Multiple description coding for SNR scalable video transmission over unreliable networks

Roya Choupani · Stephan Wong · Mehmet Tolun

Abstract Streaming multimedia data on best-effort networks such as the Internet requires measures against bandwidth fluctuations and frame loss. Multiple Description Coding (MDC) methods are used to overcome the jitter and delay problems arising from frame losses by making the transmitted data more error resilient. Meanwhile, varying characteristics of receiving devices require adaptation of video data. Data transmission in multiple descriptions provides the feasibility of receiving it partially and hence having a scalable and adaptive video. In this paper, a new method based on integrating MDC and signal-to-noise ratio (SNR) scalable video coding algorithms is proposed. Our method introduces a transform on data to permit transmitting them using independent descriptions. Our results indicate that on average 1.71dB reduction in terms of Y-PSNR occurs if only one description is received.

Keywords Scalable video coding · Multiple description coding ·
Multimedia transmission

R. Choupani · S. Wong
Computer Engineering Department, TU Delft, Delft, The Netherlands

S. Wong
e-mail: J.S.S.M.Wong@tudelft.nl

R. Choupani (✉)
Computer Engineering Department, Çankaya University, Ankara, Turkey
e-mail: roya@cankaya.edu.tr

M. Tolun
Computer Engineering Department, TED University, Ankara, Turkey
e-mail: mehmet.tolun@tedu.edu.tr

1 Introduction

With the steady increase in the Internet access bandwidth, increasingly more applications utilize streaming audio and video contents [27]. This trend has been further intensified by the appearance of small, powerful hand-held terminals (such as mobile phones, iPods, and tablet PCs) in the market. In streaming video applications, the servers normally have to serve a large number of users with different screen resolutions and network bandwidth and processing capabilities. Hence, an encoding method that makes use of a single encoded data for all types of bandwidth channels and displaying device capacities is of remarkable significance in multimedia applications. Scalable Video Coding (SVC) schemes are intended to be a solution for the Internet heterogeneity and receiving devices diversity problem by encoding the data at the highest quality but enabling the receiver to utilize it partially depending on its screen, memory, or processing capabilities, or the available bandwidth [15, 19, 26]. However, communication networks offer channels with varying bandwidth [8, 15] which together with the higher rate of frame loss or corruption in wireless networks becomes a complicated issue for video streaming. On the other hand, the main drawback of the currently available scalable video coding methods is that they are not suitable for non-reliable environments with a high rate of frame loss or corruption. This problem stems from the fact that the SVC methods are based on the Motion-Compensated Temporal Filtering (MCTF) scheme [11] where the frames are coded as the difference with a (generally prior) reference frame. In case of a reference frame loss or corruption, the whole chain of the motion compensated frames which depend on this reference frame becomes unrecoverable. To increase the error resilience of the video coding schemes, Multiple Description Coding (MDC) methods were introduced [12, 22, 25]. These methods improve the error resilience of the video by adding redundancy to the encoded data. In case a frame is lost or corrupted, redundancy is used to replace it with an estimated frame. Some researchers have considered the frame loss problem and have not addressed the scalability issue. Franchi et al. proposed a method to send a video by utilizing independent multiple descriptions. However, their method does not combine scalability features with multiple description coding and therefore does not deal with the bandwidth variation problem [10]. The combination of scalable video coding methods and multiple description coding has been addressed by some researchers recently [3, 15, 19]. In the proposed approaches the video data is partitioned into disjoint sets such as the group of odd and even frames in temporal MDC. These approaches take advantage of the correlation between the adjacent data items for estimating the lost data. However, when considering the signal-to-noise-ratio scalability, the assumption of correlated data is not valid, because the bits composing a pixel value cannot be interpolated from each other. An intuitive example is putting the more significant bits of a pixel value in one description and less significant bits in another one. The more significant bits cannot be estimated from the less significant bits in case that they are lost during the transmission. In this study, we propose a method which aims at expressing SNR scalable video coding scheme by multiple equivalent descriptions. In order to achieve this aim, we propose a transform which allows the data bits to have a contribution in each description. In this way, each description besides to conveying the most basic part of data values, is capable of refining the basic part of data. Our proposed method falls into the class of methods which combine MDC with SVC schemes. Our results

indicate that on average 1.71dB reduction in terms of Y-PSNR occurs if only one description is received. The remainder of this paper is organized as follows: Section 2 introduces the main multiple description coding methods. Section 3 describes the details of our proposed method. In Section 4, we introduce the theoretical base of our performance evaluation method and provide the experimental results and finally, in Section 5, we draw our conclusions.

