VIDEO TRANSCODING FOR PACKET LOSS RESILIENCE BASED ON THE MULTIPLE DESCRIPTIONS

Il Koo Kim and Nam Ik Cho

Inst. of New Media and Communications
School of Electrical Eng., Seoul National Univ., Seoul 151-742, KOREA
lit2eng@ispl.snu.ac.kr, nicho@snu.ac.kr

ABSTRACT

This paper proposes error resilient video transcoding structures based on the multiple description (MD) scheme. Two structures are proposed for different use, namely low complexity simple MD transcoding structure (Simple MD) and adaptive MD transcoding structure (Adaptive MD). Simple MD structure always converts and splits the given bit stream into two descriptions regardless of channel condition. On the other hands, the Adaptive MD structure considers the channel condition, and it switches between the MD and single description (SD) mode depending on the packet loss rate (PLR). In the switching process, the best mode which produces less end-to-end cost is chosen between SD and MD under given channel condition. The expected end-to-end cost is optimally estimated based on the rate-distortion theory at the time of encoding. The simulation results show that the Simple MD structure provides efficient performance than the conventional error resilient transcoding structures based on the spatial and temporal error localization. And the Adaptive MD structure shows higher PSNR than the conventional methods under the comparable complexity.

1. INTRODUCTION

So far, low computational complexity and efficient structures have been the main issues in developing video transcoding systems. However, in the error prone channel, the error resilience in the transcoded video stream should be given more emphasis than in the conventional video stream. For this reason, several error resilient video transcoding structures have been proposed. The main purpose of these algorithms is to prevent the propagation of errors, and they use the methods such as the spatial and temporal error localization with intra block update [1], SNR scalability with unequal error protection [2], error resilient entropy coding (EREC) [3] and adaptive intra update (AIR) and reference frame selection (RFS) with feedback control signaling (FCS) [4].

Recently, error resilient transcoding structure which utilized multiple description (MD) scheme by forward error correction (FEC) was proposed [5]. Video stream encoded by scalable video encoder is packetized into N description using Reed-Solomon channel code. Each description has different error protection determined by optimization algorithm. This paper showed the MD scheme as an effective tool for packet loss resilience.

In this paper, based on this MD scheme, we propose two error resilient transcoding structures, *i.e.*, Simple MD transcoding structure and Adaptive MD transcoding structure. Although, in work of [5], transcoding structure based on the MD schemes was already examined, it was suitable for scalable video codec, 3D-SPIHT, not for motion compensated prediction structure. In addition, FEC-based MD transcoding structure only utilizes channel coding features without considering source coding aspects. In order to investigate a usefulness of MD scheme in the standard compliant transcoding structures, two effective transcoding structures for packet loss resilience are discussed in this paper.

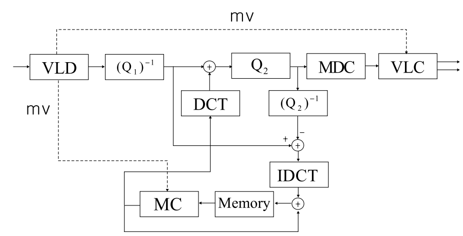

Fig. 1. Simple MD transcoding structure with reduced motion compensation loop

2. SIMPLE MD VIDEO TRANSCODING STRUCTURE

In this section, we propose a robust video transcoding structure using MD scheme, which has less computational complexity as well as higher ability to cope with severe error conditions than the single description (SD) transcoders. Fig. 1 shows a MD robust video transcoding structure for rate reduction, which is straight forward combination of efficient video transcoding structure, reduced motion compensation loop, and

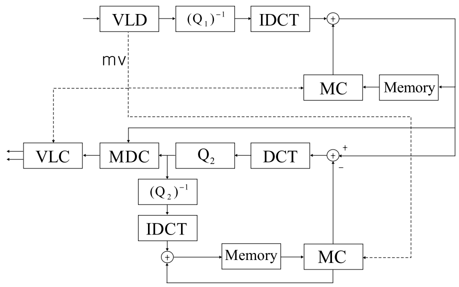

Fig. 2. Adaptive MD transcoding structure with motion vector reuse

effective error resilient tool, MDC.

