Optimal filtering of polyphase-downsampling-based multiple description coded video

Çağlar Ateş
Oğuzhan Urban
Sarp Ertürk
Kocaeli University
Umutepe Campus
Department of Electronics and Telecommunications Engineering
Laboratory of Image and Signal Processing (KULIS)
41380, Kocaeli, Turkey
E-mail: urhano@kocaeli.edu.tr

Abstract. We combine an optimal filtering approach with a multiple description coding (MDC) approach to enhance the quality of video at the receiver. Optimal filter coefficients are obtained directly from the bit stream and employed at the receiver. Optimal filter coefficients are computed at the encoder and added to each description. At the decoder, optimal filter coefficients are obtained directly from the bit stream and employed to improve the quality. Experimental results show that the proposed approach enables better MDC performance for video in case of packet losses. © 2009 SPIE and IS&T.

1 Introduction

Transmission of video over packet networks requires efficient compression due to limited bandwidth available to parties. Layered coding (LC) and multiple description coding (MDC) are two important source coding approaches for transmission of video over error-prone networks. These coding approaches produce more than one data stream whereas the conventional video coding approach outputs only a single data stream. In MDC, multiple equally important data streams, which are called descriptions, are generated. Since each description is self-decodable; there is no strict requirement for any particular description. Lower quality data is reconstructed if some descriptions are lost during transmission. Additional descriptions improve the reconstruction performance of MDC in general. Detailed discussion on LC and MDC can be found in Ref. 1.

As described in Ref. 1; it is not possible to use automatic repeat request (ARQ) for delay-sensitive real-time applications, such as interactive multimedia and video conferencing. In the case of packet loss or channel failure, conventional coding approaches may cause important quality loss. On the other hand, MDC provides reasonable performance even if some of the descriptions are lost during the transmission. If all descriptions are received, a higher quality reconstruction of the data is obtained. There are three possible situations for the decoder in a simple two-description case. If only the first or second description is received, the decoder is referred to as a side decoder. The decoder is called a central decoder when all descriptions are available to the decoder. See Ref. 2 for detailed information about MDC for speech, image, and video coding.

Although it is relatively easy to utilize MDC for speech and image coding, the situation is more complicated for video because a prediction error, called mismatch or drift, might occur at the decoder. In a typical video coder, temporal prediction based on motion estimation is carried out to exploit the temporal correlation existing in consecutive video frames to effectively compress the video. If MDC is utilized, however, three different prediction errors are possible when two descriptions are generated. At any time, if the prediction error at the encoder and decoder is not the same due to the utilization of different descriptions, it is not possible to reconstruct the video properly due to the mismatch of predictions, until an intraencoded frame is received. An easy way to eliminate the mismatch is to use two separate prediction loops, as in Ref. 3. In this case, however, the coding performance may be degraded and an additional operation may be required to combine side descriptions to obtain an improved quality video in the case when two descriptions are received. In Ref. 3, a coarse quantizer is utilized in the prediction loop, to make the predictions of the two side encoders similar to each other, so as to remove any possible mismatch. An information theoretic approach was presented in Ref. 4 to alleviate the performance degradation in Ref. 3. The quantizer in the prediction loop of Refs. 3 and 4 was removed in Ref. 5, and it was maintained that the central decoder prevents the mismatch in this way. Another way to deal with drift is to allow some drift, but keep it under control, as presented in Ref. 6, where three different strategies were proposed for this purpose.

A polyphase-downsampling (PD)-based MDC approach that employs 2-D oversampling for video transmission was presented in Ref. 7. Two different approaches are considered to deal with the mismatch in this method. The first one is to utilize the commonly used three-prediction-loop approach presented in Ref. 3 for PD-based MDC. The second approach is to create descriptions before the coding loop so as to prevent the mismatch. A 1-D oversampling approach was presented in Ref. 8 for efficient MDC of images to
improve the performance of the method described in Ref. 7. A detailed overview of MDC of video is given in Ref. 9.

Optimal filtering was utilized in Ref. 10 for improved compression of JPEG-coded images. The idea behind this approach is to improve the encoded image by simply filtering it with an optimal filter that is obtained in a least-squares sense using the original and encoded images. This approach is extended to standard video coding (without MDC) in Ref. 11 using 2-D optimal filters. A faster implementation of optimal filtering for real-time video coding applications (again without MDC) was later on presented in Ref. 12. The optimal filtering approach was then applied to MDC of still images.