2 MDC-based video coding techniques

Multiple descriptions have attracted a lot of attention as an error resilient way of encoding and communicating visual information over lossy packet networks. A multiple description coder divides the video data into several bit-streams called descriptions which are subsequently transmitted separately over the network. All descriptions are equally important and each description can be decoded independently from other descriptions which means that the loss of some of these descriptions does not affect decoding of the others. The accuracy of the decoded video depends on the number of received descriptions. Figure 1 depicts the basic framework for a multiple description encoder/decoder with two descriptions. In case of a failure in one of the channels, the output signal is recovered from the other description only. Descriptions are defined by constructing P non-empty *sets* summing up to the original signal f. Each set in this definition corresponds to one description. The sets however, are not necessarily disjoint. A signal sample may appear in more than one set to increase error resilience property of the video. Repeating a signal sample in multiple descriptions is also a way for assigning higher importance to some parts/signals of the video. The more a signal sample is repeated the more reliably it is transmitted over the network. The duplicate signal values increase the redundancy which results in a subsequent increase in the data size and reduced efficiency. Designing descriptions as partition, does not necessarily mean that there will be no redundancy in the data. In fact, designing the descriptions as partitions prevents extra bits to be added to the original data for error resilience but still a redundancy in the form of reduced coding efficiency exits. In case of a data loss, the correlation between the spatially or temporally close data can be used for estimating the lost bits. The estimation process is commonly referred

Fig. 1 Multiple descriptions coding block-diagram
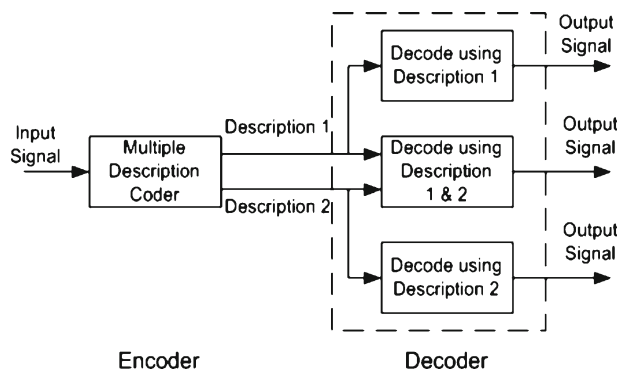
to as error concealment and relies on the preserved correlation in constructing the descriptions. MDC schemes for video transmission can be classified as below:

– Multi-layer MDC schemes partition the video into one base layer and one or several enhancement layers [5]. The base layer can be decoded independently from enhancement layers but it provides only the minimum spatial, temporal, or signal-to-noise ratio quality. The enhancement layers are not independently decodable. An enhancement layer improves the decoded video obtained from the base layer. MDC schemes based on multi-layers puts the base layer together with one of the enhancement layers at each description. This helps to partially recover the video when data from one or some of the descriptions are lost or corrupted. Repeating base layer bits in each description is the overhead added for a better error resilience. In [1] the authors propose to generate multiple scalable descriptions from a single SVC bit-stream by mapping scalability layers of different frames to different descriptions. Their scheme is intended for Peer-to-Peer (P2P) streaming over multiple multicast trees and features several encoding parameters, such as base layer rate of descriptions and overall redundancy, to optimize for mean rate-distortion performance of each description received over a packet loss network, range of extraction points of the SVC stream, and overall redundancy of their MDC scheme. In [9] the SVC is combined with MDC schemes, by sub-sampling in both horizontal and vertical directions which yields four subsequences. The authors use two approaches to combine the subsequences into two descriptions. In the first approach, each description is encoded by predicting one subsequence from the other using the inter layer prediction tools. The second approach exploits the redundancy between the subsequence with the hierarchical dyadic B frame prediction algorithm. The authors in [17] present a solution for the differences in the types of delivered services in H.264-based SVC combined with MDC by using optimization and control strategies. In [18] an algorithm is proposed to control the mismatch between the prediction loops at the encoder and decoder in MDC with motion-compensated predictions. They consider three cases when both descriptions received or either of the single descriptions is received.

– Forward Error Correction (FEC)-based MDC methods assume that the video is originally defined in a multi-resolution manner [16, 23]. This means if we have M levels of quality, each one is adding to the fidelity of the video with respect to the original one. This concept is similar to the multi-layer video coding method used by FGS scheme. The main difference, however, is that there exists a mandatory order in applying the enhancements. In other words, it is sensitive to the position of the losses in the bitstream, e.g., a loss early in the bitstream can render the rest of the bitstream useless to the decoder. FEC-based MDCs aim to develop the desired feature that the delivered quality become dependent only on the fraction of packets delivered reliably. One method to achieve this is Reed Solomon block codes. Mohr et al. [14] used Unequal Loss Protection (ULP) to protect video data against packet loss. ULP is a system that combines a progressive source coder with a cascade of Reed Solomon codes to generate an encoding that is progressive in the number of descriptions received, regardless of their identity or order of arrival. In [28] a 2-D layered multiple description coding (2DL-MDC) for error-resilient video transmission over unreliable networks is used which encodes each group of pictures (GOP) using the SVC extension of H.264

into sub-streams. First dimension of encoding uses temporal scalability while the second dimension uses SNR scalability. Assuming that the temporal scalability takes priority over the SNR scalability, they put the base layer sub-streams in one group and the rest of the sub-streams in the other one and use FEC with ULP at each group. The first x packets from the first group and y packets from the second group are gathered in description one and the rest in description two. In [13] the authors combine SVC with MDC for video multicasting over P2P networks. Their proposed method uses one base layer and two enhancement layers for SVC. They use FEC with ULP to assign a higher priority to the base layer. The main disadvantage of the FEC-based methods is the overhead added by the insertion of error correction codes.

