

# Low Delay Rate-Control in Video Transcoding

*Yu Sun, Xiaohui Wei and Ishfaq Ahmad*

Department of Computer Science & Engineering

University of Texas at Arlington

Arlington, TEXAS 76019, USA

## ABSTRACT

In order to obtain optimal visual quality while meeting the bit rate requirements, a video transcoder needs efficient rate-control with fine and dynamic adjustments. This paper proposes a new frequency domain complexity estimation scheme and an algorithm for adaptive bit allocation during transcoding. In conjunction with frame skipping mechanism, the proposed algorithm adaptively determines spatial coding parameters to realize very low target bitrate for MPEG-2 to MPEG-4 transcoding. Experiments show that the proposed algorithm performs accurate target bit rate transcoding, improves the visual quality and effectively alleviates visual quality degradation during scene changes.

## 1. INTRODUCTION

As the growth of the Internet leads to an increased number of networked multimedia applications, inter-compatibility among different systems is becoming an important issue. The heterogeneity of the current communication networks, user display-devices, and user preferences of quality demand the matching of the bit-rates and formats of the pre-encoded video to the channel constraints and application requirements [1]. Video transcoding is thus a core technology for universal multimedia access by Internet users with different access links and devices. Various transcoding techniques have been studied to convert a previously compressed video bit-stream to a different compression format, or a different bit-rate [1]–[5].

The goal of rate control in video coding is to achieve a target bit-rate with good and consistent visual quality [1]. Since the outgoing bit stream may need to be transcoded in order to fit into CBR channels, rate control is important to achieve optimal video quality in transcoding of one or multiple video streams. A conventional transcoder uses requantization methods so as to get target bit-rate, for example MPEG-2 TM5 [2], [3]. Researchers [4] agreed that TM5 method was not suitable to control the bit-rate in transcoding since the inversely proportional relationship between quantizer parameter (QP) and generated bit-rate might not be suitable in transcoding. Tudor [5] computed the complexities of pictures defined in TM5 from the input bit-stream, and then allocated target bits based on these complexities and re-coded these pictures in the transcoding process. However, Xin and Sun [1] pointed out that the TM5 complexity measures calculated from the input video bit-stream at the input bit-rate might not be suitable to directly serve as the complexities for coding pictures at the output bit-rate. According to the correlations between the input and output complexity measures, they estimated the picture complexities of the output transcoded video from the statistics extracted from the input video bit-stream. Then based on the estimated complexity,

they performed bit allocation for rate control in their MPEG-2 transcoder.

The above rate control schemes have been developed based on homogeneous MPEG-2 encoders. For a heterogeneous MPEG-2 to MPEG-4 transcoder, the conventional TM5 [6] or VM18 [7] rate control leads to several problems. First, TM5 is not suitable for controlling bit rates in transcoding a MPEG-2 bitstream to a very low bit rate MPEG-4 stream. TM5 cannot change temporal coding parameters, and cannot properly deal with scene changes. In addition, since TM5 rate control is based on MB level, and thus its computational complexity is high for a real-time transcoder. Second, MPEG-4 VM18 adopts a quadratic rate-quantizer model to describe the relation between the required bits for coding the texture and the quantization parameter. The Mean Absolute Difference (MAD) is introduced in this model to estimate the target bit allocation more accurately, it is computed after the motion compensation for the luminance component. But it may be difficult to obtain MAD in video transcoding, since motion estimation (ME) and motion compensation (MC) are normally omitted to reduce transcoding complexity and most transcoding processing is performed in frequency domain. Thus VM18 rate control is not suitable to be applied in transcoding.

In this paper, we present a simple, yet effective, rate control scheme for MPEG-2 to MPEG-4 transcoding. According to the transcoding properties, we define a DCT-based complexity and propose a new bit allocation method. Our algorithm adaptively determines spatial and temporal coding parameters to realize very low target bit rate transcoding. In addition, we adopt a robust buffer control strategy to effectively minimize the buffer overflow or underflow.

This paper is organized as follows: Section 2 proposes a DCT-based complexity estimation. We describe the basic framework of our rate-control algorithm for MPEG-2 to MPEG-4 transcoding in section 3 and present experimental results and discussions in section 4. Finally, Section 5 concludes the paper with some final remarks.

## 2. DCT-based Complexity Estimation

We develop our rate control scheme on a modified MPEG-2 to MPEG-4 transcoder based on [8], as showed in Figure 1. In this transcoder, the motion vector reused algorithm is deployed. In order to reduce the computational complexity while maintain the reconstructed video quality, motion compensation is done only when it's P frame, I and B frames are transcoded in frequency-domain. After the incoming MPEG-2 bitstream is performed VLD and inverse quantized, for P frame, the DCT coefficients are fed to perform IDCT and motion compensation in the decoder end, and then fed to the MPEG-4 encoder end, whilst for I/B frame, the

DCT coefficients are fed to perform MPEG-4 re-quantization directly. But since I frames are the anchors for subsequent P and B frames, the IDCT at the MPEG-2 decoder, inverse quantization and IDCT of MPEG-4 for I frames are still needed to reconstruct the reference frames.

