MULTIPLE DESCRIPTION VIDEO CODING USING MOTION-COMPENSATED LIFTED 3D WAVELET DECOMPOSITION

Aditya Mavlankar and Eckehard Steinbach

Institute of Communication Networks, Media Technology Group
Technische Universität München
Munich, Germany
aditya_m@mytum.de, eckehard.steinbach@tum.de

ABSTRACT

In this paper we design a multiple description (MD) video coding scheme based on the motion compensated (MC) lifted wavelet transform. A major advantage of basing MD video coding on motion-compensated lifted 3D wavelet decomposition is that it does not require any mismatch control like in a hybrid codec which was previously achieved by sending drift compensation data. We propose to perform the temporal decomposition of a group of pictures and then create multiple descriptions for each temporally transformed frame. We use polyphase sub-sampling with an appropriate amount of oversampling to form the descriptions and to control the redundancy among them. We determine the appropriate amount of redundancy as a function of description loss probability. Our experimental results show that the controlled introduction of redundancy leads to significant improvements in lossy transmission environments.

1. INTRODUCTION

Multiple description (MD) video coding leads to good error robustness, especially in the case when there are multiple paths available to reach a client and the loss events on these paths are independent. MD coding can also prove useful on a single path when a single stream is broken into multiple streams so that a stall in the decoding of one stream does not affect the other streams. The main condition for multiple description coding is that every description should be independently decodable. However, the descriptions should complement each other so that in any scenario, the reception of one more description ensures enhancement of quality. It should be noted that in the case of a high description loss probability there has to be an appropriate amount of redundancy across the descriptions which helps in reconstructing the original spatial and temporal resolution of the video signal.

In the past years a significant number of MD coding schemes for video have been proposed (e.g., [2], [7]-[12]). For an overview paper on MD coding please refer to [13]. Most of the proposals for MD video consider the traditional hybrid video coding structure with motion compensated prediction and DCT as their main building blocks. In comparison, in this work we present a multiple description video coding scheme that is based on a video encoder that uses the recently emerging motion-compensated lifted 3D wavelet transform as its basis [9], [14], [15], [16].

In [2] a MD coding approach for subband/wavelet-coded images and motion compensated 3D subband/wavelet-coded video is presented. The descriptions are formed after performing the 2D spatial wavelet transform so that the spatial encoding works more efficiently compared to the case where sub-sampling is performed on the signal itself. For this the transform coefficients are partitioned such that they are maximally separated from each other. But due to this the conventional zero-tree structures cannot be exploited and a loss in coding efficiency is observed. In [2] there is no discussion of introducing redundancy among the descriptions to ensure reasonable quality in the case of high loss probability. In addition, it is more difficult to interpolate lost descriptions because the wavelet coefficients have less correlation compared to temporally decomposed video data with added redundancy like in this paper. The authors in [3] and [4] propose to embed low quality copies of the transform coefficients across the descriptions so that the loss of a particular description allows us to retrieve at least a low quality version of the content which was lost in the description. These methods are however not very flexible from the design aspect of introducing appropriate amount of extra redundancy. To address this issue our proposal entails a design parameter for inter-description redundancy for which we find the optimal value. This value is matched to

- the description loss probability
- the signal
- the filters used for interpolating the lost descriptions
- the spatial encoding scheme used to encode every description.

This paper is organized as follows. In Section 2, we present our proposed scheme for multiple description video coding. An important aspect of this scheme is to control the amount of redundancy introduced among the individual descriptions. How to optimally introduce this redundancy as a function of loss probability is discussed in detail. Section 3 gives experimental results demonstrating the performance of our approach. Section 4 concludes the paper.

