REAL-TIME MPEG VIDEO CODING WITH INFORMATION LOOK-AHEAD

Ligang Lu and Vadim Sheinin
Multimedia Technologies,
IBM T. J. Watson Research Center,
1101 Kitchawan Road/Route 134, Yorktown Heights, NY 10598
Email: lul@us.ibm.com

ABSTRACT

There have been increasing market demands to develop more effective schemes for applying MPEG video coding standard to meet the challenges from the emerging applications. We present a real-time MPEG video coding system with information look-ahead for constant bit rate (CBR) applications, such as Video-on-Demand (VoD) over ADSL. This scheme employs two MPEG encoders. Both encoders operate at the same CBR. The second encoder has a buffer to delay the input by an amount of time relative to the first encoder to create a look-ahead window. In encoding, the first encoder collects the information of statistics and rate-quality characteristics. An on-line information processor then uses the collected information to derive the best coding strategy for the second encoder to encode the incoming frames in the look-ahead window. The second encoder uses the encoding parameters from the processor as the coding guide to execute the coding strategy and generate the final bitstream. Results on an IBM MPEG-2 encoder testing system with a 15 frame look-ahead window have shown that the scheme can obtain a 0.6 to 1.6dB improvement in PSNR over the conventional single encoder on MPEG test video sources. Moreover, measured by the Tektronix Inc.'s Picture Quality Analysis (PQA) System, this scheme achieved 0.87 ~ 1.35 reduction in PQA measurement. The visual quality improvement is also very evident.

1. INTRODUCTION

MPEG-2 video coding standard has been successfully used to reduce the transmission bandwidth and storage space requirements in many applications, such as digital satellite TV and cable TV broadcasts, DVD, etc. However, emerging applications and services are becoming more demanding for lower transmission bandwidth and smaller storage space. For example, in Video-on-Demand (VoD) service over ADSL, live news and sports events need to be transmitted in real-time to the subscribers using MPEG-2 video coding [1] at a constant bit rate (CBR) in the range of 0.6 to 2Mbits/second. Using MPEG CBR coding at such bit rate, it is very challenging for the conventional MPEG encoders available on the market to achieve acceptable picture quality. Conventional MPEG encoders employ a single encoder scheme as depicted in Figure 1. These encoders routinely adopt a coding strategy that is only based on the information obtained in coding the previous video frames and/or relied on some assumed signal models to predict or estimate the signal properties of the current input frame and to guide the encoding process of the current frame. However, natural video is a statistically non-stationary signal source. Prediction and estimation based on the past signal often cannot correctly describe the current input signal and there is no known robust signal model that can describe the natural video signal reliably. Such encoders will not be able to determine and apply the best coding strategy to encode the incoming video frames for lack of the information about the current and future input frames. In order to meet the challenges from more demanding emerging applications, more sophisticated schemes for MPEG-2 video coding are needed to improve the performance and to ensure the quality of services. In the past research efforts have been made to improve the variable bit rate (VBR) MPEG video coding for DVD application by employing two-pass and re-encoding schemes [2-4]. However there are no published research results for multi-pass CBR coding in the literature. In this paper, we present a real-time MPEG CBR video coding scheme that is able to jointly determine and apply the best coding plan to encode input video frames based on not only the information of the previous and current frames but also the information about the future input frames. In the next section we will first introduce the concept and essence of this scheme. We will then describe the algorithm and implementation issue of our scheme in details in Section 3. We will present the performance results of our scheme in Section 4. Finally in the last section we summarize and conclude our paper.

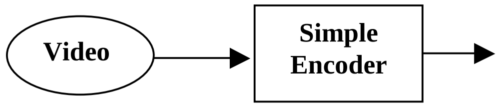

Figure 1 Conventional Video Encoder

2. VIDEO CODING WITH INFORMATION LOOK-AHEAD

As we have mentioned above, the conventional video encoder uses a coding strategy which is devised based on the information retained in coding the previous video frames and such coding strategy clearly does not fit well with the non-stationary nature of the video signal. To overcome this shortcoming and achieve better performance, we add an information look-ahead mechanism in the encoder.