The MD scheme is directly applied to the transcoded DCT coefficients which are obtained from rate reduction transcoding structure. The MD encoder splits an incoming stream into several descriptions (typically two). Some coefficients are duplicated to all descriptions, and the others are alternated into only one description considering the channel environment. More specifically, the DCT coefficients up to a certain index (from 0 to 63 in zig-zag order) are duplicated to every stream and the rest are alternated to each stream. So, we need to define the "start index" from which the alternation of coefficient begins. When the PLR is high, the stream should be made more robust by having more duplicated coefficients (large "start index"), because the lower band coefficients are generally more important. Hence, the start index is defined to be related with PLR as

$$\text{start index} = a \log_{10}(PLR) + b, \tag{1}$$

where $a = 64$ and $b = 128$, and is clipped to integer between 0 and 63. Although more complicated split algorithms based on the rate-distortion (RD) optimization can be applied to the transcoding scheme [6, 7], it seems impractical in transcoding applications due to high computational complexity. Simple MD achieves not only error robustness which is derived from MDC, but also efficient performance which is derived from low complexity transcoding structure by adding reasonable amount of computations.

3. ADAPTIVE MD VIDEO TRANSCODING STRUCTURE

Although the Simple MD proposed in the previous section has low complexity as well as robustness to errors, it can be further optimized by using SD/MD switching with additional computations. The basic idea of SD/MD switching is that SD is used when the PLR is low for reducing the redundancy, and MD when the PLR is high for the robustness to errors. Fig. 2 shows this SD/MD switching structure which is based on the motion vector reuse structure instead of the reduced motion compensation loop structure. We use the motion vector reuse structure because we need reconstructed pixel values for estimating distortion in the optimal SD/MD switching process.

Actual SD/MD switching is operated by rate-distortion sense, which means optimal mode that produces less rate-distortion costs is selected. Therefore, we need to find the criterion whether to use SD or MD according to the source characteristics and channel condition. In this paper, the optimal mode is selected by defining a cost function for the given mode and channel error as

$$mode^* = \min_{mode \in \{sd, md\}} J_{n,mode}. \tag{2}$$

Here, $J_{n,mode}$ is a Lagrange cost function defined as

$$J_{n,mode} = D_{n,mode} + \lambda R_{n,mode} \tag{3}$$

where λ is the Lagrange multiplier, $D_{n,mode}$ is the overall reconstruction distortion of the n-th frame, and $R_{n,mode}$ is the n-th frame's overall bit rate. The Lagrange multiplier selection is discussed in [8], where the best one is determined as

$$\lambda = c \cdot (QP)^2. \tag{4}$$

The theoretical justification is also given in the same paper, and the constant c is known to be 0.85 in the case of H.263 video encoder. More specifically, $D_{n,mode}$ and $R_{n,mode}$ is defined as

$$D_{n,mode} = \sum_{i=0}^{M-1} d^i_{n,mode}, \tag{5}$$

$$R_{n,mode} = \sum_{i=0}^{M-1} r^i_{n,mode} \tag{6}$$

where M is a total number of pixel within a frame, $d^i_{n,mode}$ is the i-th pixel's reconstruction distortion and $r^i_{n,mode}$ is the i-th pixel's bit rate.

In order to switch between SD and MD, it is necessary to estimate decoder distortion of SD transcoded stream, $d^i_{n,sd}$ and MD transcoded stream, $d^i_{n,md}$ at the time of encoding. For this reason, we use a Recursive Optimal per Pixel Estimation (ROPE) [9, 10]. Given the PLR, we estimate the decoder condition, especially the decoder reconstruction distortion. Considering the PLR, distortion d^i_n is defined as

$$\begin{aligned} d^i_n &= E\{(f^i_n - \tilde{f}^i_n)^2\} \\ &= (f^i_n)^2 - 2f^i_n E\{\tilde{f}^i_n\} + E\{(\tilde{f}^i_n)^2\} \end{aligned} \tag{7}$$

where f^i_n denotes the original value of pixel i in the n-th frame and $\tilde{f}^i_n$ is the reconstructed value at the decoder, possibly after the error concealment. For the encoder, $\tilde{f}^i_n$ is a random variable. This concept is first introduced for the SD video coding [9] and extended to MD video coding [10]. Also, ROPE with half pixel accuracy was proposed to estimate the decoder distortion at the encoder [11].