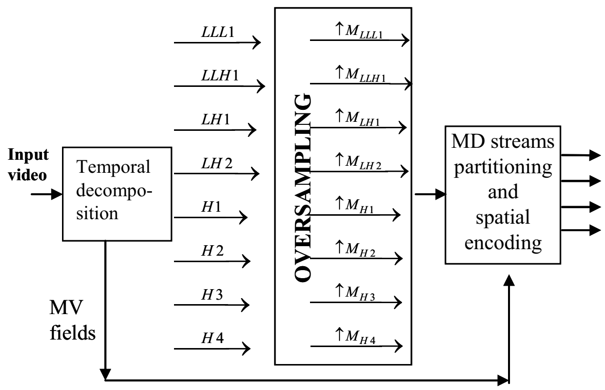

Figure 1: Overview of the proposed MD coding scheme

2. OVERVIEW OF THE SCHEME

The recent years saw the emergence of motion-compensated wavelet transform techniques with the use of lifting steps. These techniques provide a signal decomposition which is invertible and can employ arbitrary MC methods [14]-[16]. Since recursive temporal prediction is replaced by a motion-compensated transform, it also eliminates the dependency quantization framework which is an inherent part of hybrid video codecs. In traditional hybrid codecs quantization is embedded in the recursive prediction loop, whereas in 3-D wavelet codecs quantization and spatial encoding are applied after the temporal decorrelation of a group of pictures (GOP). Figure 1 shows our proposed MD video codec which is based on MC lifted wavelet transform. This scheme extends the Drift Compensation Multiple Description Video Codec (DC MDVC) proposed for a hybrid video codec in [1] to MC-lifted 3-D wavelet coding. A big advantage of employing the 3-D wavelet scheme is that there is no need to send additional drift compensation data. DC MDVC has to send the drift compensation data to compensate for mismatch between the recursive prediction loops at the encoder and the decoder in case of loss. In fact the number of drift compensation streams increases exponentially with the total number of descriptions. It is $(2^N - 2)$, where N is the total number of descriptions. This can be easily understood by considering that a drift compensation stream has to be created for every scenario except when all or no descriptions are received.

In our scheme the temporal decomposition of a group of pictures yields the temporally decomposed frames LLL1, LLH1,…,H4 as shown in Figure 1. Then multiple descriptions are created out of each of these temporally decomposed frames. This is done by first performing the pre-processing proposed in [1] which results in an appropriate amount of oversampling or frame expansion and then forming the descriptions based on spatial sub-sampling. The pre-processing is shown in Figure 2.

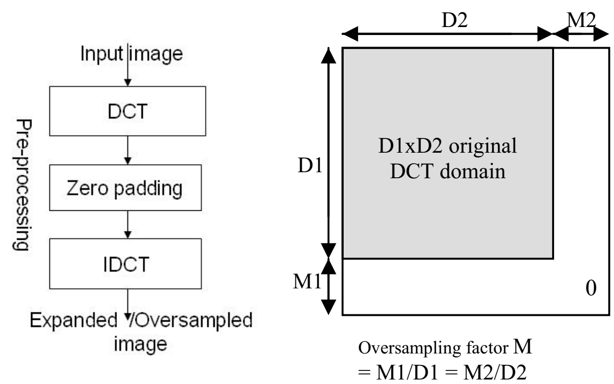

Figure 2: Pre-processing to introduce flexible amount of redundancy across descriptions [1]

2.1. Appropriate amount of redundancy

The encoder has to determine the appropriate amount of oversampling to be performed on each frame. In [1] the effect of the pre-processing in the spectral domain is shown. Aliasing in every individual description is reduced but at the same time there are more samples to be coded and hence the optimum tradeoff has to be found for a given loss probability and total rate of encoding. The result depends primarily on the loss probability and the signal since different signals have a different amount of correlation among neighboring samples. For example, if the loss probability is zero then we can allow a high amount of aliasing in every individual description because all the descriptions will be received and put together in the end anyway. And hence in this case oversampling is not needed in the pre-processing step.