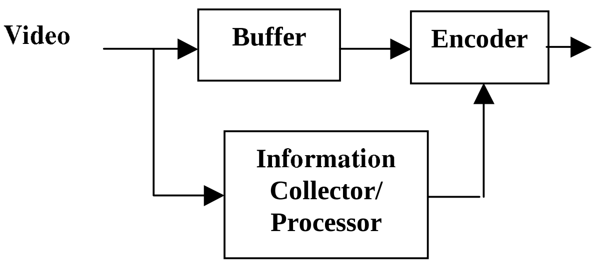

Figure 2. Video Coding with Information Look-Ahead

As shown in Figure 2, the video input is fed in parallel to a buffer in front of the encoder and an information collector/processor. The buffer delays the input video frames by a fixed amount of time so that the information collector/processor can have the operation time to extract useful information about the incoming frames in the delay buffer and process the information to determine a coding strategy for encoding these frames and pass the coding strategy in the form of the coding parameters to the encoder for execution. In this video coding scheme, the buffer effectively creates a look-ahead time window for the information collector/processor to gather and process the information. Given the information and the processing algorithm, the best coding plan can be determined jointly based on the information about the past, current, and future input video frames. However, to implement this video coding scheme in a cost effective way and to achieve efficient performance, issues of what appropriate information to collect and how to collect, how the information should be used to devise the coding plan (i.e., the processing algorithm), what is the proper buffer size (or the look-ahead window size) need to be determined. In the next section, we will address these crucial issues.

3. REAL-TIME MPEG VIDEO CODING WITH INFORMATION LOOK-AHEAD: THE SYSTEM IMPLEMENTATION AND ALGORITHM

The system diagram of our video coding scheme is depicted in Figure 3. Obviously there are several key issues that will have significant impacts on the performance and the cost of the system. These issues are what and how the look-ahead information should be collected, how the information should or can be used to devise the best coding strategy, and how the coding strategy can be carried out. In the following subsections we discuss these issues in details.

3.1. Look-Ahead Information Collection

Clearly the most useful information for determining the best coding strategy for the incoming video frames are the signal statistics and characteristic variables, rate-quality measure, and coding parameters that are directly used in various steps of the encoding process with the dominant impact in the coding results. The most effective approach to collect such information is to use a collector that emulates the encoder operation. Therefore in order to gather the most pertinent and useful information to derive the best coding strategy, we employ another MPEG encoder, Encoder 1, as the information collector in our scheme as shown in Figure 3. The main benefit of using an encoder as the information collector is that the direct signal information and intermediate results in various encoding stages can be obtained in the same encoding operation conditions as the intended encoding process.

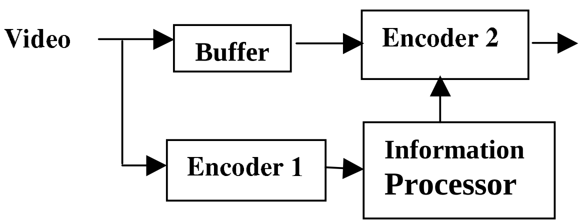

Figure 3. Real-Time MPEG Video Coding with Information Look-Ahead

The exact items of information to be collected depend on the need of the processing algorithm and the availability of the information in the encoder chip on the fly and the real-time output capability of the encoder. In our system, we use the IBM MPEG-2 encoder chip and collect the information regarding to the picture complexity, motion magnitude, and picture quality index.

The length of the look-ahead window determines the input delay buffer size. The more frames to look-ahead, the larger the buffer size and in turn the longer the delay. The cost also increases with the buffer size. For the convenience of bit allocation and rate control in CBR coding, we make the look-ahead window size as the multiple of the size of Group of Pictures (GOP) [1] so that the numbers of Intra-coded (I), Predictive Coded (P), and Bi-direction Predictive Coded (B) pictures [1] in the look-ahead window are constants. After some experiments, we determine the look-ahead window size W_s as

$$W_s = K * GOP_s \qquad (1)$$

where K = 1 or 2; GOP_S is the size of Group of Picture in MPEG video coding. The input delay buffer size is then

$$B_s = W_s + \Delta_p \qquad (2)$$

where Δ_p is the information processing time which depends on the complexity of the algorithm.