The ways to evaluate the first and second expectation are

different according to prediction modes (Intra or Inter) and split strategy (SD or MD). The general equations for SD and MD is already suggested [9, 10]. In this paper, simplified equations are given for duplication and alternation. These equations can reduce the overall complexity by removing multiplications. For duplication, $(\hat{f}_n^i) = (\hat{f}_n^i)_1 = (\hat{f}_n^i)_2$ for Intra and $(\hat{e}_n^i) = (\hat{e}_n^i)_1 = (\hat{e}_n^i)_2$ for Inter. If we plug these equalities into the equations of [10], we obtain simplified equations for duplication cases.
Intra mode :

$$E\{\tilde{f}_n^i\} = (1-p^2)(\hat{f}_n^i) + p^2 E\{\tilde{f}_{n-1}^i\},$$

$$E\{(\tilde{f}_n^i)^2\} = (1-p^2)(\hat{f}_n^i)^2 + p^2 E\{(\tilde{f}_{n-1}^i)^2\},$$

Inter mode :

$$E\{\tilde{f}_n^i\} = (1-p^2)\left(\hat{e}_n^i + E\{\tilde{f}_{n-1}^j\}\right) + p^2 E\{\tilde{f}_{n-1}^i\},$$

$$\begin{aligned} E\{(\tilde{f}_n^i)^2\} &= (1-p^2)\left((\hat{e}_n^i)^2 + 2\hat{e}_n^i E\{\tilde{f}_{n-1}^j\} + E\{(\tilde{f}_{n-1}^j)^2\}\right) \\ &\quad + p^2 E\{(\tilde{f}_{n-1}^i)^2\}. \end{aligned}$$

For the alternation, one of $(\hat{f}_n^i)_1$ and $(\hat{f}_n^i)_2$ is zero and another is same with $\hat{f}_n^i$. Also either $(\hat{e}_n^i)_1$ or $(\hat{e}_n^i)_2$ is zero and the other is same with $\hat{e}_n^i$. Modified equations for alternation cases are summarized as follows.
Intra mode :

$$E\{\tilde{f}_n^i\} = (1-p)(\hat{f}_n^i) + p^2 E\{\tilde{f}_{n-1}^i\},$$

$$E\{(\tilde{f}_n^i)^2\} = (1-p)(\hat{f}_n^i)^2 + p^2 E\{(\tilde{f}_{n-1}^i)^2\},$$

Inter mode :

$$E\{\tilde{f}_n^i\} = (1-p)\left(\hat{e}_n^i + E\{\tilde{f}_{n-1}^j\}\right) + p^2 E\{\tilde{f}_{n-1}^i\},$$

$$\begin{aligned} E\{(\tilde{f}_n^i)^2\} &= (1-p)\left((\hat{e}_n^i)^2 + 2\hat{e}_n^i E\{\tilde{f}_{n-1}^j\} + E\{(\tilde{f}_{n-1}^j)^2\}\right) \\ &\quad + p^2 E\{(\tilde{f}_{n-1}^i)^2\}. \end{aligned}$$

The transcoder with the switching scheme in this section is called Adaptive MD transcoder. This method requires more computations than the SD transcoder, which is less than three times of DCT per block that is needed for the computing the SD-ROPE and MD-ROPE [9]. Thus, the overall computational complexity of the Adaptive MD is comparable to that of SD transcoder. The overall computational complexity can be further reduced by simplified equations by utilizing duplication and alternation characteristics. As a result, Adaptive MD transcoder provides more efficient and robust performance regardless of channel conditions.