Since the encoder has to determine the oversampling factor for every individual temporally transformed frame, the search would be as follows:

$$\{M_{LLL1}, M_{LLH1}, \cdots, M_{H4}\} = \arg \max_{\{M_{LLL1}, M_{LLH1}, \cdots, M_{H4}\}} \left\{ \sum_{i=1}^{2^N-1} \left[PSNR_{GOP}\left(i, \left\{ \begin{matrix} M_{LLL1}, M_{LLH1}, \\ \cdots, M_{H4} \end{matrix} \right\} \right) \times p(i) \right] \right\}$$

where, $\{M_{LLL1}, M_{LLH1}, \cdots, M_{H4}\}$ is the set of the values of the oversampling factor of temporally decomposed frames in a GOP. The expression maximized is the expected value of the PSNR of reconstruction of the GOP. This is obtained by cycling through all loss patterns except the case that all descriptions are lost and weighting the PSNR by the probability $p(i)$ of the particular loss pattern occurring. $p(i)$ can be easily calculated from the description loss probability which is assumed to be known to the MD encoder.

2.2. Simplifying the search for M

It is shown in [5] that for a MC lifted temporal decomposition, the total distortion in a reconstructed GOP is a weighted sum of the distortions in each temporally transformed frame. It is also true that the distortion in a temporally transformed frame is independent of the distortion in any other temporally transformed frame. The expected value of the PSNR of reconstruction of every temporally transformed frame can therefore be individually maximized to yield a set of values of M. In other words

$$M_x = \arg\max_{M_x}\left\{\sum_{i=1}^{2^N-1}\left[PSNR(i, M_x)\times p(i)\right]\right\}$$ where

$x = LLL1, LLH1, \cdots, H4$.

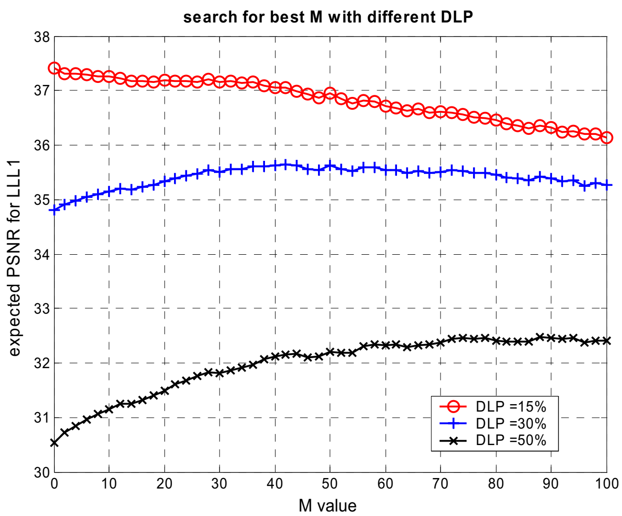

Figure 3: Search for M for LLL1. Description loss probability (DLP) = 15%, 30%, and 50%.

The higher the description loss probability, the higher is the amount of extra redundancy to be added. This is illustrated by Figure 3 for the LLL1 frame. (Note the interesting fact that for DLP=15%, it does not pay off to add any extra redundancy since the signal has enough correlation for this case.) Knowing the description loss probability, the range for sweeping the values of M can be narrowed down and does not need to be between 0 and 100 %. In our scheme we use JPEG2000 [6] for the spatial encoding of every temporally transformed frame. This allows us to specify the rate for encoding every description of each temporally transformed frame. While sweeping M we keep the total number of bits for a particular frame constant by using this feature of the Kakadu software [6].

3. RESULTS

The performance of multiple description coding schemes is often compared using the so called redundancy rate distortion (RRD) plot. Figure 4 shows the RRD plot averaged over 88 frames of the Foreman sequence. The image size is 288x352 and only the Y component is transmitted. The GOP size is 8 frames and altogether 4 descriptions are formed from every temporally transformed frame. The RRD plot shows the effect of losing two descriptions completely. The upper curve shows the performance when all the M are optimally searched by setting description loss probability (DLP) equal to 0.5. The lower curve is obtained by using no pre-processing and hence no extra redundancy. Table 1 shows the comparison in more detail. The left half of the table shows the performance when all the M are optimally searched using description loss probability 0.5. The right half of the table shows the performance with all M equal to zero. All columns marked with "A" depict the situation when no loss occurs on the network. All columns marked with "B" depict loss of two descriptions. It can be seen that when no loss occurs there is a penalty to be paid in performance due to the added redundancy. But when loss occurs according to the loss probability assumed by the MD encoder then the performance is much higher compared to the case of all M set to zero.

