2005

Robust video transmission based on multiple description scalable coding with EREC

Haifeng Zheng

Lun Yu

Chang Wen Chen

Robust Video Transmission based on Multiple Description Scalable Coding with EREC [*]

Haifeng Zheng[a], LunYu[a], Chang Wen Chen[b]

[a]College of Information Technology, Fuzhou University, Fuzhou, Fujian 350002 CHINA
[b]Dept. Electrical & Computer Eng., Florida Institute of Technology, Melbourne, FL 32940 USA

ABSTRACT

This paper presents a multiple description scalable video coding scheme based on overcomplete motion compensated temporal filtering, named MD-OMCTF, for robust video transmission over wireless and packet loss networks. The intrinsic nature of the structure of OMCTF and embedded coding with modified SPIHT algorithm enable us to provide fully scalable properties for the proposed scheme. We show that multiple description coding is very effective in combating with channel failures in both Internet and wireless video. The integration of MD with OMCTF allows us to achieve both loss resilience and complete scalability. In order to further improve error-resilience to channel bit error for this scheme and reduce error propagation in error-prone network, we apply error resilient entropy coding (EREC) to the multiple bitstreams to gain additional error resilience. With EREC, multiple bitstreams are reorganized into fixed-length slots so that synchronization of the beginning of each bitstream can be automatically obtained at the receiver. The integration of scalable coding and EREC with MDC enables the coded video bitstream to be adaptive to the varying channel condition and to be resilient to both transmission losses and bit errors. We also develop corresponding error concealment scheme to recover the lost or erroneous information during video transmission. Experimental results show that the proposed scheme is able to achieve robust video transmission over both wireless and packet loss networks.

Keywords: Multiple description video coding, scalable video coding, overcomplete motion-compensated temporal filtering, wireless video, packet loss networks, EREC, error concealment

1. INTRODUCTION

Video communication over error-prone network such as Internet and wireless networks has been a significant challenge for current multimedia communication technology. Since channels in such networks are unreliable and their bandwidth varies with time, it is very much desired for video coding schemes to be able to adapt to the channel conditions, including scalability and error resilience. In general, compressed video is very sensitive to error-prone environment; a single bit error may cause the loss of decoding synchronization and severe degradation to received video quality. To ensure reasonable video quality during transmission over data loss channel such as Internet, multiple description coding (MDC) has been proposed to improve the robustness of image and video transmission [1-5]. The objective of MDC is that if all bit-streams have been received correctly, a high quality video sequence can be reconstructed, whereas, if some bit-streams have been lost, a low-quality, but acceptable reconstruction can still be reconstructed from the received description. As a result, this loss resilient source coding strategy is very suitable for wireless packet network and multiple antennas system. Since wireless packet network are very sensitive to time delay in video transmission and the decoder terminal cannot wait for all packets to be received to start decoding. The decoder reconstructs video quality only from received descriptions within a limited time. With MDC strategy, we also can reconstruct the whole video information by estimating lost descriptions from the correctly received ones.

Furthermore, video signals may be transmitted over heterogeneous networks where a wide variety of users are serviced by different available bandwidths and different memory and computational power in their terminal devices. To this end, scalable video coding schemes have been proposed which produce bit-streams decodable at different spatial resolution, different quality resolution and different frame-rates. Besides the advantage for the heterogeneous networks, we argue

---

[*] Send Correspondence to:
Haifeng Zheng: Email: zhenghf@fzu.edu.cn
Lun Yu: Email : yulun@fzu.edu.cn
Chang Wen Chen: Email: cchen@fit.edu

Visual Communications and Image Processing 2005, edited by Shipeng Li,
Fernando Pereira, Heung-Yeung Shum, Andrew G. Tescher, Proc. of SPIE Vol. 5960
(SPIE, Bellingham, WA, 2005) · 0277-786X/05/$15 · doi: 10.1117/12.633197

that the scalable feature of the compressed video is also useful to guarantee robust video transmission. As is well known, the wireless channel is time-varying. With scalable coding, we can enable the compressed video bitstream to be adaptively decoded according to channel condition. This scalable coding property, combined with multiple description coding, would allow us to obtain the basic reconstructed video quality, even when some channel links are corrupted by errors and losses. Once we have successfully combined multiple description coding and scalable coding, we believe we can achieve substantial improvement in terms of received video quality, flexible adaptability and transmission robustness over Internet and wireless packet video transmission networks.

A great amount of research dedicated to MDC has been proposed for image and video transmission over error-prone channel [1-5]. However, little research has been done to address simultaneously multiple description coding and scalable coding, two equally important features for the robust video transmission over wireless networks. More recently, a multiple description scalable coding (MDSC) [6] was proposed by Bajic and Woods. The scheme, denoted as domain-based MDC, partitions the wavelet coefficients after wavelet decomposition into multiple descriptions by a maximally separated strategy. The advantage of this scheme is that simple error concealment methods can be employed to estimate lost descriptions by maximally separated strategy. Another scheme for MDSC was proposed by Van der Schaar, which is denoted as Multiple Description Motion Compensated Temporal Filtering (MD-MCTF) [7]. The proposed MDSC scheme is based on the framework of motion compensated temporal filtering, which can provide scalable (temporal, SNR, spatial and complexity) property for bit-streams. The scheme splits bit-streams into two substreams using MCTF-based wavelet coding, and then use inherent nature of MCTF scheme, such as the lifting algorithm of temporal filtering to recover missing descriptions.

In this paper, we propose a novel multiple description scalable coding scheme, called MD-OMCTF, based on the framework of overcomplete motion compensated temporal filtering [8,9]. Unlike MD-MCTF scheme, which employs MCTF on original image, our scheme applies MCTF on the wavelet-domain of a frame after discrete wavelet transform. It has been well recognized that some problems like occlusion and photometric distortion can be handled more effectively by overcomplete MCTF than spatial-domain MCTF [9]. With the structure of overcomplete MCTF, we can easily obtain temporal scalability for this scheme. Moreover, in order to further provide spatially and SNR scalable feature, we develop a modified SPIHT algorithm without sacrificing other important features of the original SPIHT, including coding efficiency and SNR scalability. As a result, the proposed scheme enables us to provide fully scalable functionalities for the compressed video.