3.2. Information Processing

Once the information about the video frames in the look-ahead window is available, the processing algorithm will determine a coding strategy for these frames using the information. In our scheme, we jointly determine a target bit allocation plan for the video frames so that the available bits can be used efficiently and the decoding buffer defined as Virtual Buffer Verifier (VBV) in MPEG-2 standard [1] can be exploited sufficiently. Assume there are N frames in the look-ahead window. Let P_i, i=1, … N, be the i-*th* frame in the window. Denote the picture complexity, motion magnitude, picture quality index, and target number of bits for P_i as C_i, M_i, Q_i and T_i, respectively. Let R be the bit rate and F the frame rate. Our algorithm performs the following steps:

Step 1. Calculate the dynamic weighted picture complexity:

$$C_i^{'} = C_i W(\frac{M_i}{\overline{M}_l}, \frac{Q_i}{\overline{Q}_l}, S_i, D_i), \qquad (3)$$

where $W(.)$ is a real function; $S_i \in \{I, P, B\}$ is the picture coding type of P_i; $\overline{M}_l$ and $\overline{Q}_l$ are the average motion magnitude and average picture quality index of all frames in the look-ahead window with the same picture coding type as S_i; and D_i is the distance from P_i to the most recent I frame. Note that in our measurement, the larger the value of Q_i, the worse the picture quality.

Step 2. Jointly determine the target number of bits for all frames in the look-ahead window:

$$T_i = \frac{RNC_i^{'}}{F \sum_{k=1}^{N} C_k^{'}}; \qquad (4)$$

Step 3. Rate control to prevent decoder buffer overflow and underflow: Denote V the decoder buffer size (1835008 bits for MP@ML case) as defined in MPEG-2 standard [1] and V_i the decoder buffer fullness just before the picture P_i is drawn from the decoder buffer for decoding. Let G be a guard band, usually G = 3~5% of V. The MPEG-2 decoder buffer model for CBR operation can be described by the recurrence

$$V_0 = V_{init}, \qquad V_i = V_{i-1} + R/F - T_i, \qquad (5)$$

where V_{init} is the initial buffer fullness. To prevent any overflow and underflow, the buffer fullness must always satisfy

$$T_i + G \le V_i \le V - G. \qquad (6)$$

If V_i underflow or overflow the buffer requirement by an amount of δ, then adjust the target bit allocation as

$$T_k^{'} = T_k \frac{\sum_{m=1}^{i} T_m \mp \delta}{\sum_{m=1}^{i} T_m}, \qquad k = 1, 2, \ldots, i;$$
$$T_k^{'} = T_k \frac{\sum_{n=i+1}^{N} T_n \pm \delta}{\sum_{n=i+1}^{N} T_n}, \qquad k = i+1, \ldots, N. \qquad (7)$$

The processing algorithm can either be implemented in a general processor or a DSP chip or resides on the host PC. We use the host PC to perform the information processing function in our implementation.

3.3. Execution of the Coding Strategy

After the coding strategy is determined for the video frames in the look-ahead window, it needs to be passed to Encoder 2, in the form of the coding parameters and necessary information to carry out. In practice, the real-time communication bandwidth between the encoder and information processing may limit the amount of the coding parameters and information to be transmitted to the encoder on frame-by-frame basis and may have an impact on the execution of the coding strategy. In our implementation, we use 4 16bit integers to pass parameters to Encoder 2 for every frame's encoding. They are 16 bits for the target number of bits, 16bits for the weighted picture complexity, and 32 bits for the sum of the weighted complexities of the remaining un-encoded frames in the look-ahead window. The last two parameters are used by Encoder 2 to reallocate any excess bits over the remaining frames.