4. EXPERIMENTAL RESULTS

First, we compare the performance of SD transcoder, Simple MD and Adaptive MD transcoder under constant packet loss

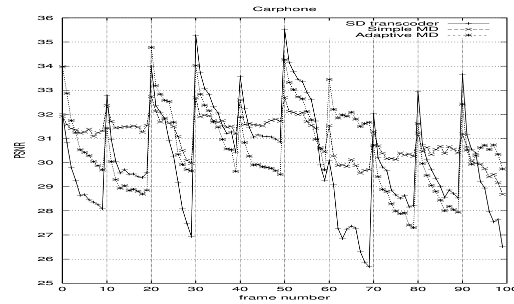

Fig. 3. Packet loss performance for Carphone sequence when PLR is 0.1. SD transcoder (QP = 8, 100.2 kbps), Simple MD (QP = 13, 96.8 kbps), Adaptive MD (QP = 9, 96.9 kbps)

environment. In order not to be influenced by the rate control and Lagrange multiplier selection method, we use fixed QP for the transcoding. All the transcoders receive the pre-encoded H.263 stream and generate another H.263 stream including the additional data for the error resilience. The SD transcoder puts more frequent sync marker for the spatial error localization, and the intra blocks are inserted randomly with the frequency reciprocally proportional to the PLR for temporal error localization. In the proposed MD schemes, each frame is split into two streams and intra frame update is performed every 10 frame.

After transcoding, each frame is split into two packets. In case of the SD transcoder, odd GOBs are assigned to one packet, and even GOBs to another. If a packet is lost, the decoder hides errors using motion compensated error concealment techniques. In case of MD transcoders, a split frame is packetized as one packet. For the channel setting, we assumed that every packet goes through the same channel, that is, same PLR is applied to each packet regardless of SD or MD schemes. Packet loss is generated by uniform random distribution. After transcoded packets are generated, we inject 30 different packet loss patterns for each sequence and average the results.

Fig 3 shows the result for Carphone (QCIF) sequence with 10 fps when PLR equals 0.1. In general, SD transcoder has dramatic quality degradation and severe error propagation. For example, the frames between 60-th and 70-th in Fig. 3 show the lowest PSNR among them. However, frames between 30-th and 40-th in Fig. 3 show higher PSNR than the other methods. This demonstrates that if the packet loss rate is small or error concealment is effective enough, SD transcoder attains better quality because it uses small QP, that is, large bit rate for those frames. In case of Simple MD, the fluctuation of PSNR curve is small. Even if packet loss is occurred, quality degradation is minimized. This is because the added redundancy is used for preventing error damage and propagation. This ability is easily seen by the frames between 0-th and 20-th in Fig. 3. Adaptive MD shows better quality than SD

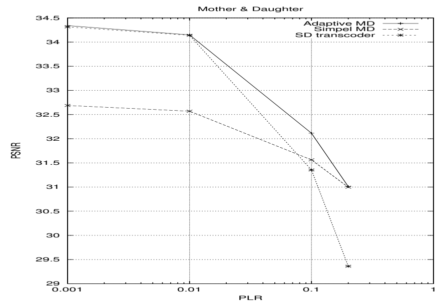

Fig. 4. Packet loss performance for M & D sequence

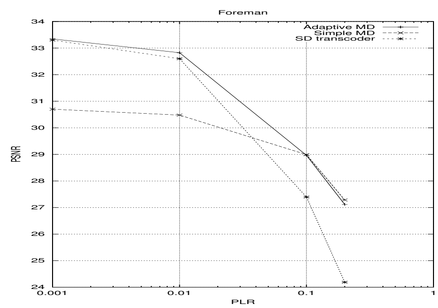

Fig. 5. Packet loss performance for Foreman sequence

transcoder and Simple MD between 60-th and 70-th frames in Fig. 3. In this interval, MD is selected for error resilience. As a result, Adaptive MD has better quality because of smaller QP than Simple MD and higher PSNR than SD transcoder because of MD's error robust property.