– Discrete Wavelet Transform (DWT)-based video coding methods are convenient for applying multiple description coding. In the most basic method, wavelet coefficients are partitioned into maximally separated sets, and packetized so that simple error concealment methods can produce good estimates of the lost data [3, 7, 20, 21, 29]. More efficient methods utilize MCTF which is aimed at removing the temporal redundancies of video sequences. In [4] MDC-SVC based on MCTF and 2D DWT is used for video streaming over P2P networks. The receiving peer can measure the channel conditions such as the packet loss rate and bandwidth of each sending peer's path in each GOP period and then calculates the optimal encoder parameters for that GOP through a post-encoding procedure. The resultant encoding parameters are sent to the sending peers through the feedback control channels. Also in [2] an adaptive P2P video streaming system with a flexible multiple description coding (F-MDC) framework is proposed, so that the number of base and enhancement descriptions, and the rate and redundancy level of each description can be adapted. They combine their F-MDC framework with SVC by using JPEG2000 based T+2D DWT which lets them truncate each code-block at any point of bit-plane codes.

– If a video signal f is defined over a domain D, then the domain can be expressed as a collection of sub-domains {S1;..;Sn} where the union of these sub-domains is a cover of domain D. Besides, a corrupt sample can be replaced by an estimated value using the correlation between the neighboring signal samples. Therefore, the sub-domains should be designed in a way that the correlation between the samples is preserved. Domain-based multiple description schemes are based on partitioning the signal domain. Each partition, which is a sub-sampled version of the signal, defines a description. Chang and Sang [5] utilize the even-odd splitting of the coded speech samples. For images, Tillo et al. [20] propose splitting the image into four sub-sampled versions prior to JPEG encoding. There, domain partitioning is performed first, followed by discrete cosine transform, quantization and entropy coding. The main challenge in domain-based multiple description methods is designing sub-domains so that the minimum distance between values inside a domain (inter-domain distance) is maximized while preserving the auto-correlation of the signal.

Our proposed method falls into the group of multi-layer MDC schemes. We have proposed a transform to minimize the base layer size which is the main source of redundancy in these schemes. The proposed method allows us to split the video data into two descriptions although the method can be extended to 4, 8, and more descriptions by repeatedly applying the transform on the data.

3 Our proposed method

Our proposed method involves splitting video into two descriptions each representing video in a lower quality. In our previous work [6], we split each frame of the video spatially into four descriptions. In case of loss or damage in one of the descriptions we estimated the missing data from the remaining descriptions. The data belonging to one of the descriptions is not repeated in other descriptions and the redundancy introduced was in the form of inefficiency in motion compensation. We used the correlation between the adjacent pixels to estimate the missing data. Expressing SNR scalability however, is not feasible using the same method. The bits representing a pixel value do not show any correlation with each other. In SNR scalable video coding techniques, the video is split into two or more layers where the first layer, called the base layer, includes the most essential information and the remaining layers, called enhancement layers, improve the base layer data. The main drawback of these techniques is that the enhancement layers cannot be used whenever the base layer is damaged or lost. This means that when SNR scalability techniques are combined with MDC methods the base layer should be repeated in all descriptions which introduces a large redundancy and cause a decrease in bit rate efficiency. A second problem is that the importance levels of the enhancement layers are not the same. This characteristic arises from the fact that bits at different positions convey different values. Hence, descriptions with equal importance cannot be defined by simply distributing the bits between the descriptions. The solution proposed here defines the base layer in a way that each bit has a contribution in it. Figure 2 depicts the block diagram of our proposed method. The left side blocks refer to the MCTF video encoding where ME/MC indicate motion estimation/motion compensation, DCT is the discrete cosine transform, and Q refers to the quantization step. The right side blocks show the MDC coder proposed in our paper, and variable length coder (VLC). The output of our transform is sent to variable length coder where the descriptions are created. The process indicated by Description Encoder in the block diagram of Fig. 2 gets as input the quantized coefficients of the cosine transform and splits them into two descriptions. We propose a transform $\tau(.)$ to create the descriptions as specified in (1) where A is the original data, and B_1 and B_2 are the data transmitted in each description.

$$\tau(A) = [B_1, B_2] \tag{1}$$

The inverse of this transform reconstructs the original data value as indicated by (2).

$$\tau^{-1}([B_1, B_2]) = A \tag{2}$$

Fig. 2 Block-diagram of the proposed method
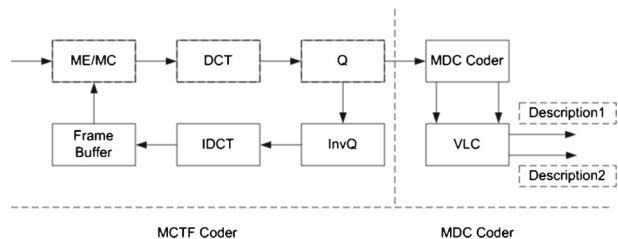
Hence in case of a damage or loss in one these descriptions, we should be able to reconstruct the original value partially as expressed in (3).