4. RESULTS

Our scheme is tested with a 15 frame look-ahead window. We have compared it with the conventional single MPEG encoder scheme. Table 1 compares the performances of the two schemes measured by Tektronix's video Picture Quality Assessment (PQA) system. The measurements were done by a third party. The two schemes used same encoders but one employed the information look-ahead scheme described above. Our proposed scheme reduced the PQA in value by 0.87 to 1.31. In industry, a 0.5 reduction in PQA value is considered significant. Table 2 compares the performances of the two schemes in PSNR measurements. Again our scheme achieved a gain of 0.60~1.63 dB in the average PSNR. Examples of PSNR plots are also shown in Figures 4 and 5. Visual picture quality improvement is also very significant.

Table 1. Picture Quality Assessment (PQA) Comparison

720x480 Video Coded @2Mbits/s.	Cheer Leader+	Football +	Mobile Calendar+
Convention Encoder	12.81	12.02	9.94
Proposed Scheme	11.60	10.71	9.07

Table 2. Peak Signal to Noise (PSNR) Comparison

720x480 video @2Mbits/s.	Cheer Leader+	Football +	Mobile Calendar +	Basket Ball
Convention Encoder	27.15	31.34	26.99	27.49
Proposed Scheme	27.85	31.94	27.91	29.12

5. CONLUSIONS

We presented a scheme for MPEG CBR video coding with information look-ahead. Our scheme uses two MPEG encoders. The first encoder acts as an information collector. The information is then used by an on-line processor. Taking the advantage of the time delay between the inputs of the two encoders, the processor employs an efficient algorithm to jointly derive the best coding strategy for the all incoming frames in the look-ahead window by exploiting the information not only about the past and current frames but also the future frames. The second encoder, which operates at the same constant bit rate as the first encoder, uses the coding strategy from the processor as the guide to encode the incoming frames. This scheme achieved significant both objective and visual picture quality improvement over the conventional single encoder scheme and can be used in many emerging applications.

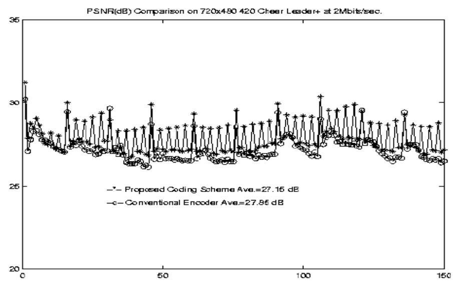

Figure 4 PSNR Plot on Cheer Leader+ at 2 Mbits/sec.

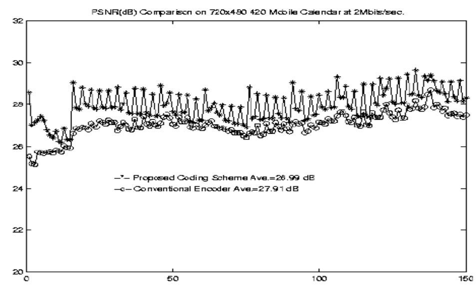

Figure 5 PSNR Plot on Mobile Calendar+ at 2Mbits/sec.

REFERENCES

[1] *Informaiton Technology*—Generic Coding of Moving Pictures and Associated Audio:Video, ISO/IEC 13818-2, 1995.

[2] Y. Yu, J. Zhou, Y. Wang, and C. W. Chen, "A Novel Two-Pass VBR Coding Algorithm for Fixed-Size Storage Application," *IEEE Trans. On CSVT*, Vol. 11, No. 3, pp. 345-356, March 2001.

[3] L. Lu, et al. "Segment Reencoding of Buffer Constrained Variable Bit Rate Video Streams," *IEEE ICME*, July 2000.

[4] P.H.Westerink, et al. "Two-pass MPEG-2 Variable-bit-rate encoding", IBM J. of R&D, V. 43, No.4, pp. 471-488, July 1999.