Next, we compared the performance of the transcoders with constant bit rate using same rate control method. In this experiment, we used 10 seconds of QCIF Mother & Daughter and Foreman sequences as input streams and generated transcoded streams using H.263 TMN-8 rate control algorithm. The target rates of each sequence are 64 kbps and 128 kbps, respectively. Fig. 4 and Fig. 5 show the packet loss performance for Mother & Daughter and Foreman sequences. These results show that the Simple MD transcoder scheme is more robust than the SD transcoder scheme at high PLR, whereas it shows lower PSNR at low PLR due to unavoidable redundancy. The Adaptive MD transcoder shows better performance than SD transcoder and Simple MD transcoder over the all PLR range. These results strongly support that the proposed MD scheme with SD/MD switching is effective in robust video transcoding.

5. CONCLUSIONS

In this paper, we have proposed robust video transcoding algorithms using multiple description schemes, namely Simple and Adaptive MD. The Simple MD splits a stream into two independent streams using simple split strategy. And Adaptive MD generates one or two descriptions considering the channel condition based on the optimal SD/MD switching. The experimental results show that proposed MD transcoding schemes are more robust than the error resilient SD transcoder scheme at high PLR rate. Moreover, the Adaptive MD shows better performance over all the PLR ranges, while it requires less computational complexity than the SD transcoder. The proposed robust transcoding structures can be easily applied to other video standard like H.264/MPEG-4 AVC.

6. REFERENCES

[1] G. Reyes, A. Reibman, S.-F. Chang, and J. C.-I Chuang, "Error-resilient transcoding for video over wireless channels," *IEEE Journ. on Sel. Areas in Commun.*, vol. 18, pp. 1063–1074, 2000.

[2] Y. Yu and Chang Wen Chen, "SNR scalable transcoding for video over wireless channes," in *Wireless Communications and Networking Conference*, Sept. 2000, pp. 1398–1402.

[3] L. Manus and A. Kokaram, "An improved error resilience scheme for highly compressed data," in *Wireless Communications and Networking Conference*, Mar. 2002, pp. 233–238.

[4] S. Dogan, A. Cellatoglu, M. Uyguroglu, A. Sadka, and A. Kondoz, "Error-resilient video transcoding for robust internetwork communications using gprs," *IEEE Trans. Circ. and Syst. Video Tech.*, vol. 12, pp. 453–464, 2002.

[5] R. Puri, K. W. Lee, K. Ramchandran, and V. Bhargavan, "An integrated source transcoding and congestion control paradigm for video streaming in the internet," *IEEE Trans. Multimedia*, vol. 3, pp. 18–32, Mar. 2001.

[6] A. Reibman, H. Jafarkhani, Y. Wang, and M. Orchard, "Multiple description video using rate-distortion splitting," in *Proc. IEEE Int. Conf. Image Proc.*, Oct. 2001, pp. 978–981.

[7] I. K. Kim and N. I. Cho, "Error resilient video coding using optimal multiple description of DCT coefficients," in *Proc. IEEE Int. Conf. Image Proc.*, Sept. 2003, pp. 663–666.

[8] T. Wiegand and B. Girod, "Lagrange multiplier selection in hybrid video coder control," in *Proc. IEEE Int. Conf. Image Proc.*, Oct. 2001, pp. 542–545.

[9] R. Zhang, S. Regunathan, and K. Rose, "Video coding with optimal Inter/Intra-mode switching for packet loss resilience," *IEEE Journ. on Sel. Areas in Commun.*, vol. 18, pp. 966–976, June 2000.

[10] A. Reibman, "Optimizing multiple description video coders in a packet loss environment," in *Proc. Int. Packet Video Workshop*, Mar. 2002.

[11] H. Yang and K. Rose, "Recursive end-to-end distortion estimation with model-based cross-correlation approximation," in *Proc. IEEE Int. Conf. Image Proc.*, Sept. 2003, pp. 469–472.