A NOVEL RESYNCHRONIZATION METHOD FOR SCALABLE VIDEO OVER WIRELESS CHANNEL

Yu Wang, Lap-Pui Chau, Kim-Hui Yap

School of Electrical and Electronic Engineering
Nanyang Technological University, Singapore, 639798

ABSTRACT

A scalable video coder generates scalable compressed bit-stream, which can provide different types of scalability depend on different requirements. This paper proposes a novel resynchronization method for the scalable video with combined temporal and quality (SNR) scalability. The main purpose is to improve the robustness of the transmitted video. In the proposed scheme, the video is encoded into scalable compressed bit-stream with combined temporal and quality scalability. The significance of each enhancement layer unit is estimated properly. A novel resynchronization method is proposed where joint group of picture (GOP) level and picture level insertion of resynchronization marker approach is applied to insert different amount of resynchronization markers in different enhancement layer units for reliable transmission of the video over error-prone channels. It is demonstrated from the experimental results that the proposed method can perform graceful degradation under a variety of error conditions and the improvement can be up to 1 dB compared with the conventional method.

1. INTRODUCTION

Recent advances in technology have led to an increasing interest in video services over wireless networks, which have different characteristics and do not guarantee a Quality of Service (QoS). In such noisy channels, bit error rate (BER) can be quite high. When compressed video bit-stream is sent over these channels, the effect of channel errors can be very severe. Transmission errors, together with lossy source coding techniques, lead to the distortion of the video sequences at the decoder. Therefore, it is demanded to develop robust video coding techniques to ensure the quality of the decoded video.

The robustness can be achieved by adopting error resilience techniques into the compressed bit-stream [1]. Provisions for error resilience have been included in previous video coding standards such as H.263 and MPEG-4. Resynchronization has been proven to be a very effective tool among the state-of-art error-resilient techniques.

Through the use of resynchronization markers, error can be localized and error propagation is limited. In H.263, resynchronization markers are inserted at certain position in the bit-stream such as the starting point of the group of block (GOB) [2]. This kind of synchronization may result in the problem that the resynchronization markers are not evenly distributed so that some regions of the frame are more prone to be affected by errors. In MPEG-4, the above problem is overcome and resynchronization markers are inserted with an approximately constant interval [3].

Insertion of resynchronization markers has been investigated in many literatures [4]-[7] besides the approaches mentioned above. Picture level resynchronization methods have been deeply studied by researchers [4][5]. The schemes in these works all aims at inserting resynchronization marker in the bit-stream of one picture. For example, in [4], Fang and Chau propose a content-based resynchronization method so that the areas of interest can achieve a better protection. Besides picture level resynchronization methods, GOP level approach has also been explored in [6], where different sizes of slice are assigned for different pictures in a GOP considering the type and the order of the picture. In addition, the insertion of resynchronization markers has also been applied for the scalable video. In [7], Yan, *et al.*, propose to include error-resilience tools into the fine granularity scalability (FGS) enhancement layer bit-stream. A hierarchical enhancement layer bit-stream structure is designed with resynchronization markers and Header Extension Code (HEC).

The new developed scalable video coding (SVC) [8] provides a full scalability including spatial, temporal and quality (SNR) scalability with fine granularity. In this paper, for the scalable encoded video with combined temporal and quality scalability, we introduce a joint GOP level and picture level resynchronization method. A GOP is decomposed into pictures with different temporal levels and each picture is encoded into quality base layer and enhancement layers. Our scheme is to efficiently insert resynchronization markers to each unit in the enhancement layer considering both the significance of the unit and the channel condition to perform a graceful degradation of the video over error-prone channels.
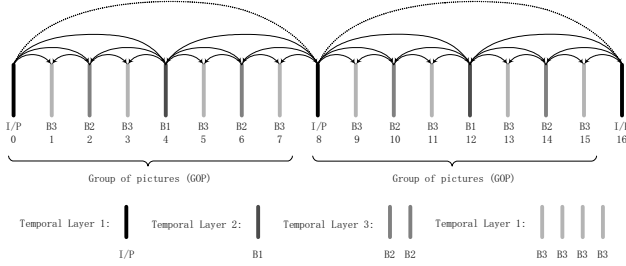
Figure 1. Structure of hierarchical B pictures and generation of temporal layers.
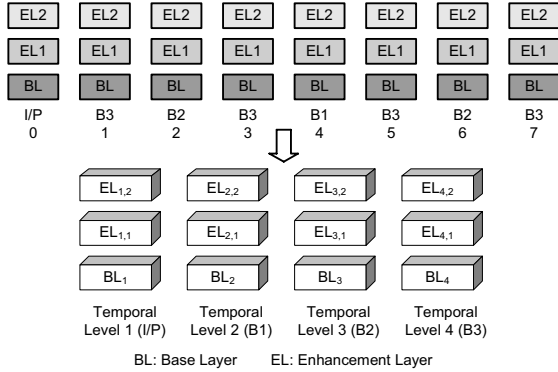

BL: Base Layer   EL: Enhancement Layer

Figure 2. Generation of quality (SNR) layers and illustration of scalable units in a GOP.

