Proceedings of SPIE: Multimedia Systems and Applications VII*, vol. 5600, Philadelphia, PA, USA, Oct. 25 – 27, 2004, pp. 120 - 127.

[6] Y. Fan, J. Wang, J. Sun, P. Wang, and S. Yu, “A Novel Multiple Description Video Codec Based on Slepian-Wolf Coding,” in *Proc. of the Data Compression Conference (DCC 2008)*, Snowbird, UT, USA, Mar. 25 – 27, 2003, p. 515.

[7] A. Wang, Y. Zhao, and H. Bai, “Robust multiple description distributed video coding using optimized zero-padding,” *Science in China Series F: Information Sciences*, vol. 52, no. 2, pp. 206 – 214, Feb. 2009.

[8] C. Brites, J. Ascenso, and F. Pereira, “Modeling Correlation Noise Statistics at Decoder for Pixel Based Wyner-Ziv Video Coding,” in *Proc. of PCS 2006*, Beijing, China, Apr. 24 – 26, 2006.

[9] A. Aaron, R. Zhang, and B. Girod, “Wyner-ziv coding for motion video,” in *Proceedings of Asilomar Conference on Signals, Systems and Computers 2002*, vol. 1, Pacific Grove, CA, USA, Nov. 2002, pp. 240 – 244.

[10] R. Puri and K. Ramchandran, “PRISM: A new robust video coding architecture based on distributed compression principles,” in *Proc. of the* 40^{th} *Allerton Conference on Communication, Control and Computing*, Allerton, IL, USA, Oct. 2002, pp. 402–408.

[11] T. Wiegand, “Version 3 of H.264/AVC,” in *Joint Video Team (JVT) of ISO/IEC MPEG & ITU-T VCEG (ISO/IEC JTC1/SC29/WG11 and ITU-T SG16 Q.6),* 12^{th} *Meeting*, Redmond, WA, USA, July 17 – 23 2004.

[12] S. Milani, J. Wang, and K. Ramchandran, “Achieving H.264-like compression efficiency with distributed video coding,” in *Proc. of VCIP 2007*, San Jose, CA, USA, Jan. 28 – Feb. 1, 2007, pp. 65 082Z–1 – 65 082Z–12.

[13] S. Milani, G. Calvagno, R. Bernardini, and P. Zontone, “Cross-Layer Joint Optimization of FEC Channel Codes and Multiple Description Coding for Video Delivery over IEEE 802.11e Links,” in *Proc. of IEEE FMN 2008 (co-located with NGMAST2008)*, Cardiff, Wales, GB, Sept.17 – 18, 2008, pp. 472 – 478.

[14] F. Xiao. (2000, Winter) DCT-based Video Quality Evaluation. MSU Graphics and Media Lab (Video Group). [Online]. Available: http://compression.ru/video/quality_measure/vqm.pdf

[15] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, “Image Quality Assessment: From Error Visibility to Structural Similarity,” *IEEE Trans. Image Processing*, vol. 13, no. 4, pp. 600 – 612, Apr. 2004.