Each description of the compressed video will be further decomposed into multiple bitstreams. With multiple bitstreams transmission, we have shown that it is very effective in combating error propagation in both Internet video streaming and mobile wireless video [20]. Furthermore, in order to achieve high performance in video coding while still maintain error resilience to channel errors, we also introduced EREC [10] for each description of the video bitstream. The idea of EREC is to re-group the variable-length data blocks into fixed-length slots so that re-synchronization can be obtained naturally at the receiver. The adoption of EREC enables us to generate entropy coded video bitstreams that are self-synchronized for robust decoding at the receiving end.

The rest of the paper is organized as follows. In Section 2, we summarize the overall system structure of the proposed scheme. In Section 3, we discuss the proposed scheme in detail. This section includes the discussion on video coding based on overcomplete MCTF, multiple description scalable coding, EREC and its implementation in the proposed scheme. The error concealment algorithm to recover the missing information will also be discussed in this section. In Section 4, we present the experimental results to verify the performance of the proposed scheme. Finally, we conclude with a summary in Section 5.

2. SYSTEM OVERVIEW

Figure 1 shows the block diagram of the proposed MD-OMCTF system. At the encoder side, the input video sequence is first divided into groups of pictures (GOPs). In this case, we use 8 frames per GOP. Then, for each GOP, the frames are hierarchically decomposed by the critically-sampled discrete wavelet transform (DWT). The number of spatial decomposition level is determined by the input sequence format. Furthermore, we adopt the motion compensation based approach to de-correlate the video signal along the temporal dimension using 3D-lifting scheme [11,12], which jointly implements the temporal WT and MC. To avoid the shift-variant problem existed inherently in critically-sampled discrete wavelet transform and to achieve high coding performance, motion estimation and motion compensation is performed in the overcomplete wavelet domain. This process is called overcomplete motion compensated temporal filtering.

After the overcomplete motion compensated temporal filtering step, the generated texture information and motion vector information are sent into the MDC coder. The MDC coder distributes the bitstream into each description. Meanwhile, we developed modified SPIHT algorithm without sacrificing other important features of the original SPIHT [13], such as coding efficiency and SNR scalability. Using modified SPIHT for subband coding, the encoded bitstream not only has spatial scalability feature but also keeps the full SNR embedded features at any required resolution level. Furthermore, since motion estimation is performed in overcomplete wavelet domain in a level by level fashion, the produced motion vectors can also be easily used to perform motion compensation at the desired resolution.

Meanwhile, in order to make source coding resilient to channel errors, we also develop a modified SPIHT algorithm, borrowed from Creusere's work [14], for partitioning wavelet coefficients into a number of independent spatial orientation trees, so that an error in one tree does not affect the others. However, the modified SPIHT algorithm is still sensitive to bit errors, because it produces variable-length bitstreams. A single bit error may lead to loss of synchronization between encoder and decoder. In order to generate entropy coded video bitstreams that are self-synchronized for robust decoding at the receiving end, we also introduce EREC [10] to generate fix-length bitstreams. In the following sections, we shall discuss more technical details for this scheme.

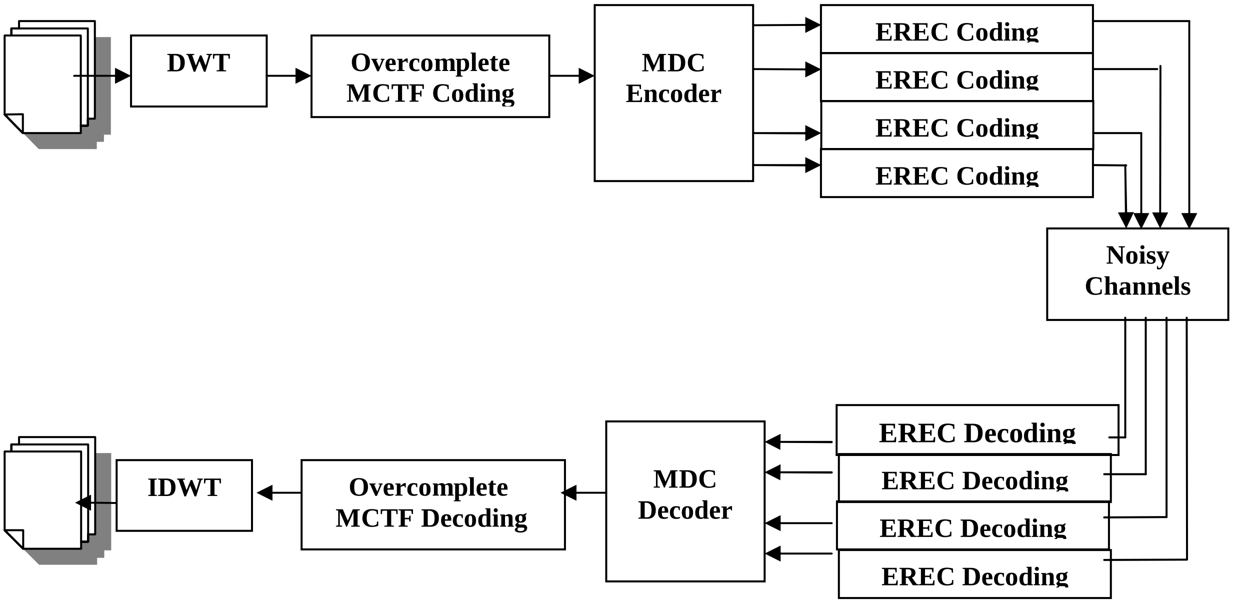

Figure1: The block diagram of the proposed MD-OMCTF system.