The rest of the paper is organized as follows. In Section 2, an overview of the proposed scheme is presented. The optimization problem is formulated and properly solved. The experimental results are given in Section 3 and conclusions are drawn in Section 4.

2. PROPOSED SCHEME

In this section, the structure of the proposed scheme is firstly described. After that, a method is introduced to measure the significance of different enhancement layer units. Finally the optimization problem is formulated and properly solved.

2.1. System Overview

In scalable video coding, temporal scalability is achieved using hierarchical B pictures [8]. A typical hierarchical coding structure is depicted in Fig. 1.

In our scheme, the whole video sequence is divided into groups of pictures (GOP) with fixed length m ($m=2^n$). Through dyadic decomposition, each GOP is decomposed into n+1 temporal layers. In Fig. 1, the GOP size is 8. Therefore 4 temporal layers are achieved and the pictures of the same type (I/P, B1, B2 or B3) are grouped into different temporal layers. In Fig. 1, different temporal levels are illustrated using different colors.

To attain quality (SNR) scalability [8], texture of each picture is encoded to produce a quality (SNR) base layer,

which provides a minimum quality at a given quantization parameter (QP). The quality enhancement layers are generated by repeatedly decreasing the value of QP and encoding the refinements from the base layer. Within each enhancement layer, a modified entropy coding process is applied to achieve fine grain scalability. This process is illustrated in Fig. 2, where one base layer cell and two enhancement layer cells are generated for each picture in a GOP. The cells with the same quality level and from the same type of pictures are organized together to form a unit. Therefore, a set of units is generated from a GOP. In Fig. 2, $EL_{i,j}$ represents the Enhancement Layer (EL) unit with temporal level i and quality level j and BL_i denotes the Base Layer (BL) unit with temporal level i in the quality base layer.

In the proposed scheme, we assume that an error-free transmission of the base layer information can be guaranteed. To efficiently insert resynchronization markers in the EL units within a GOP, some factors need to be considered. First is the channel condition. With more resynchronization markers, the error resilient property of the video stream will improve while the coding efficiency will drop. We assume that under different channel conditions, there must be a tradeoff between the bits used for source coding and those used for resynchronization markers. Second is the importance of different EL units. More resynchronization markers should be inserted in the more important EL units to make the video stream more robust to channel errors. The measurement of importance of each EL unit will be described in the following Subsection.

2.2. Measurement of Significance of Different Enhancement Layer Units

EL units with different temporal or quality levels are not equally important. To estimate the importance of a unit, two factors are taken into consideration in our scheme.

The first factor is utility, which is the impairment caused when the slices belonging to the unit are lost. Here the impairment is defined as decreased PSNR of the decoded video compared to the original one due to the loss of the information. We use $U_{i,j}$ to denote the utility of the unit with temporal level i and quality level j.

The second factor is cost, which is the amount of source bits in the unit. $B_{i,j}$ represents the cost of a certain unit.

The utility-cost ratio $R_{i,j}$ is defined to measure the importance of each unit. It is calculated as

$$R_{i,j} = U_{i,j} / B_{i,j} \qquad (1)$$

The utility-cost ratio is computed for each unit. The unit with a larger utility-cost ratio is assumed to be more important in our scheme. Through this way, the significance of each EL unit is properly determined.

2.3. Analysis of the Optimization Problem

In this Subsection, we will discuss the problem of insertion of resynchronization markers to different EL units in a GOP to optimize the quality of the decoded video. It should subject to an overall target bit rate R_{Budget}.

We assume that there are T temporal layers in one GOP and each picture is encoded into a base layer and Q quality enhancement layers. Thus there are totally $T \times Q$ EL units to be transmitted. Given the overall coding rate, we want to efficiently insert resynchronization markers such that the quality of the decoded video is maximized. It also means that the total distortion $D_{overall}$ is minimized. The problem is formulated as

$$Min\ D_{overall}, \text{ subject to } R_{overall} \leq R_{Budget} \tag{2}$$

The overall bit rate $R_{overall}$ is defined as

$$R_{overall} = R_S + R_{RM} \tag{3}$$

with R_S being the source rate and R_{RM} being the rate used up by resynchronization markers.