$$\tau^{-1}([B_1, null]) = A' \text{ where } \quad |A - A'| < \epsilon$$
$$\tau^{-1}([null, B_2]) = A'' \text{ where } \quad |A - A''| < \epsilon \tag{3}$$

The error threshold value ϵ is determined by a tradeoff between efficiency and accuracy as described below. The proposed transform creates a base layer and an enhancement layer parts for each description. The base layer is repeated in both descriptions and hence introduces a redundancy to the coding. Each description D_i therefore can be given as:

$$D_i = a \bigoplus b_i \qquad for \quad i = 1, 2$$

where a is the base layer, b is the enhancement layer, and $\bigoplus$ is the operation of combining data from these layers. It should be noted that the reconstruction error rate in presence of damage or loss in one of the descriptions depends on the amount of information present in the enhancement layers. Hence a smaller enhancement layer tends to increase the accuracy. On the other hand, a smaller enhancement layer results in a large base layer which will increase the data redundancy. We have considered the following metrics in designing our transform:

– To minimize the redundancy, the base layer size should be minimized,
– Reconstruction error using the base layer only, should be minimized,
– The enhancement layer data size should be a function of the transmitted value.

The last item in the list above is the result of the observation that most of the quantized values are small numbers. Since enhancement layer data in split between the descriptions, reconstruction with one description only results in a large error when the base layer is small. Hence, we prefer an adaptive enhancement layer which grows with increasing data values. The above-mentioned metrics can be expressed mathematically as shown in (4) and (5):

$$Min(A - \tau^{-1}(\tau_B(A)) + \tau_B(A)) \tag{4}$$

$$Min(A - \tau^{-1}(\tau_B(A) \bigoplus \tau_{Ei}(A))) \qquad for \quad i = 1, 2 \tag{5}$$

where $\tau(.)$ is the intended transform, $\tau_B(.)$ is the base layer after the transform, $\tau_{Ei}(.)$ is the enhancement layer *i* after applying the transform, and $\tau^{-1}(.)$ is the inverse transform. Figures 3 and 4 depict the reconstruction error using the base layer only, and the reconstruction error using one description only, for an inverse quadratic and logarithmic functions respectively. We have used the proposed method with inverse-quadratic and logarithmic functions as transforms. Then we reconstructed the encoded value using different cases when one or both descriptions are received. The figures serve to verify the effectiveness of the proposed method in terms of the generated error.

In designing the transform we considered the issue of minimizing the reconstruction error for all cases of reconstruction using one description only, reconstruction using base layer only. The last case arises when the channels used for transmission of the descriptions suffer from the limited bandwidth problem and a down-scaled

Fig. 3 Reconstruction error
with inverse-quadratic
transform
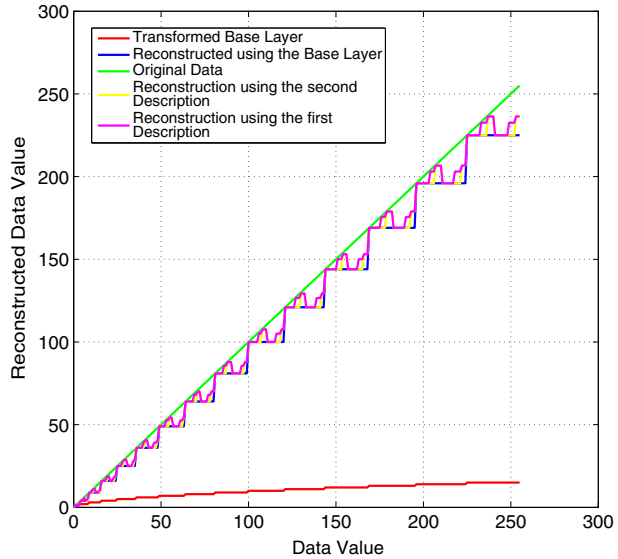
stream is received through each channel. The optimum solution considering the
above-mentioned criteria is an inverse quadratic transform as given in (6):

$$Base = Trunc(\sqrt{Coef}) \qquad\qquad Enhancement = \sqrt{Coef} - Base \qquad (6)$$

where Coef is the quantized DCT coefficients of the macro-blocks Fig. 2. The fraction
part after applying the transform is divided into two parts and used as enhancements
to the base layer data. The enhancement layer bits are coded separately. This
feature provides the multi-layer scalability characteristic for each description. The