3. THE PROPOSED MD-OMCTF SCHEME

3.1 Video coding based on OMCTF

Recently, motion compensated temporal filtering (MCTF) has been proposed to replaced motion compensated prediction (MCP) for developing scalable coder. In particular, MCTF with lifting scheme has a nice property of reversibility, which could not only effectively eliminate the temporal redundancy but also handle the problem of error propagation [9]. In this proposed scheme, since MCTF is performed after spatial decomposition of the input video sequences, due to the problem caused by shift-variant property, important information of motion accuracy would be lost if motion estimation and motion compensation (ME/MC) is performed only in critically-sampled wavelet domain. Consequently, in order to achieve high coding efficiency, ME/MC should be performed in overcomplete wavelet domain. Some techniques have been proposed to solve the shift-variant problem with wavelet transform for MC/ME purpose, such as Low-Band-Shift (LBS) method [15], Phase-Shifting-Filter (PSF) [16] and Prediction-Filters [17].

As mentioned before, after the video sequence is split into GOPs and the spatial analysis of the video by WT for each frame, overcomplete MCTF is performed among frames within a GOP. Figure 2 shows overcomplete MCTF operation with lifting scheme along temporal axis. In this scheme, we use Haar filters for lifting algorithm with prediction and updating operations. The lifting formulation can be written as follows [9]:

Prediction:

$$\boldsymbol{H(m,n) = \frac{1}{\sqrt{2}}(x_{cur}(m,n) - P_f[x_{ref}(m,n)])} \tag{1}$$

Updating:

$$\boldsymbol{H(m,n) = \sqrt{2}x_{ref}(m,n) - P_b[H(m,n)])} \tag{2}$$

Where $\boldsymbol{P_f}$ and $\boldsymbol{P_b}$ are forward and backward prediction operations, respectively; $\boldsymbol{H(m,n)}$ and $\boldsymbol{L(m,n)}$ are the high-pass and low-pass frames, respectively.

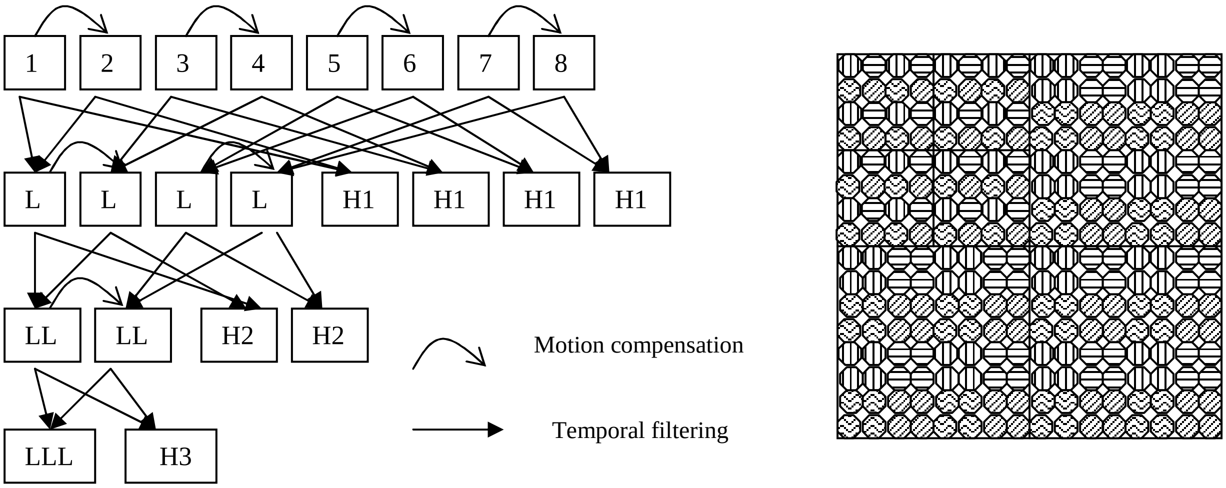

Figure 2: Left: Overcomplete MCTF structure with lifting scheme; Right: An example of dispersive partitioning.

There have been various existing schemes for motion estimation in overcomplete discrete wavelet domain, in a fashion either band by band or level by level. That is, motion vectors are determined separately either for each individual subband or for each spatial transform level [18]. However, the estimation of motion within the individual subbands LH, HL and HH does not work satisfactory, since minimizing the sum of absolute differences (SAD) does not result in the estimation of true motion and a homogeneous vector field [18]. Therefore, we choose the level by level fashion to perform motion estimation, in which case the triple subbands (LH, HL, HH) are jointly considered for best-matching estimation. Thus, only one motion vector of each triple macroblocks for each level is needed. This approach not only reduces computational complexity but also reduces coding cost. Furthermore, motion estimation occurs in wavelet domain, which facilitates the hierarchical block matching algorithm (HBMA) with variable block-size.

Meanwhile, we also develop a modified SPIHT algorithm for coding the spatio-temporal subband coefficients. This algorithm not only extents spatial scalability for the proposed video coding scheme but also provides error resilience for compressed video. Since the original SPIHT algorithm is very sensitive to bit errors in that any single-bit transmission error may lead to corruption of the whole decoded video quality, in this scheme, we adopt a method developed by Creusere [14] for partitioning the wavelet coefficients into groups and independently encoding and decoding with SPIHT algorithm. With such a modification, a bit error in the bitstream may corrupt only one block and does not affect the others. The details of the modified SPIHT algorithm will be described in Section 3.3.

3.2 Multiple description coding

A challenging task for a multiple description coding scheme is how to distribute the total bitstream into each description. Bit allocation strategy is important for a good MDC scheme. In our previous work [20], we have found out that motion vectors are crucial component in a video codec in that any bit error in the motion vectors information would cause serious error propagation within a GOP and greatly degraded video quality. Furthermore, based on the structure of overcomplete MCTF, after 3 level temporal filtering, the generated LLL frame remains the most important texture information within a GOP. As a result, in each description, it is desired to send redundant information, including the first frame and motion vectors information in order to guarantee basic video quality in the case of channel error corruption. Consequently, the overall redundancy of the proposed scheme includes three extra frames and three copies of motion vectors in the case of four descriptions. Of course, we can encode the extra frames in different bit rate according to different channel condition. As for the LLL frame and H frames, we then use a dispersive partitioning strategy to partition the wavelet coefficients into four descriptions. The illustration of the partitioning method is shown in Figure 2. As demonstrated in Figure 2, dispersive partitioning strategy groups adjacent wavelet coefficients in the lowest subband into different descriptions. Since each spatial orientation tree is encoded and decoded independently with modified SPIHT algorithm, wavelet coefficients in the high subband are grouped into corresponding wavelet trees according to their roots. The main advantage of dispersive partitioning method is to achieve error resilience and to facilitate the error concealment to be performed for lost descriptions. For example, when the wavelet coefficients in one wavelet tree are lost, we can employ appropriate interpolation algorithm to recover them using the surrounding coefficients in the other trees. Even when some descriptions are lost entirely, we can still recover missing coefficients from the other correctly received descriptions. In this case, each description shall have the basic information for reconstructing the coarse video quality.