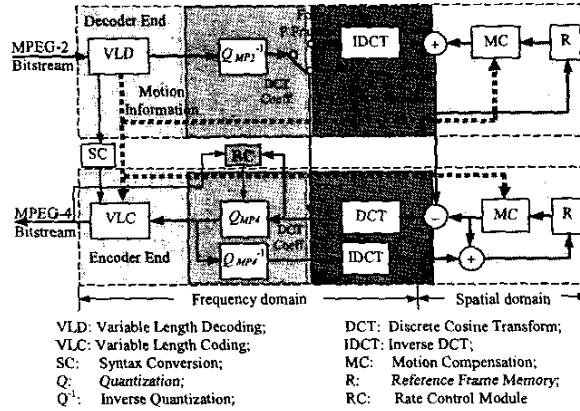


Figure 1: The MPEG-2 to MPEG-4 transcoder.

We adopt the model which describes the relation between the required bits for texture coding and QP [7]:

$$T = S \cdot \left( \frac{X_1}{QP} + \frac{X_2}{QP^2} \right) \quad (1)$$

where  $T$  is the target bits for a frame;  $S$  is the mean absolute difference which is computed after the motion compensation for the Y component, , it represents the complexity of a frame;  $X_1$  and  $X_2$  are the first and second order model parameters. Using this model, QP can be solved if  $T$ ,  $S$ ,  $X_1$ ,  $X_2$  are known.

However, in the transcoding architectures without ME, it's difficult to get MAD. Thus, we cannot obtain QP by (1). From the energy conservation property of DCT, the signal energy calculated in the frequency-domain is equal to the energy in the spatial domain, the complexity can be computed from DCT coefficients. Thus, we define a new DCT-based complexity  $S'_t$  for a frame at time  $t$ .

$$S'_t = \frac{1}{N} \sum_{m \in M} \sum_{i=1}^{64} W_i \cdot |P_m(i)| \quad (2)$$

where  $N$  is the number of pixels in a frame,  $m$  is the block index in the set  $M$  of blocks in a frame,  $P_m(i)$  is the  $i^{th}$  DCT coefficient in the  $m^{th}$  block .  $W_i$  is the weight factor of  $P_m(i)$  and is set to 1 in experiments.

The proposed DCT-based complexity (2) indicates the energy of DC and AC coefficients, therefore, can reflect the instantaneous characteristics of this frame and dictate how many bits can be appropriate for the frame before really encoding it. This is especially useful when a scene change happens, because the complexity of a frame can reflect the change, according to it, this frame can obtain appropriate target bits. However, TM5 rate control can't deal with scene changes properly. The reason is that, target bits are allocated to the current frame according to the statistical information of its previous frame, without any consideration to the real complexity of the current frame. This may

result in inappropriate allocation of bits to the current frame, which can lead to fluctuating and overall degraded visual quality.

### 3. Rate Control for Video Transcoding

**1) Bit Allocation for Video Transcoding:** According to the type of the current frame, its target number of bits is initially set to a weighted average bitcount:

$$T_{ave} = \alpha_k \cdot \frac{R_r}{\sum_{i=1}^M \alpha_i \cdot N_i}$$

where  $M$  is the number of picture types,  $N_i$  is the remaining number of frames with picture type  $i$ ,  $R_r$  is the remaining number of bits,  $\alpha_i$  is a weight factor of picture type  $i$ ,  $\alpha_k$  is the weight corresponding to the current picture type.

According to the complexity of current frame, its target bit budget  $T_t$  is then estimated by:

$$T_t = T_{ave} \times \sqrt{\frac{S'_t}{S'_{ave}}}$$

where  $S'_{ave}$  is the average complexity of previous  $n$  frames before time  $t$ . The number of target bits is estimated only for P and B frames. We do not estimate target bits for I frames [9]. This means if the current coding complexity  $S'_t$  is higher than the average coding complexity  $S'_{ave}$ , more bits should be allocated to the current frame than  $T_{ave}$ ; on the contrary, if  $S'_t$  is lower than  $S'_{ave}$ , fewer bits should be allocated. Therefore, appropriate bits can be adaptively allocated to the current frame and coding quality can be kept constant. Since it allows bits to be saved in easy scenes so that more bits can be used for difficult scenes.