In this paper, we extend our optimal filtering approach to MDC of video. Instead of using an error drift control approach, we created descriptions before the coding loop, as in Ref. 7, to eliminate drift. The optimal filtering approach is considered as a postprocessing stage and combined with the 1-D oversampling approach presented in Ref. 8 to improve the performance of MDC of video.

2 Optimal Filtering for MDC of Video

Figure 1 shows the proposed multiple description approach that incorporates the optimal filtering strategy proposed in this paper into the PD-based MDC approaches utilized in Refs. 7 and 8. As seen from this figure, the original input data was initially oversampled to introduce controlled redundancy to the descriptions. This oversampling was performed in the discrete cosine transform (DCT) domain. In Ref. 7, zero padding was applied to both directions after the DCT of the image was taken. Next, are inverse DCT was carried out to return to the image domain. On the other hand in Ref. 8, zero padding was carried out in only one direction, which was determined according to the effectiveness of the zero padding direction. After zero-padding-based oversampling of the data, multiple descriptions were generated by simply downsampling the oversampled data, as in Ref. 13. Finally, all descriptions were compressed using separate encoders and transmitted to the receiver. At the receiver, the initial decoding was carried out to obtain available descriptions. Next, the received descriptions were combined to create output data.

In this paper, PD for MDC of video is utilized, using 1-D oversampling, which was demonstrated to improve the performance of MDC image coding. After multiple descriptions are generated by simply downsampling video frames, these are encoded separately. After computing optimal filter coefficients of the descriptions at the encoder, these coefficients are multiplexed into the data stream of each description (see Fig. 1). At the decoder, simple linear filtering, using these filter coefficients extracted from the data stream, is performed to obtain higher quality representations of the encoded video in case of packet loss.

Before giving any detail of optimal filtering for MDC of video, we describe the optimal filtering strategy for conventional video coding, so that the explanation of this concept for MDC will be easier. The optimal filtering approach used

![Fig. 1 Implementation of PD-based MDC.](image1)

![Fig. 2 MDC results for different test sequences and several methods: (a) “Foreman” sequence and (d) “Coastguard” sequence.](image2)
in Refs. 10–14 aims to improve visual quality of images or video frames in the form of a postprocessing operation. The coefficients of the optimal filter are computed at the encoder by finding the filter kernel that minimizes the difference between encoded (reconstructed) and original video frames in least squares sense. These coefficients are multiplexed into the bit stream so that the decoder simply performs linear filtering using these coefficients to obtain a higher quality version of the encoded data.

If \( \hat{I} \) shows the encoded video frame, then the optimal filtered version of it can be computed as

\[
\hat{I} = \hat{I} * G,
\]

where \( G \) is a 2-D kernel of size \( l \times l \), \( \hat{I} \) represents the filtered final video frame, and \( * \) denotes the convolution operation. Coefficients of the optimal filter can be computed as

\[
\min_{G} \| I - \hat{I} \|^2 = \min_{G} \| I - \hat{I} * G \|^2, \tag{2}
\]

where \( I \) denotes the original video frame, and \( \| \cdot \|^2 \) is used to represent a least-squares solution. Note that an unrolled implementation of the iterative preconjugated gradients is utilized due to its potential of solving large systems and providing quick convergence to obtain \( G \), as described in Ref. 11. The kernel size is chosen as \( 5 \times 5 \), empirically, based on our past experience.

Now, let us consider only the two-descriptions case in MDC for simplicity. In this case, the original video frame \( I \) is initially oversampled \( (I_{os}) \) with an oversampling ratio \( v \). Next, this oversampled video frame is partitioned into two descriptions \( I_{os}^1 \) and \( I_{os}^2 \) using alternative downsampling patterns. These descriptions are fed into the video encoders and encoded descriptions, i.e., \( \hat{I}_{os}^1 \) and \( \hat{I}_{os}^2 \), are obtained. The approach proposed in this paper aims to utilize a single-stage optimal filtering MDC of video, which was presented for still image coding. In this approach, two optimal filters are computed for each description. The computations for the first and second descriptions are given as

\[
\min_{G_{1,1}} \| I_{os}^1 - \hat{I}_{os}^1 * G_{1,1} \|^2, \quad \hat{I}_{os}^1 = \hat{I}_{os} * G_{1,1}, \tag{3}
\]

\[
\min_{G_{1,2}} \| I_{os}^1 - \hat{I}_{os}^1 * G_{1,2} \|^2, \quad \hat{I}_{os}^1 = \hat{I}_{os} * G_{1,2}, \tag{3}
\]

\[
\min_{G_{2,1}} \| I_{os}^2 - \hat{I}_{os}^2 * G_{2,1} \|^2, \quad \hat{I}_{os}^2 = \hat{I}_{os} * G_{2,1}, \tag{3}
\]

\[
\min_{G_{2,2}} \| I_{os}^2 - \hat{I}_{os}^2 * G_{2,2} \|^2, \quad \hat{I}_{os}^2 = \hat{I}_{os} * G_{2,2}. \tag{3}
\]