3.3 Scalability of MD-OMCTF scheme

In this subsection, we shall address the scalability of MD-OMCTF scheme. Since multimedia information such as compressed video may be transmitted over heterogeneous networks and the receiving terminals may decode bitstreams according to network condition at that time, scalability is very much desirable for contemporary multimedia application. However, the original SPIHT algorithm is unable to provide spatially scalable functionality because it encodes the entire wavelet tree in a bit-plane by bit-plane fashion and the encoded bitstream does not assign information in an order according to the different spatial resolutions. To support the desired spatial scalability, we modified the original SPIHT algorithm such that the bit-stream not only has spatial scalability feature but also keeps the full SNR embedded features at any required resolution level.

Since the SPIHT algorithm is based on the multi-resolution wavelet decomposition, it is relatively easy to add spatially scalable feature. As we can see from SPIHT algorithm, it sorts the wavelet coefficients in three ordered lists: the list of insignificant sets (LIS), the list of insignificant pixels (LIP), and the list of significant pixels (LSP). During the sorting pass and refinement pass stages, each wavelet coefficient to be processed with its coordinate is sent into corresponding list. To enable spatial scalability, in the encoder, what we need to do is to allocate several output buffers according to spatial decomposition, and then put the encoded bits into corresponding output buffer according to the location of wavelet coefficients. At the decoder, what we need to do is to decode only the bitstream in the desirable resolution level and skip the processing of the pixels in three lists beyond that resolution level. This idea is quite simple: we only partition the bitstream into portions according to their corresponding spatial locations and decode the bitstream belonging to the desired resolution level. The advantage of this method is that we need not to change the structure of three lists and add any additional list for the encoding and decoding algorithms.

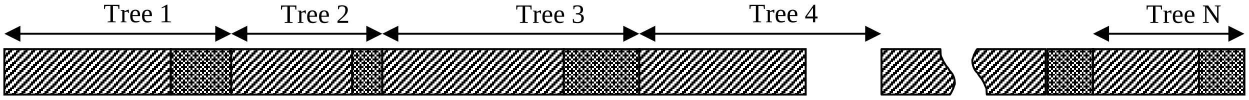

Figure 3: Partitioning the bitstream of each wavelet tree into portions according to its corresponding spatial locations.

As mentioned before, in order to ensure error resilience for source coding, we partition wavelet coefficients into a number of independent spatial orientation trees to avoid error propagation across from one tree to the others. As a result, we employ the modified SPIHT algorithm to this partitioning strategy. Further more, since motion estimation is performed in overcomplete wavelet domain using a level by level fashion, the produced motion vectors are also easy to be used to perform motion compensation at the desired resolution level. Figure 3 shows the illustration of partitioning the bitstream into portions according to resolution level.

Furthermore, due to the nature of hierarchical temporal decomposition structure using overcomplete MCTF, the temporal scalability is intrinsic in MD-OMCTF scheme. As seen from Figure 2, the scheme enables the desired temporal scalable functionality. In the scheme with three-level temporal decomposition, when only the LLL-frame in the third level of temporal decomposition is decoded, we can obtain eighth of original frame rate. When LLL-frame, H3-frame and corresponding motion vectors in the third level of temporal decomposition can be decoded, we obtain quarter of the original frame rate. When the information in the third level and the second level of temporal decomposition can be decoded, we obtain half of original frame rate. When the entire bit-stream is decoded, we can achieve full frame rate.

The scalability of the scheme also makes the coding system more error resilience. With spatial scalability, if we detect corrupted wavelet coefficients in the high subband, we can decode the bitstream and reconstruct the video quality with only low subband wavelet coefficients at a low resolution. Furthermore, with overcomplete MCTF, if LLL-frame at the last stage of temporal decomposition of only one description is received correctly, we can still obtain the basic video quality based the received LLL-frame.

3.4 EREC and error concealment

Error resilient entropy coding (EREC) is an effective means to recover lost or erroneous information to enhance error resilience for coding variable-length blocks of data. The scheme is originally proposed by Redmill et al [10] to handle the sequential transmission of DCT coded data blocks over noisy channels. The key of the EREC is to re-organize the variable-length data blocks into fixed-length slots with negligibly increased data size. Figure 4 shows EREC bit re-organization algorithm. It has been shown that the EREC can significantly reduce the channel error propagation effects and that remaining channel error propagation is most likely to affect data from the end of longer blocks [10]. The characteristic of EREC decoding is very suitable for some embedded coding methods, such as SPIHT algorithm. As we know, SPIHT encodes the wavelet coefficients from high bit-plane to low bit-plane by threshold. This suggests that the encoded bitstream contains more important information in the beginning of the bitstream.