**2) Buffer Control for Video Transcoding:** The initial bit budget is further refined based on the buffer fullness to get a more accurate target bit estimation. We adopt our Proportional-Integral-Differential (PID) buffer control technique [9] in transcoding. The PID coefficient is computed as:

$$E_t = (B_s / 2 - B_f) / (B_s / 2),$$

$$P_t = K_p \times E_t + K_i \times \int E_t \cdot dt + K_d \times (dE_t / dt),$$

where  $B_s$  is the buffer size,  $B_f$  is the current buffer fullness at time  $t$ ,  $K_p$ ,  $K_i$  and  $K_d$  are the Proportional, Integral and Differential control parameters respectively, and are set to 1.0, 0.25 and 0.3 respectively by experiments. Then the initial target bits,  $T_t$ , can be further adjusted by:

$$T_t = T_t \times (1 + P_t).$$

To maintain a minimum acceptable visual quality for each frame, the target is not lower than  $T_{ave}/C_{MIN}$ , and to avoid buffer overflow, a maximum number of bits  $T_{ave} \cdot C_{MAX}$  is given,  $C_{MIN}=2.0$  and  $C_{MAX}=2.8$  are constant factors decided by experiments.

**3) QP Calculation, Encoding and Model Updating:** Once the target number of bits is obtained, QP is solved based on the model (1). The initial QPs of the first 3 frames in MPEG-4 encoder are obtained by the following ways:

$$QP_i = QP_{ave,i} \times K_i \times \sqrt{\frac{T_1}{T_2}}, \quad (i = 1, 2, 3)$$

where  $QP_{ave,i}$  is the average QP adopted in the  $i^{\text{th}}$  frame by MPEG-2 encoder,  $K_1$  is a adjusting coefficient ( $K_1 = 0.2$ ,  $K_2 = 1.1$ ,  $K_3 = 0.5$ ),  $T_1$  is the bitrate of the input MPEG-2 bitstream,  $T_2$  is the target bitrate adopted by the transcoder.

The MPEG-4 encoder in the transcoder starts to encode the current frame after QP is obtained. After encoding, the encoder updates the model based on the encoding results, for example, modifies the model parameters  $X_1$  and  $X_2$  [7].

**4) Frame-Skipping Control for Video Transcoding:** To transmit video over low bandwidth channels, a high transcoding ratio or a very low target bit rate is required. In conjunction with determining suitable spatial coding parameters, our rate control scheme can resort to changing the temporal coding parameters to obtain the desired transcoding bitrate.

In the transcoder, the frame type for each frame in MPEG-4 encoder should be as same as that of the incoming stream. Since I & P frames are anchors, in addition, B frames could not be encoded if their reference I or P frames are skipped. In order to avoid transcoding quality degradation by frame skipping, we only skip B frames. The encoder examine the current buffer fullness before encoding the next B-frame, if the buffer occupancy exceeds 80% of the buffer size, the next B-frame is skipped.

#### 4. Experiments Results and Discussions

We have implemented the proposed rate control algorithm named FDRC (Frequency-Domain Rate Control) in the MPEG-2 to MPEG-4 transcoder. To demonstrate the performance of FDRC, we have implemented TM5 and VM18 rate control on the same transcoder. Since MAD in the model (1) can't be obtained in our transcoder, we made a little change to this model, that is, we replace MAD by our DCT-based complexity in the model.

Different spatial resolution sequences are encoded in MPEG-2 at various bit rates using a public-domain MPEG-2 simulation software. These pre-encoded video streams are transcoded into lower bit-rate MPEG-4 bitstream with rate control. All sequences are in YUV format 4:2:0, the number of P frames is set to 4 between two I frames, the number of B frames is set to 2 between two P frames or between two I & P frame, GOP-size is 15, and the frame-rate is 30 fps, the number of frames in each bitstream to be transcoded is 150.

**1) Experimental Results Without Frame Skipping:** MPEG-2 is used to provide high quality in video digital broadcast, and thus it should have a fixed group of pictures, and cannot skip frames in TM5. To fairly compare results among these three algorithms, we disable frame skipping function in FDRC and VM18.

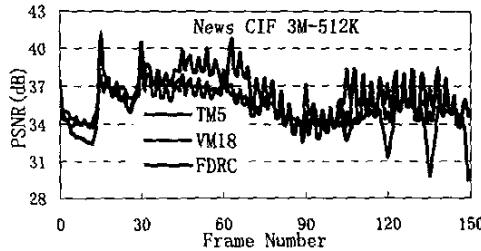


Figure 2: PSNR curves for transcoding without frame skipping.