Fig. 4 Reconstruction error
with logarithmic transform
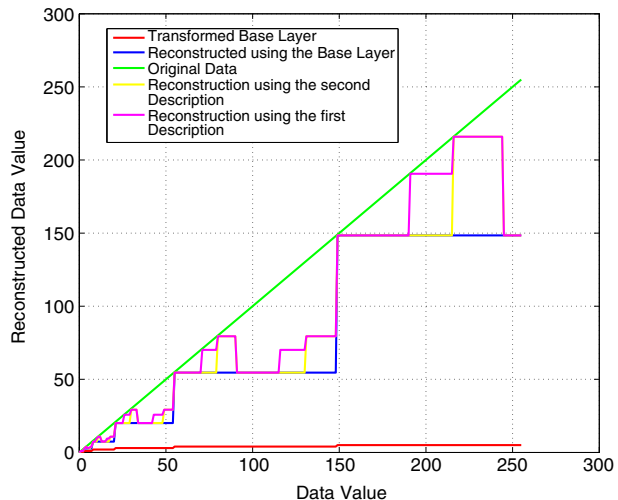
descriptions go through entropy coding later on, so that each layer present in the descriptions is entropy coded separately.

In the following discussions we labeled the descriptions as D1 and D2. The fraction bits at position 2^{-1} and 2^{-4} are packed and entropy coded at the enhancement layers of D1 and the fraction bits at position 2^{-2} and 2^{-3} are packed and entropy coded at the enhancement layers of D2, respectively. In this way, we tried to balance the bit rate and accuracy of both descriptions. Algorithm 1 describes the reconstruction of the video when both descriptions are received or in case of failure in one of the descriptions. The proposed method provides the possibility of ignoring one or both enhancement layers data in each description in case of communication bandwidth restrictions. This scalability feature when the reconstruction is carried out using the base layer only, has not been considered in Algorithm 1.

Algorithm 1 Reconstructing video

---

if $Base\,D1 = NULL$ **then**
   {If Description 1 is lost}
   $fraction \leftarrow Enhance\,D2_1 \times (2^{-2}) + Enhance\,D2_2 \times (2^{-3})$
   $Coef \leftarrow Round((Base\,D2 + fraction)^2)$
else if $Base\,D2 = NULL$ **then**
   {If Description 2 is lost}
   $fraction \leftarrow Enhance\,D1_1 \times (2^{-1}) + Enhance\,D1_2 \times (2^{-4})$
   $Coef \leftarrow Round((Base\,D1 + fraction)^2)$
else
   {Both Descriptions are received}
   $fraction_1 \leftarrow Enhance\,D1_1 \times (2^{-1}) + Enhance\,D1_2 \times (2^{-4})$
   $fraction_2 \leftarrow Enhance\,D2_1 \times (2^{-2}) + Enhance\,D2_2 \times (2^{-3})$
   $fraction \leftarrow fraction_1 + fraction_2$
   $Coef \leftarrow Round((Base\,D2 + fraction)^2)$
end if

---

4 Experimental results

For evaluating the performance of our proposed method, we have considered measuring Peak Signal to Noise Ratio of the Y component of YcbCr color space from the macro-blocks (Y-PSNR). Equations (7) and (8) describe PSNR used in our implementation mathematically.

$$PSNR = 20 \log_{10} \frac{\text{Max}_I}{\sqrt{MSE}} \qquad (7)$$

$$MSE = \frac{1}{mn} \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} ||I(i,\,j) - I'(i,\,j)||^2 \qquad (8)$$

where MSE is the mean square error, Max_I indicates the largest possible pixel value, I is the original frame, I' is the decoded frame at the receiver side, and m and n are number of rows and columns respectively. Y-PSNR is applied to all frames of video segments listed in Table 1 by comparing the corresponding frames of the original video segment and retrieved video using one or both descriptions from our

**Table 1** Average Y-PSNR values when loss is in only one frame of each GOP.

Sequence name	Resolution	Frame rate	Avg. Y-PSNR using D1 & D2 (dB)	Avg. Y-PSNR using D1 (dB)	Avg. Y-PSNR using D2 (dB)
Foreman	352 × 288	30	36.474	34.192	33.966
Stefan & Martin	768 × 576	30	34.078	32.472	32.105
City	704 × 576	60	34.643	32.671	31.978

proposed coding method. We place 32 frames in each GOP and a diadic hierarchical temporal structure has been used for motion compensated coding. Furthermore, we have imposed the same reference frame for all macro-blocks of a frame for simplicity although H.264 supports utilizing different reference frame for macro-blocks of a frame. As the proposed method has both error resilience characteristic through implementing multiple description coding, and scalable video coding, we have considered the following test scenarios.