The coefficients of the two optimal filters that are obtained from Eqs. (3) and (4) must be multiplexed into the bit streams for each description. At the decoder, each description and the corresponding optimal filter coefficients are utilized to obtain an improved version of the description itself and an estimation of the other description. Therefore, if one of the descriptions is lost, the remaining description can be used to estimate the lost one using the corresponding optimal filter. After each received description is reconstructed properly the descriptions are merged and the output video is obtained at the decoder.

3 Experimental Results

The experiments are carried out using common intermediate format (CIF) sized (352 × 288) “Foreman” and “Coastguard” test sequences of length 300 frames. The oversampling parameter is set to \( v = 0.25 \), since this value is obtained to balance the performance for all bit rates. Only the two-description case is considered in the experimental results because usually this is the case used in literature. We employed-the state of the art video encoding standard H.264/AVC (advanced video coding) to evaluate the performance of the proposed optimal filtering approach. JM version 10.1 of the reference software was employed for tests.

Figure 2 shows rate-distortion curves for the “Foreman” and “Coastguard” sequences. Note that in the case of channel failure, all data in this channel are assumed to be lost. The 1D OS and 2D OS labels in the figure denote the oversampling based approaches presented in Refs. 8 and 7, respectively, and these do not contain the proposed optimal filtering strategy. These approaches are applied to video coding in this paper with some suitable modifications for comparison purposes. In Ref. 7, after 2-D oversampling, there are two possible downsampling directions, i.e., horizontal and vertical. Based on our experiments, horizontal downsampling is chosen in the results presented in this paper because of its better performance. In Ref. 8, oversampling and downsampling can be performed in both directions, resulting in four possible scenarios. Experiments have shown that zero padding in the vertical direction followed by horizontal downsampling gives better results for the 1-D oversampling approach, and therefore results are presented for this case only. Note that linear spatial estimation was used to reconstruct the lost descriptions for 1-D OS and 2-D OS.

The approach proposed in this paper utilizes the optimal filtering concept to reconstruct lost descriptions. It is important to note that the filter coefficient overhead is relatively small and filter coefficients are represented with 16 bits each without compression and multiplexed into the bit-stream, which results in a \( 5 \times 5 \times 16 \) bit/s overhead, which is included in the results given in Fig. 2. It is furthermore possible to compress the optimal filter coefficients, as described in Ref. 12, but the overhead is still small compared to the other video data. As described in Ref. 14, computation of optimal filter coefficients can be performed in real time and further improvements are possible.

When we evaluate the results shown in Fig. 2, we clearly see that the 1-D oversampling approach of Ref. 8 provides better results compared to the 2-D oversampling method of Ref. 7 for all cases. The proposed approach, on the other hand, improves the performance of the 1-D-oversampling-based approach significantly in the case of packet losses when only one description is received. If there is no packet loss and all descriptions are received, the proposed approach provides results similar to those in Ref. 8. Note that average values are given for the case if only one description is received.
Since the approach proposed in this paper can be considered as a postprocessing approach to Ref. 8, its additional computational load should be taken into consideration. In Ref. 11, it is shown that the computation of optimal filter coefficients takes only about 50 ms/frame on a PC with a 1.7-GHz processor. It is furthermore possible to accelerate this computation by making use of 1-D filters, as in Ref. 12. Thus, the additional computational load introduced by optimal filtering is quite low compared to the computational load of the video encoder.

4 Conclusions
An MDC-based video compression approach was presented. The proposed approach is based on PD of video frames before encoding. Hence, any mismatch error common to MDC of the video is prevented. The approach presented in this paper can be considered as a postprocessing approach and can be applied to other schemes as well. Experimental results show that the proposed optimal filtering approach is capable of significantly improving the performance of a conventional MDC method given in the literature, if descriptions are likely to be lost.

Acknowledgments
This work was supported by the Scientific and Technological Research Council of Turkey (TUBITAK) under Grant No. EEEAG/107E155.

References