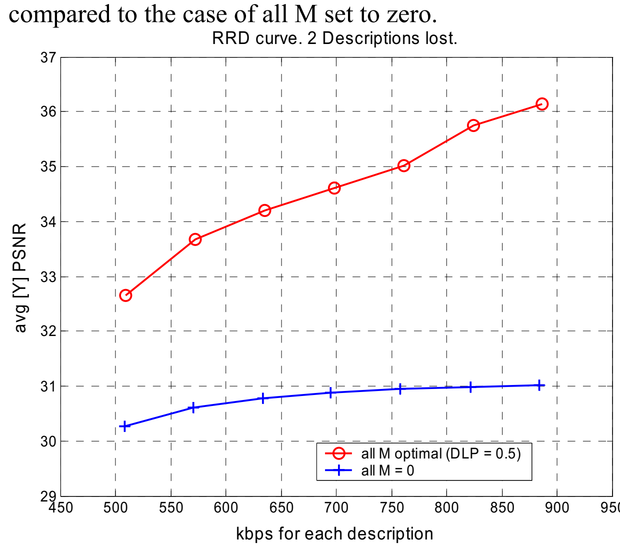

Figure 4: RRD plot for the test sequence Foreman (CIF) at a DLP of 50%.

In our scheme (as well as in DC-MDVC of [1]), the MV data has to be repeated across all descriptions. The

MVs are encoded using Huffman tables and account for a large portion of the bit-rate. Fig. 5 shows the performance details for a smaller image size (144x176).

Table 1: Performance Details

All M optimal			All M = 0		
kbps per descript.	A Avg [Y] PSNR	B Avg [Y] PSNR	A Avg [Y] PSNR	B Avg [Y] PSNR	kbps per descript.
508.87	33.68	32.65	35.35	30.27	508.10
572.12	34.73	33.68	36.76	30.61	570.32
634.90	35.42	34.21	37.76	30.78	633.77
697.92	36.08	34.61	38.46	30.88	695.15
761.02	36.49	35.02	38.99	30.94	758.35
824.47	36.88	35.75	39.44	30.99	821.77
886.30	37.18	36.14	39.79	31.02	884.07

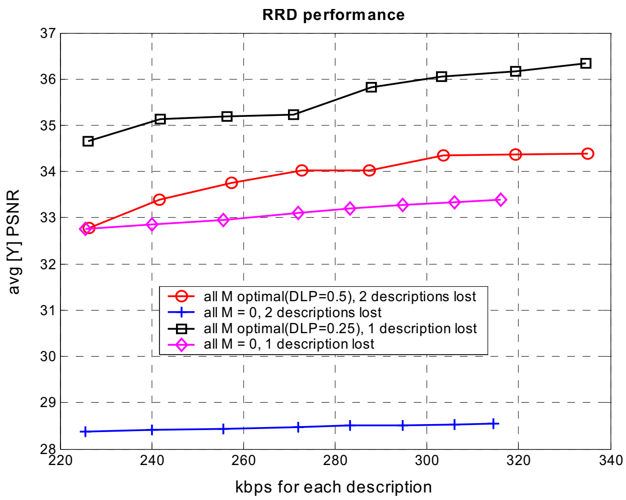

Figure 5: RRD plot for the test sequence Foreman (QCIF).