The overall expected distortion is defined as

$$D_{overall} = D_S + D_{RM} \tag{4}$$

The total distortion $D_{overall}$ consists of two parts. The first part D_S is due to the loss of source bits caused by channel errors during transmission. The second part D_{RM} is the distortion resulted from insertion of resynchronization markers. The more bits consumed by resynchronization markers, the larger the value of the distortion D_{RM}.

D_S is calculated as follows

$$D_S = \sum_{i=1}^{T} \sum_{j=1}^{Q} \sum_{l=1}^{k_{i,j}} \delta_{i,j,l} P_{i,j,l} \tag{5}$$

In equation (5), i denotes the temporal level and j denotes the quality level with $i = 1,2,\ldots,T$ and $j = 1,2,\ldots,Q$. l is the number of slices lost in the enhancement layer unit with temporal level i and quality level j ($ELU_{i,j}$) and $k_{i,j}$ is the total amount of resynchronization markers in the $ELU_{i,j}$. $\delta_{i,j,l}$ represents the decreased PSNR due to loss of l slices in $ELU_{i,j}$. $\delta_{i,j,l}$ can be obtained during encoding by calculating the difference between the decoded video and the original one. $P_{i,j,l}$ is the probability that l slices are lost in the specified unit. The calculation of $P_{i,j,l}$ is shown below:

$$P_{i,j,l} = \binom{k_{i,j}}{l} SER_{i,j}^{l} (1 - SER_{i,j})^{k_{i,j}-l} \tag{6}$$

where $SER_{i,j}$ is the slice error rate, which is the probability that one slice in $ELU_{i,j}$ will be lost. $SER_{i,j}$ can be calculated as

$$SER_{i,j} = 1 - (1 - BER)^{b_{i,j}} \tag{7}$$

with BER being the bit error rate and $b_{i,j}$ being the average number of bits in each slice of $ELU_{i,j}$.

To calculate the second part distortion D_{RM}, it should be noticed that during inserting of resynchronization markers, the same amount of source bits is required to be removed to satisfy the bandwidth requirement. This portion of source bits is taken from the least important enhancement layer unit and will not be transmitted. The measurement of the importance of different units in a GOP has been described in Subsection 2.2. Through this method, The unit with the smallest utility-cost ratio will be selected as the least important unit denoted as ELU^*. Thus the distortion due to insertion of resynchronization markers can be calculated as the distortion caused by loss of information in the specified unit, that is

$$D_{RM} = U^* \times \frac{m \sum_{i=1}^{T} \sum_{j=1}^{Q} k_{i,j}}{B^*} \tag{8}$$

where U^* and B^* are the utility and the cost of ELU^*, respectively. m is the number of bits consumed by one resynchronization marker.

Till now, the optimization problem can be expressed as

$$Min\ D_{overall}(\boldsymbol{K}), \text{ subject to } R_{overall} \leq R_{Budget} \tag{9}$$

with

$$\boldsymbol{K} = \begin{bmatrix} k_{1,1} & k_{1,2} & \cdots & k_{1,Q} \\ k_{2,1} & k_{2,2} & \cdots & k_{2,Q} \\ \vdots & \vdots & \vdots & \vdots \\ k_{T,1} & k_{T,2} & \cdots & k_{T,Q} \end{bmatrix}$$