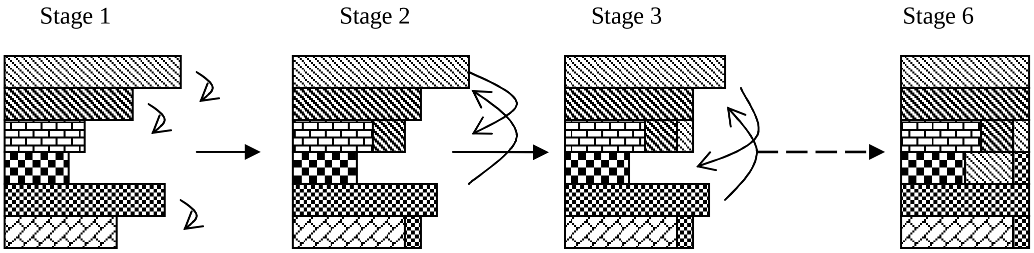

Figure 4: EREC bit re-organization algorithm

In the EREC application in image based on wavelet transform, Cao and Chen have proposed a method to use wavelet tree coding based on SPIHT to re-organize the data into multiple independent short bitstreams, and then apply EREC to re-organize the bitstreams into fixed-length slots so that synchronization of the beginning of each bitstream can be automatically obtained at the receive [19]. In this paper, we extend the EREC to the MD-OMCTF scheme. After we complete the wavelet transform in each frame and the overcomplete MCTF operation among frames within a GOP, the modified SPIHT is applied for the generated wavelet coefficients. Then, we apply EREC to reorganize variable-length bitstreams coded with the modified SPIHT into fixed-length data slots in each description. As indicated in previous section, with the modified SPIHT, each bitstream of wavelet tree has several portions according to their spatial locations, thus, when a data slot is corrupted, the probability of bit error occurring in higher subband location is higher than that in low subband.

If errors are detected in high subband, we could decode bitstream with the modified SPIHT by skipping the high subband. Even if an error occurs in the first several bits, we could still estimate those lost wavelet coefficients from adjacent wavelet coefficients. In this scheme, we adopt the method proposed by Cao and Chen [19]. First, in the EREC encoding stage, we add one bit for the first several bits in each slot data for the purpose of error detection. This is because the first several bits coded by SPIHT have higher energy than other bits. Since MDC strategy assumes that not all bitstreams experience failures simultaneously, thus, if we detect the errors occurring in one slot in some description, most likely we can replace the slot data with corresponding uncorrupted slot in the other descriptions. Also, even if the slot data in the same location in all descriptions are corrupted, we can still apply error concealment method as the last resort.

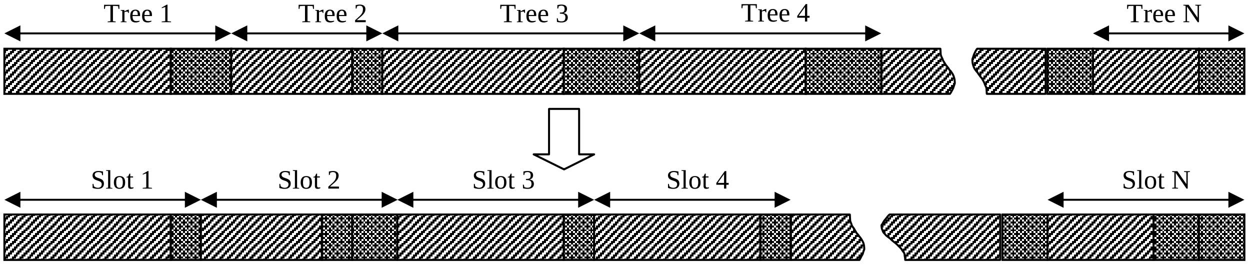

Figure 5: Partitioning each variable-length bitstream with spatial scalability into fix-length segment by EREC.

4. SIMULATION RESULT

In this section, we will first verify the scalability of the proposed scheme. As described in the previous section, the scalability features are intrinsic in MD-OMCTF scheme due to the structure of overcomplete MCTF, and the nature of the modified SPIHT algorithm. Therefore, MD-OMCTF can provide not only the temporal scalability but also the spatial and SNR scalabilities. As a result, the decoder should be able to decode the bitstreams from the received descriptions at any desirable spatial resolution, quality resolution and frame rate. In this paper, we present the experiment on the spatial scalability. Figure 6 shows one frame of the Foreman sequence decoded in 1/4QCIF format from the bitstream corresponding to QCIF format. From the figure we can see that overcomplete MCTF have the advantage in spatial scalability than the spatial-domain MCTF because there exists no drift problem in spatial scalability. We simply skip the high band information decoding instead of sub-sampling from the reconstructed decoded image.

Figure 6: The right frame is decoded in 1/4 QCIF format from its corresponding frame in the left in QIF format.

Next, we conduct experiments to show the performance of the proposed scheme over wireless channels. The simulations are tested on standard video sequence “Foreman”, whose frames are in QCIF format with only luminance component. In this scheme, four descriptions are generated from the complete bitstream. For spatial decomposition we apply the 9-7 bi-orthogonal filter and three level of decomposition. For overcomplete MCTF, we apply Haar filter and three level of temporal decomposition. For wavelet tree coding, we adopt the modified SPIHT algorithm without arithmetic coding. In order to enhance the error resilience of the coding system, coding efficiency is slightly decreased. As indicated earlier, the LLL frame and motion vectors information are contained in each description. Thus, the compressed bit rate for each description is about 1 bpp. Simulations are performed using BSC channels at different BERs. There BERs are 0, 0.001, 0.002, 0.005, and 0.01 respectively. As discussed in the previous section, we apply the error detection strategy so as to facilitate the corresponding error concealment scheme. In this approach, we add one parity bit for the first 16 bits in every slot. Error concealment scheme is only applied to the LLL frame. When errors are detected in some slot data, we will then check whether or not the other slot data at the same location in the other descriptions remains uncorrupted. If so, we shall replace the corrupted slot data with the uncorrupted one from other descriptions. Even when we are not able to find the correct slot data, then, we can still compensate these “bad” wavelet coefficients in the corrupted slot from their surrounding wavelets coefficients in the lowest frequency subband.

Figures 7 and 8 present frame by frame comparison of PSNRs of the reconstructed frames at different BERs before and after error concealment. Figure 9 shows the comparison of average PSNRs for “Foreman” sequence at different BERs before and after error concealment. From these Figures, we can demonstrate that visual quality of the reconstructed frames have been gradually improved as channel BER decreases, and the average PSNRs of the reconstructed frames with error concealment are higher than those without error concealment. Moreover, the improvement in both visual quality and the PSNR due to error concealment increases distinctly as the channel BER increases. These results show the importance of error concealment especially in the case of high channel BER. Figure 10 shows the comparison of visual results of the reconstructed frames with different channel conditions. From these sample frames, we can conclude that the proposed scheme can obtain reasonable video quality even when suffered from high bit error rate. Figure 11 shows both visual quality and PSNR results of sample frames before and after error concealment to demonstrate the importance of error concealment for the proposed multiple description scalable video coding for error prone channels.