4. CONCLUSION

For high loss probability we need to insert redundancy across the descriptions generated by a MD video encoder. With the pre-processing proposed in [1] a flexible amount of redundancy can be inserted which allows graceful degradation. We apply this concept to video coding which is based on the MC lifted wavelet transform. This eliminates sending drift compensation data which has to be transmitted in case of the hybrid codec. We show that determining the global optimum of the amount of redundancy among the descriptions is straightforward given that the description loss probability is known. Our experimental results show that our MD codec significantly outperforms a codec without redundancy if the actual loss probability on the network is close to the loss probability assumed during optimization.

5. REFERENCES

[1] N. Franchi, M. Fumagalli, R. Lancini, “Multiple Description Coding For Scalable and Robust Transmission over IP,” Proc. of 13th Packet Video Workshop, Nantes, France, Apr. 2003.

[2] I.V. Bajic, J.W. Woods, “Domain-Based Multiple Description Coding of Images and Video,” IEEE Transactions on Image Processing Vol. 12, No. 10, Oct. 2003.

[3] W. Jiang, A. Ortega, “Multiple Description Coding via Polyphase Transform and Selective Quantization,” Proc. of VCIP ‘99, San Jose, USA, Jan. 1999.

[4] A.C. Miguel, A.E. Mohr, E.A. Riskin, “SPIHT for Generalized Multiple Description Coding,” Proc. of ICIP ‘99, Kobe, Japan, Oct. 1999.

[5] A.A. Mavlankar and E. Steinbach, “Distortion Prediction For Motion-Compensated Lifted Haar Wavelet Transform and its application to Rate Allocation,” Proc. of Picture Coding Symposium, San Francisco, Dec. 2004.

[6] D. Taubman and M.W. Marcellin, “*JPEG2000: Image Compression Fundamentals, Standards and Practice,*” Kluwer Academic Publishers, Nov. 2001.

[7] J.G. Apostolopoulos, “Reliable Video Communication over Lossy Packet Networks using Multiple State Encoding and Path Diversity,” Proc. Visual Communications and Image Processing, pp. 392-409, Jan. 2001.

[8] V. Vaishampayan and S. John, "Interframe Balanced Multiple Description Video Compression," in Proc. Packet Video 1999, New York, USA, April 1999.

[9] A.A. Mavlankar, “Robust Video Communication using motion-compensated lifted 3-D wavelet coding,” Master’s Thesis, Technische Universität München, Sept. 2004.

[10] 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 Trans. Circuits Syst. Video Technol., vol. 10, pp. 1403-1412, Dec. 2000.

[11] C.-S. Kim and S.-U. Lee, “Multiple description motion coding algorithm for robust video transmission,” in Proc. IEEE Int. Symp. Circuits and Systems, Lausanne, Switzerland, May 2000, vol. 4, pp. 717-720.

[12] A.R. Reibman, H. Jafarkhani, Y. Wang, M.T. Orchard, and R. Puri, “Multiple description coding for video using motion compensated prediction,” Proc. IEEE Int. Conf. Image Processing, Kobe, Japan, Oct. 1999.

[13] V. K Goyal, Multiple Description Coding: Compression Meets the Network,”, *IEEE Signal Processing Mag.*, vol. 18, no. 5, pp. 74-93, Sept. 2001.

[14] B. Pesquet-Popescu, V. Bottreau, “Three-dimensional lifting schemes for motion compensated video compression,” Proc. IEEE Int. Conference on Acoustics, Speech and Signal Processing, Vol. 3, Salt Lake City, UT, 2001, pp. 1793–1796.

[15] L. Luo, J. Li, S. Li, Z. Zhuang, Y.-Q. Zhang, “Motion compensated lifting wavelet and its application in video coding,” Proc. IEEE International Conference on Multimedia and Expo, Tokyo, Japan, 2001, pp. 481–484.

[16] A. Secker, D. Taubman, “Motion-compensated highly scalable video compression using an adaptive 3D wavelet transform based on lifting,” Proc. IEEE Int. Conference on Image Processing, Vol. 2, Thessaloniki, Greece, 2001, pp. 1029–1032.