The problem is deduced to find $\boldsymbol{K}_{opt}$,

$$\boldsymbol{K}_{opt} = \arg\min D_{overall}(\boldsymbol{K}) \tag{10}$$

There are two key points in the optimization problem. One is the trade-off between the bits for source coding and those for resynchronization markers. The other is how to distribute the resynchronization markers in different units. Exhaustive searching can be applied to solve the above problem. However it is unfeasible in reality because of the large amount of computation. In this paper, we carried out the hill-climbing algorithm to quickly solve the optimization problem and a detailed description about the algorithm can be found in [9]. During applying the algorithm, a constraint needs to be satisfied. That is, the slice size of the more important unit should be smaller than that of the less important one. Through initializing the number of resynchronization markers in each unit as 1 and searching in a pre-defined range, the constrained minimization problem can be properly solved and the result is satisfied with tolerable computation complexity.
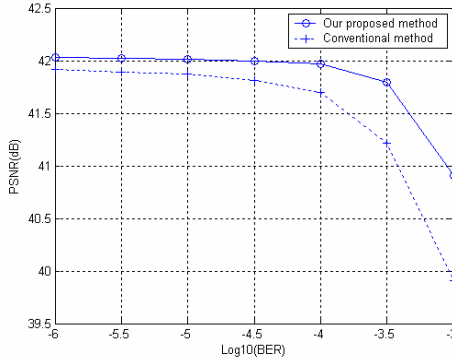
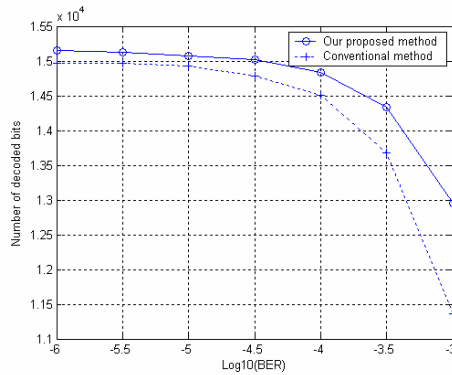Figure 3. Average PSNR of "Foreman" under different BER.

Figure 4. Average number of decoded enhancement layer bits of "Foreman" under different BER.

3. SIMULATION RESULTS

In the experiment, we tested four sequences: "Foreman", "City", "Football" and "Crew". The frame rate is 15 Hz and the spatial resolution is QCIF. They were encoded by scalable video codec. Because of the stringent space, we only use "Foreman" to show the experimental results.

The sequence length is 144 and the GOP size is defined as 16 so that five temporal layers are generated. Each picture is encoded into one quality base layer and two enhancement layers where the base layer is encoded using QP30. Hill-climbing method is applied to efficiently insert re-synchronization markers to different enhancement layer units. For comparison, conventional method [3] is also implemented, where the resynchronization markers are inserted into the bit-stream with an approximately constant interval.

Experiments were performed to transmit video sequences under a wide range of BER. 50 different runs were conducted under each BER. The performance of our proposed method is compared with the conventional method over a variety of error conditions. The results are illustrated in Fig. 3 and Fig. 4. The figures show the average PSNR of the video sequence and the average amount of decoded bits of the enhancement layer bit-stream under different BER, respectively. In contrast, our proposed scheme exhibits superiority over the conventional method. The value of the average PSNR is improved up to 1 dB.

4. CONCLUSIONS

In this paper, a joint GOP level and picture level resynchronization method is proposed for the scalable video with combined scalability. The scheme is based on the structure of the joint scalable video model (JSVM), where combined temporal and quality scalability can be easily achieved. Considering the significance of different enhancement layer units and the time-varying channel condition, resynchronization markers are efficiently inserted. Through this way, the video exhibits robustness against the transmission errors and performs a graceful degradation over error-prone channels. The experimental results demonstrate that our proposed method is superior to the conventional method under various error conditions and the improvement is up to 1dB.

5. REFERENCES

[1] Y. Wang, S. Wenger, J. Wen and A. K. Katsaggelos, "Error resilient video coding techniques," *IEEE Signal Processing Magazine,* vol.17, no.4, pp. 61-82, Jul 2000.

[2] ITU-T Recommendation H.263 Version 2, "Video coding for low bitrate communication," Jan. 1998.

[3] "Coding of audio-visual objects: visual, final draft international standard," ISO/IEC 14496-2, ISO/IEC JTC1/SC29/WG11 N2502, Oct. 1998.

[4] T. Fang, L.-P. Chau, "Content-based resynchronization for robust video transmission," *IEEE trans. Broadcasting*, vol.50, no.4, pp. 390-395, Dec. 2004.

[5] J.-H. Jeong, H.-S. Kang and J.-K. Kim, "Optimal resynchronization marker positioning method using a novel information measure," *Signal Processing: Image Communication* 17, pp. 799-806, 2002.

[6] L. O.-Barbosa and T. Han, "On the use of frame-based slice size for the robust transmission of MPEG video over ATM networks," *IEEE Trans. Broadcasting*, vol. 46, no.2, pp. 134-143, Jun. 2000.

[7] R. Yan, F. Wu, S. Li and R. Tao, "Error resilience methods for FGS video enhancement bitstream," *IEEE PCM*, Dec. 2000.

[8] "Joint scalable video model (JSVM) 4.0 reference encoding algorithm description," ISO/IEC JTC 1/ SC 29/ WG 11 N7556, Nice, France, Oct. 2005.

[9] E. Mohr, E. A. Riskin and R. E. Ladner, "Unequal loss protection: graceful degradation of image quality over packet erasure channels through forward error correction," *IEEE J. Select. Areas Comm.*, vol.18, no. 6, pp. 819-828, Jun. 2000.