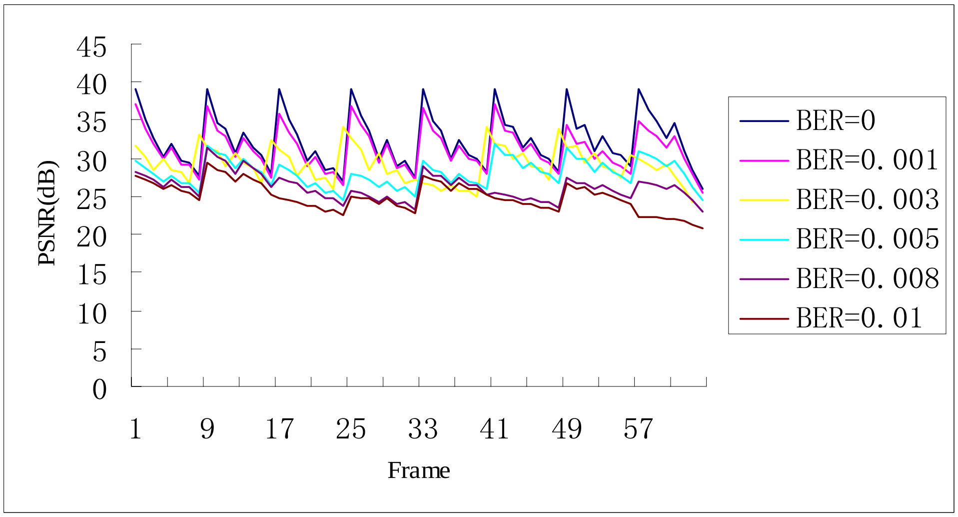

Figure 7: PSNRs comparison of frame by frame of “Foreman” sequence with different BERs and without Error Concealment.

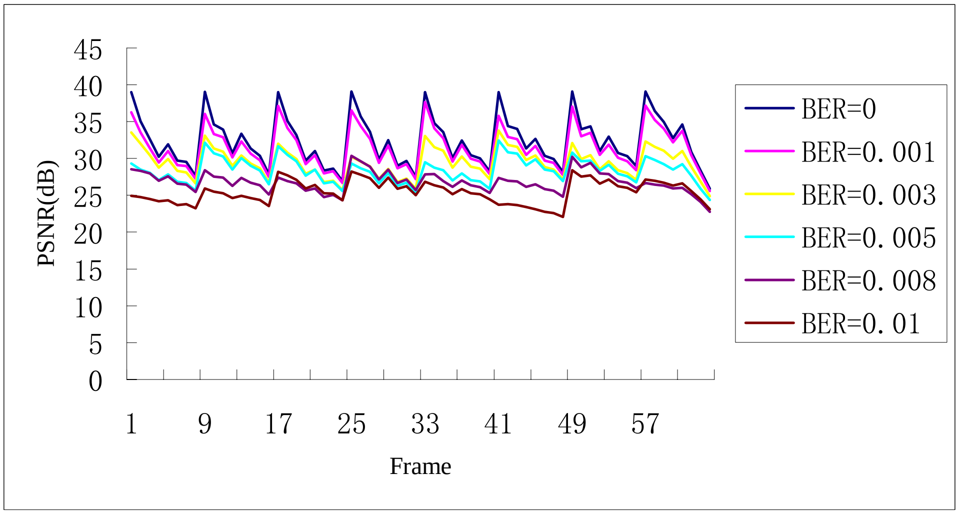

Figure 8: PSNRs comparison of frame by frame of “Foreman” sequence with different BERs and with Error Concealment.

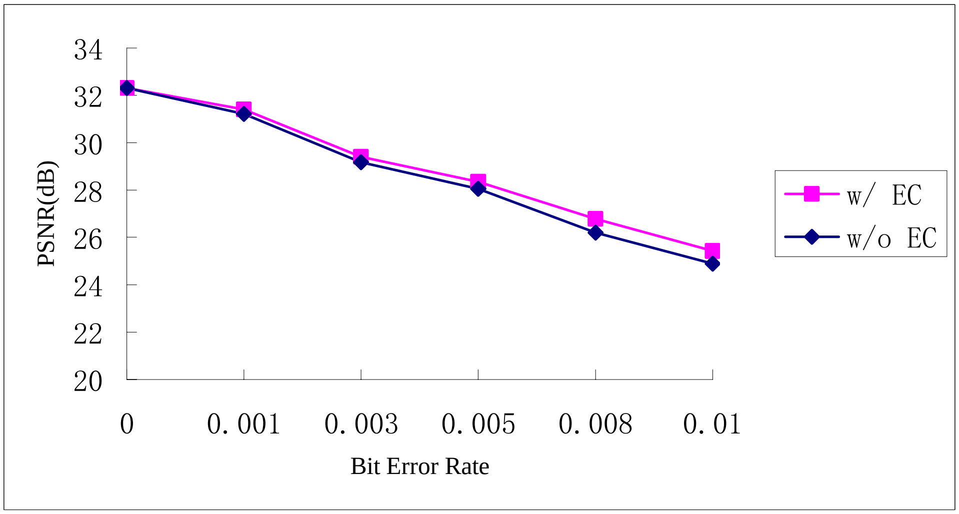

Figure 9: Average PSNRs comparison of “Foreman” sequence with different BERs with and without Error Concealment.