– Measuring redundancy imposed by error resilience of MDC,
– Performance measurement when only the base layer is received,

Fig. 5 Retrieved frames using proposed method: *Upper-left* original frame, *upper-right* retrieved using both descriptions, *lower-left* retrieved using first description, *lower-right* retrieved using second description
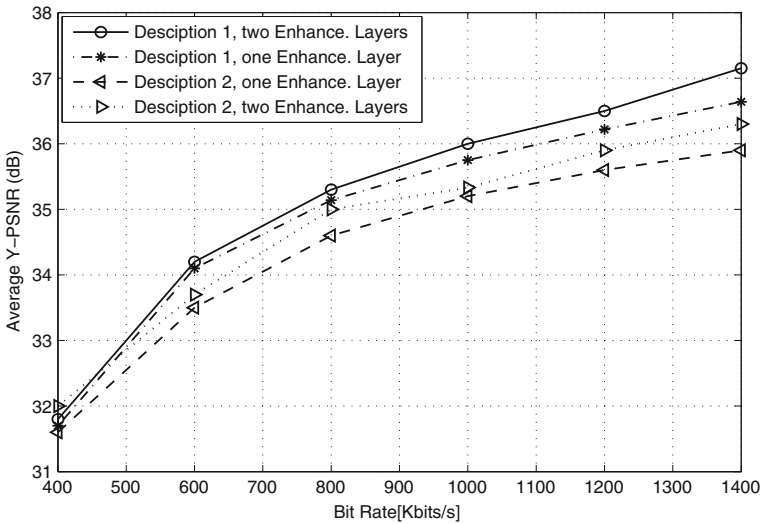
Fig. 6 Comparison of the rate distortion when one description is received

- Performance measurement when only one enhancement layer from each description is received,
- Performance measurement when only one description is received,
- Performance measurement when one description with one enhancement layer is received.
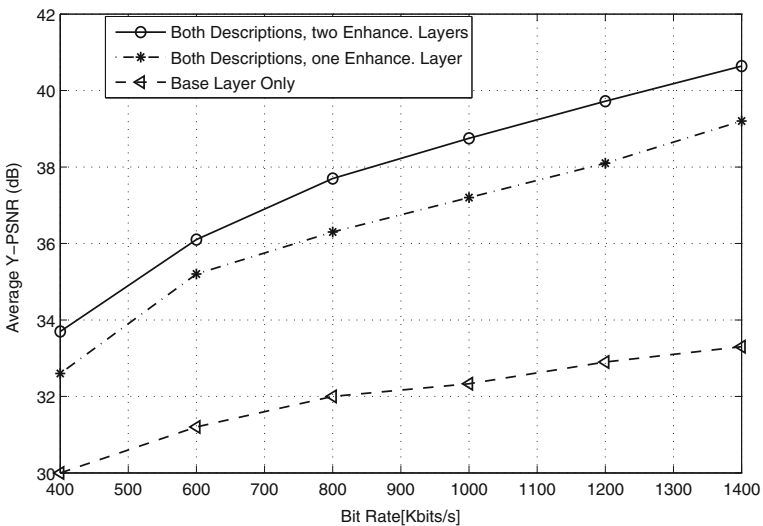
Fig. 7 Comparison of the rate distortion when both descriptions are received

Table 2 Redundancy added by the proposed method

Sequence name	Target bit rate (Kbits/s)	Bit rate using the proposed method (Kbits/s)	Redundancy percentage (%)
Foreman	1,000	1,292	29.2
Stefan & Martin	1,000	1,372	37.2
City	1,000	1,341	34.1

Figure 5 depicts a frame of the first test video and its corresponding reconstructions using one description, and using both descriptions. The visual inspection of the retrieved frames also indicates that the proposed method provides acceptable results even in presence of transmission error. The redundancy caused by repeating base layer information in both descriptions is partially compensated by organizing the data bits as mentioned in Section 3. To optimize the distortion with respect to the bit rate, the enhancement layer in each description has been entropy coded separately. The optimization is related to the observation that a large number of coefficients after quantization are small integers. Hence, entropy coding encodes the base layer more efficiently after applying the transform. Since the redundancy is arising from the repetition of the base layer in both descriptions, the total performance improves. Meanwhile, this feature provides the flexibility of having scalability at each description.

In our testing scenario, we have considered transmission over a packet loss network. The bitstreams of the two descriptions are separated in packets of maximal size of 1,500 bytes for compatibility with the maximum frame size of Ethernet. For each description, separate packets are created. If the packet is lost, we consider that the corresponding description is not available for reconstructing the block and hence, the block is reconstructed using the other description. In the second scenario we assume that as a result of bandwidth fluctuations, the receiver can receive the data in a description partially. This means that the enhancement layer of the data in a description is dropped. Figure 6 compares the rate distortion of the proposed method when one or two enhancement layers are received within a single description. With

Fig. 8 Comparison of H.264 performance with the proposed method when both descriptions are received
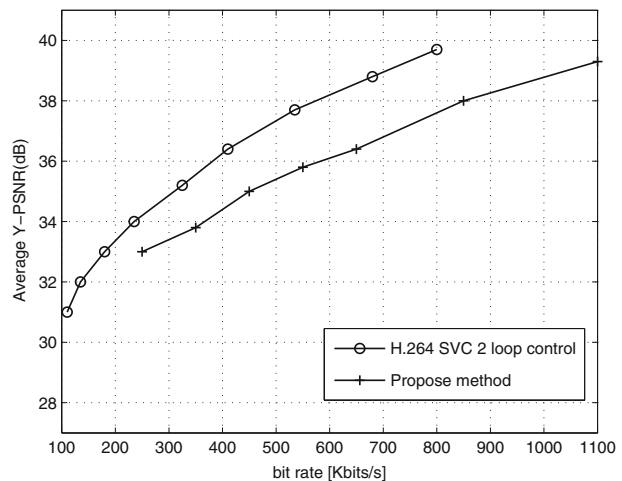
Table 3 Performance comparison of the proposed method with MDTC in terms of Y-PSNR (dB)

Proposed method	MDTC	Loss (%)
29.11	29.03	10
27.87	28.53	20

only one description received, the video quality is still acceptable with an average PSNR reduction of less than 2 dB. Figure 7 depicts the case when both descriptions are received in down-scaled form. The extreme case of receiving base layer only is computed by considering a duplication in data. Table 2 shows the average bit rates and the related data inflation percentage due to the redundancy added by the proposed method. The higher rate of redundancy in Stefan and City sequences can be related to their higher spatial detail and amount of movements.

A comparison with SNR scalability of H.264 standard has been given in Figure 8. The video sequence 'City' has been used for comparison in CIF spatial resolution, at a temporal rate of 15 fps, and 16 frames in each GOP. The coarse grain quality scalable mode with three layers has been utilized. The multi-layer structure of H.264 encoder allows it to minimize redundancy which is present in our proposed method which is optimized for a noisy channel. Meanwhile, it should be noted that the results presented in Figure 8 for H.264 standard are obtained from a single description while our proposed method is tested with two descriptions which imposes a redundancy of 34.1% in 'City' video sequence. We have also compared our method with the multiple description transform method (MDTC) method proposed in [18] which compresses the video using SNR scalability, duplicates the base layer so that it appears in both descriptions, and alternates blocks (i.e., GOBs) of the enhancement layer between the two descriptions and hence has a similarity with our proposed method. For comparison we use 'Foreman' QCIF video sequence with 144 Kbps and 7.5 fps. The comparison is for 10% and 20% of frame losses. The results of the comparison are given in Table 3. Despite having almost similar PSNR performance, it is worth noting that in 144 Kbps the redundancy imposed by our proposed method is 41.2% whereas the redundancy rate is 45% in the method proposed in [18].

5 Conclusion

A new method for handling the data loss during the transmission of video streams has been proposed. Our proposed method is based on multiple description coding combined with signal to noise ratio (SNR) scalable video coding and hence it has the capability of being used as a scalable coding method where any data loss or corruption is reflected as reduction in the quality of the video. The multi-layer structure of data in each description provides the feasibility of reducing data rate by scaling down the video whenever the connection suffers from a low bandwidth problem. In order to measure the performance of the proposed coding method, distortion rate imposed by data loss and scaling down for rate efficiency, have been utilized. Except for the case when all descriptions are lost, the video streams do not experience a major quality loss at play back. Utilizing the motion compensated temporal filtering structure of video coding standards, we managed to preserve the compatibility of the proposed method with major standards such as H.264. Our

proposed method is based on SNR scalability of video coding standards, however, a reasonable extension of the work is going to be its combination with temporal and spatial scalabilities.