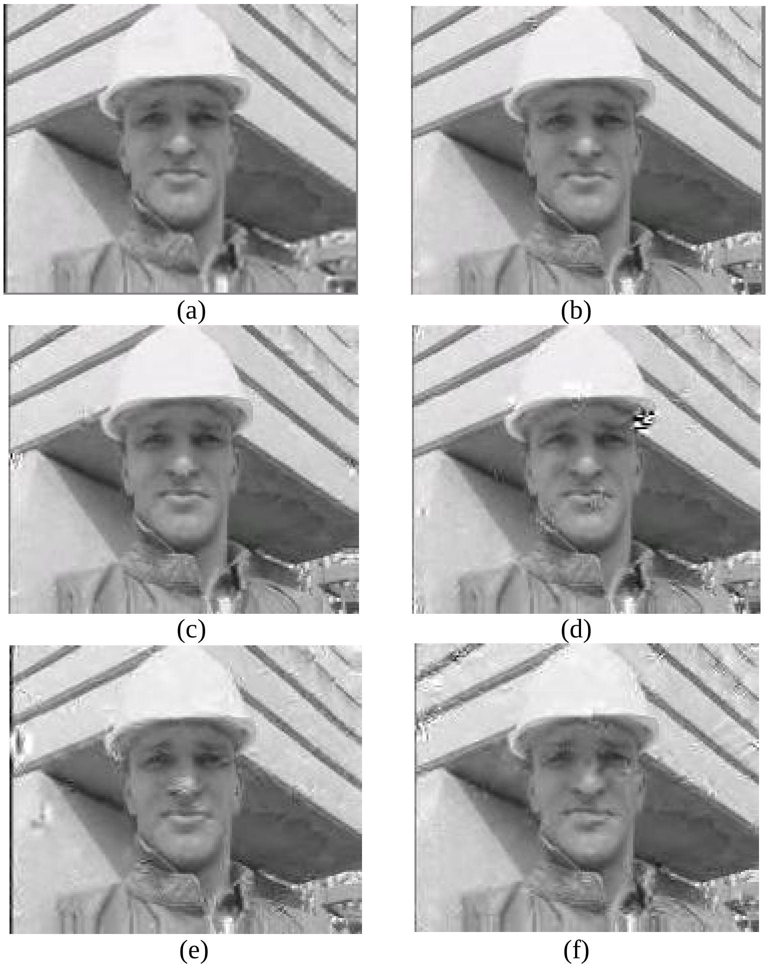

Figure 10: Samples of reconstructed frame over BSC channels with BERs at (a) 0, (b) 0.001, (c) 0.003, (d) 0.005, (e) 0.008 and (f) 0.01, respectively. The PSNR for these frames are (a) 29.70dB, (b) 29.21dB,(c) 28.42dB, (d)27.89dB, (e) 26.33dB, and (f) 25.78dB.

PSNR=30.193 (Before EC) PSNR=31.228 (After EC) BER=0.001

PSNR=26.695 (Before EC) PSNR=27.845 (After EC) BER=0.003

PSNR=24.585 (Before EC) PSNR=27.478 (After EC) BER=0.005

PSNR=21.308 (Before EC) PSNR=24.901 (After EC) BER=0.008

PSNR=20.703 (Before EC) PSNR=23.383 (After EC) BER=0.01

Figure 11: Simulation results of corresponding error concealment scheme based on the proposed multiple description scalable video coding. Left column: Visual quality and PSNR results of sample frames before error concealment; Right column: Visual quality and PSNR results of sample frames after error concealment.

To demonstrate the performance of the proposed scheme for video transmission over packet loss neworks, we also conduct the simulations under various packet loss environments. In this experiment, after EREC encoding, the wavelet coefficients of each frame are distributed into slots with the same number bits and each packet contains 11 slots. In this simulation, we also add one parity bit for each slot. When packet loss occurs, we set the parity bit to one for each slot in a packet. And in the decoding stage, when we detect the packet is lost, the decoding for the wavelet trees in the lost packet will be skipped. Figure 12 shows the average PSNR comparisons at different packet loss rates for the Foreman sequence at the packet loss rate of 0%, 1%, 5%, 10%, 15%, and 20%. From Figure 12, we can see that the video transmission performance degrades graciously as the packet loss rate increases. Table 1 presents the numerical results before and after error concealment. From these results, we can conclude that error concealment is necessary to provide enhanced video quality for video transmission over packet loss environments. The improvement in terms of PSNR can be up to 3dB for high packet loss environment.

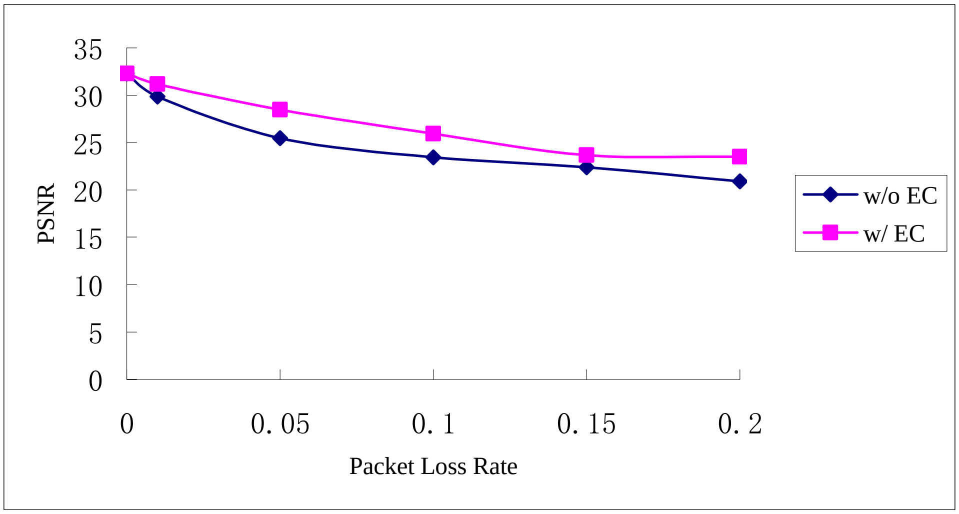

Figure 12: Average PSNRs comparisons at different packet loss rates.

Table 1: The average PSNRs results with EC and without EC at different packet loss rates.