References

 1. Abanoz TB, Tekalp AM (2009) SVC-based scalable multiple description video coding and optimization of encoding configuration. Signal Process Image Commun 24(9):691–701
 2. Akyol E, Tekalp AM, Civanlar MR (2007) A flexible multiple description coding framework for adaptive peer-to-peer video streaming. IEEE Select Topics Signal Proc 1(2):231–245
 3. Andreopoulos Y, van der Schaar M, Munteanu A, Barbarien J, Schelkens P, Cornelis J (2003) Fully-scalable Wavelet Video Coding using in-band Motion-compensated Temporal Filtering. In: Proc. IEEE International Conference on Acoustics, Speech and Signal Processing, pp 417–420
 4. Ardestani MR, Shirazi AAB, Hashemi MR (2011) Low-complexity unbalanced multiple description coding based on balanced clusters for adaptive peer-to-peer video streaming. Signal Process Image Commun 26(3):143–161
 5. Chang SK, Sang L (2001) Multiple description coding of motion fields for robust video transmission. IEEE Trans Circuits Syst Video Technol 11(9):999–1010
 6. Choupani R, Wong JSM, Tolun MR (2009) Multiple description scalable coding for video transmission over unreliable networks. In: Embedded computer systems: architectures, modeling, and simulation. 9th international workshop, SAMOS 2009. Samos, Greece, pp 58–67
 7. Choupani R, Wong JSM, Tolun MR (2011) Scalable video transmission over unreliable networks using multiple description wavelet coding. In: The 7th International Conference on Digital Content, Multimedia Technology and its Application (IDCTA2011). Busan, Korea, pp 5–10
 8. Conklin G, Greenbaum G, Lillevold K, Lippman A, Reznik Y (2001) Video coding for streaming media delivery on the internet. IEEE Trans Circuits Syst Video Technol 11:269–281
 9. Folli M, Favalli L (2008) Scalable multiple description coding of video sequences. GTTI'08—Sessione elaborazione dei segnali
10. Franchi N, Fumagalli M, Lancini R, Tubaro S (2003) A space domain approach for multiple description video coding. In: ICIP 2003, vol 2, 253–256
11. Girod B (1987) The efficiency of motion-compensated prediction for hybrid video coding of video sequences. IEEE J Sel Areas Commun 5(7):1140–1154
12. Goyal VK (2001) Multiple description coding: compression meets the network. IEEE Signal Process Mag 18(5):74–93
13. López-Fuentes FA (2011) P2P video streaming combining SVC and MDC. Int J Appl Math Comput Sci 21(2):295–306
14. Mohr AE, Riskin EA, Ladner RE (2000) Unequal loss protection: graceful degradation of image quality over packet erasure channels through forward error correction. IEEE J Sel Areas Commun 18(6):819–828
15. Ohm J (2005) Advances in scalable video coding. Proc IEEE 93(1):42–56
16. Puri R, Ramchandran K (1999) Multiple description source coding using forward error correction codes. IEEE Conf Signal Syst Comp 1:342–346
17. Reguant VD, Prats FE, De Pozuelo RM, Margalef FP, Ublergo GF (2008) Delivery of H264 SVC/MDC streams over Wimax and DVB-T networks. In: Proc. Int. Symp. Consumer Electronics, ISCE, pp 1–4
18. Reibman AR, Jafarkhani H, Wang Y, Orchard MT, Puri R (2002) Multiple-description video coding using motion-compensated temporal prediction. IEEE Trans Circuits Syst Video Technol 12(3):193–204
19. Schwarz H, Marpe D, Wiegand T (2007) Overview of the scalable video coding extension of the H.264/AVC standard. IEEE Trans Circuits Syst Video Technol 17(9):1103–1120
20. Tillo T, Olmo G (2004) A novel multiple description coding scheme compatible with the JPEG2000 decoder. IEEE Signal Process Lett 11(11):908–911

21. Tillo T, Grangetto M, Olmo G (2007) Multiple description image coding based on Lagrangian rate allocation. IEEE Trans Image Process 16(3):673–683
22. Venkataramani R, Kramer G, Goyal VK (2003) Multiple description coding with many channels. IEEE Trans Inf Theory 49(9):2106–2114
23. Wang Y, Lin S (2002) Error-resilient video coding using multiple description motion compensation. IEEE Trans Circuits Syst Video Technol 12(6):438–452
24. Wang Y, Orchard MT, Vaishampayan V, Reibman AR (2001) Multiple description coding using pairwise correlating transforms. IEEE Trans Image Process 10:351–366
25. Wang Y, Reibman AR, Shunan L (2005) Multiple description coding for video delivery. Proc IEEE 93(1):57–70
26. Wien M, Schwarz H, Oelbaum T (2007) Performance Analysis of SVC. IEEE Trans Circuits Syst Video Technol 17(9):1194–1203
27. Wu D, Hou Y, Zhu W, Zhang YQ, Peha J (2001) Streaming video over the internet: approaches and directions. IEEE Trans Circuits Syst Video Technol 11(3):282–300
28. Xiang W, Zhu C, Siew CK, Xu Y, Liu M (2009) Forward error correction-based 2-D layered multiple description coding for error-resilient H.264 SVC video transmission. IEEE Trans Circuits Syst Video Technol 19(12):1730–1738
29. Xuguang Y, Ramchandran K (2000) Optimal subband filter banks for multiple description coding. IEEE Trans Inf Theory 46(7):2477–2490

Roya Choupani received her BSc. and MSc. degrees in computer engineering in 1998 and 2002 respectively. Since 2006 she has started her PhD study at the Computer Engineering department of Delft University of Technology, in Delft—the Netherlands. Her research interests are video coding, data communications and networking.

Stephan Wong was born in Paramaribo, Suriname on October 20th, 1973. He obtained his PhD from the Delft University of Technology in December 2002 after which I started as an assistant professor at the same university. His PhD thesis entitled "Microcoded Reconfigurable Embedded Processor" describes the MOLEN polymorphic processor, organization, and (micro-)architecture. His research interests include: Reconfigurable Computing, Distributed Collaborative Computing, High-Performance Computing, Embedded Systems, Hardware/Software Co-Design, Network Processing.

Mehmet Tolun main research interests are: Artificial Intelligence and Expert Systems, Machine Learning Algorithms and Symbolic Inductive Learning. His Inductive Learning Algorithm-ILA-was the subject of many MS theses and two Ph.D. dissertations. Mehmet Tolun was past Dean of the EMU Faculty of Engineering for nearly 4 years. During his term of office two of the engineering programs at EMU have been applied for ABET "substantial equivalency". He contributed to IEEE as the IEEE Turkey Section Chair, Computer Society Chapter Chair, and GOLD Affinity Group Chair. He is the recipient of the IEEE Third Millennium Medal. He is currently a Senior Member of the IEEE.