Packet Loss Rate	Average PSNRs(dB)	
	w/ EC	w/o EC
0	32.310	32.310
1%	31.191	29.873
5%	28.481	25.462
10%	25.953	23.434
15%	23.673	22.399
20%	23.500	20.895

5. CONCLUSION

In this paper, we have presented a multiple description scalable coding scheme for video transmission over error-prone wireless networks and Internet environment. The scheme integrates multiple description coding, scalable coding, EREC, and corresponding error concealment to guarantee robust video streams transmission over error prone environments. The scheme is based on overcomplete motion compensated temporal filtering with lifting scheme, which provides not only fully scalable features but also error-resilient capabilities. Meanwhile, in order to allocate bit budget between sub-streams, we use a dispersive partitioning strategy to divide the total bitstream into different descriptions, so that the most important information is contained in each descriptions. This method enables the proposed scheme to combat the error propagation as the error propagation will be reduced to not only within one description but also between multiple descriptions. This scheme can also be extended to more descriptions. EREC adopted in this scheme re-groups the variable-length bitstreams into fixed-length slots, which can obtain re-synchronization naturally at the receiver, thus reducing sensibility of the compressed video bitstream to channel errors. We have conducted extensive simulations to verify the proposed multiple description scalable video coding. These simulation results show that the proposed scheme is able to obtain acceptable visual quality over wireless channel and packet loss network, even at high BER or high packet loss rate. We are currently investigating the integration of this scheme with MIMO-based mobile wireless video and QoS-based adaptive resource allocation for scalable video streaming over wireless and packet loss networks.

ACKNOWLEDGEMENTS

This work is supported by China National Natural Science Foundation under Research Grant No. 60328103.

REFERENCE

1. V. A. Vaishampayan, "Design of multiple description scalar quantizers," *IEEE Trans. Inform. Theory*, vol. 39, pp. 821-834, May 1993.
2. Y. Wang, M. T. Orchard, and A. R. Reibman, "Multiple description image coding for noisy channels by pairing transform coefficients", in *Proc. IEEE Workshop on Multimedia Signal Processing*, pp. 419-424, June 1997.
3. S. D. Servetto, K. Ramchandran, V. Vaishampayan, and K. Nahrstedt, "Multiple-description wavelet based image coding", *Proc.ICIP'98*, vol.1, pp. 659-663, Oct. 1998.
4. D.-M. Chung and Y. Wang, "Multiple description image coding based on lapped orthogonal transforms," *Proc. ICIP'98*, vol. 1, pp. 674-668, Oct. 1998.
5. Y. Wang and S. Lin, "Error resilient video coding using multiple description motion compensation," *IEEE Trans. Circus and Systems for Video Technology*, vol.12, no.6, pp.438-452, June, 2002.
6. I. V. Bajic and J. W. Woods, "Domain-based multiple description coding of images and video," *Proc. SPIE VCIP 2002*, vol.4671, pp. 124-135 San Jose, CA, Jan. 2002.
7. M. Schaar and D. S. Turaga, "Multiple Description Scalable Coding Using Wavelet-Based Motion Compensated Temporal Filtering," in *Proc. ICIP'03*, vol. 3, Barcelona, Spain, September 2003, pp. 489–492, Oct. 2003.
8. Y. Andreopoulos, A. Munteanu, J. Barbarien, M. van der Schaar, J. Cornelis and P. Schelkens, "In-band motion compensated temporal filtering**,**" *Signal Processing: Image Communication* (special issue on "Subband/Wavelet Interframe Video Coding"), vol. 19, no. 7, pp. 653-673, Aug. 2004.
9. X. Lin, "Scalable video compression via overcomplete motion compensated wavelet coding", *Signal Processing: Image Communication* (special issue on "Subband/Wavelet Interframe Video Coding"), vol. 19, no. 7, pp. 637-651, Aug. 2004.
10. D. W. Redmill and N. G. Kingsbury, "The EREC: An error-resilient technique for coding variable-length blocks of data," *IEEE Trans. on Image Processing*, vol. 5, pp. 565–574, April 1996.
11. B. Pesquet-Popescu, V. Bottreau, "Three-dimensional lifting schemes for motion compensated video compression", *Proceedings of International Conference on Acoustic Signal and Speech Processing*, pp. 1793-1796, May 2001.
12. A. Secker, D. Taubman, "Motion compensated highly scalable video compression using an adaptive 3D wavelet transform based on lifting", *Proc. ICIP'01*, pp. 1029-1032, Oct. 2001.
13. A. Said and W. A. Pearlman, "A new, fast, and efficient image codec based on set partitioning in hierarchical trees," *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 6, pp. 243–250, June 1996.
14. C. D. Creusere, "A new method of robust image compression based on the embedded zerotree wavelet algorithm," *IEEE Trans. on Image Processing*, vol. 6, pp. 1436–1442, October 1997.
15. H. W. Park and H. S. Kim, "Motion estimation using low-band-shift method for wavelet based moving-picture coding," *IEEE Trans. Image Processing*, vol.9. no.4, pp.577-587, April 2000.
16. X. Li, "New results of phase shifting in the wavelet space," *IEEE Signal Processing Lett*ers, vol.10, no.7, pp.193-195, July 2003.
17. Y. Andreopoulos, A. Munteanu, G. Van der Auwera, J. Cornelis and P. Schelkens, "Complete-to-overcomplete discrete wavelet transforms: theory and applications," *IEEE Trans. on Signal Processing*, vol. 53, no. 4, pp. 1398-1412, April 2005.
18. C. Mayer, "Motion compensated in-band prediction for wavelet-based spatially scalable video coding," *Proc. IEEE Int. Conf. on Acoust., Speech, Signal Processing*, Hong-Kong, CN, vol. 3, pp. 73-76, April 2003.
19. L. Cao and C. W. Chen "Robust image transmission based on wavelet tree coding, error resilient entropy coding, and error concealment," *Journal of Electronic Imaging*, vol.13, pp. 646-653, July 2004
20. H. Zheng, L. Yu, C. W. Chen, "A Novel Multiple Description Scalable Coding Scheme for Mobile Wireless Video Transmission" *Proc. SPIE Multimedia over Mobile Devices*, Vol. 5684, pp. 1-11, San Jose, CA, Jan. 2005.