TM5 rate control, realized at MB level, decides QP for each MB; while FDRC, realized at frame level, only needs to compute QP for a frame. Thus, FDRC has lower delay and needs less computation than TM5. Under this case, from the results in Table 1 and Fig. 2, one can see that FDRC can still achieve a little more accurate target bit rates and higher average PSNRs than TM5. In addition, when compared with VM18, FDRC also achieves more accurate target bit rates and higher average PSNRs. These results show that our scheme is effective.

**2) Experimental Results With Frame Skipping:** By examining the results in Table 2, it is obvious that FDRC achieves more accurate target bit rates and target frame rate (30fps) with higher average PSNRs as compared to the VM18 solution. One may note that the actual bit rate to transcode Foreman using VM18 (84.18Kbps) is little closer to the target (84Kbps) than that of FDRC (84.54 Kbps), however VM18 skip 33 frames to get the rate while FDRC does not skipping any frames to realize 84.54 Kbps. From Fig. 3(b), the buffer occupancy curve of FDRC is quite stable, it is around 50% of the buffer size with a small variation. However, by examining the buffer occupancy curve produced by VM18, it is evident that VM18 has less control ability, e.g. the buffer occupancy of Fig. 3(b) is out of control, and results in more frame skipping cases. In addition, visual quality of VM18 is more fluctuated than that of FDRC (See Fig. 3(a)).

**3) Experimental Results of Scene Change Handling:** In order to test scene change handling abilities of FDRC and TM5, combined sequences are used. For the combined sequence "Mother\_daughter-Foreman" (Fig.3(c)), the first 42 frames are from the "Mother\_Daughter" and the remaining 48 frames are from the "Foreman". It is encoded in QCIF format and transcoded from 1.5Mbps to 256Kbps. In Fig. 3(c), TM5 performs poorly at scene change (frame 42) and its subsequent frames, since it only utilizes information obtained from previously coded pictures in estimating target bits for the current picture, when a scene change occurs, information obtained from previous coded pictures is no more suitable for the current frame and causes visual quality degradation in the pictures following the scene change. However, our DCT-based complexity increases when a scene change happens, and the scene change frame can obtain enough bits according to our bit allocation method. As a result, FDRC generally obtains higher average PSNRs than TM5 through the whole combined sequences. In addition, the visual quality at scene change is significantly improved, meanwhile, the performance of its following frames is also improved. These results show our complexity estimation scheme can get better visual quality and bit allocation.

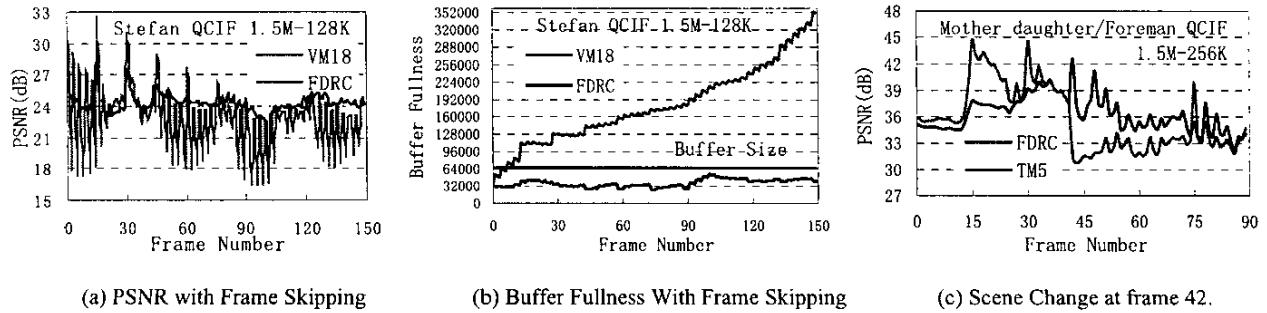
#### 5. SUMMARY

In this paper, we have proposed a rate control scheme for an efficient MPEG-2 to MPEG-4 transcoder. Considering transcoding properties, we define a complexity estimation in the frequency domain, perform effective bit allocation among frames, use a frame skipping mechanism to achieve very low target bitrates of transcoding, and adopt a PID buffer control strategy to effectively handle buffer fullness. Experimental results authenticate that the proposed algorithm outperforms VM18 in: (a) providing more accurate target bitrate transcoding; (b) achieving better picture quality; (c) maintaining a more stable buffer level. At the mean time, our algorithm achieves better performance than TM5 by: (a)

providing a little more accuracy in target bitrate transcoding; (b) obtaining higher PSNRs; (c) having lower delay and computation; (d) handling scene change more effectively. This algorithm can also be used in other transcoding architectures.

## 6. REFERENCES

- [1] J. Xin, M. T. Sun, and K. Chun, "Bit-allocation for Transcoding of Pre-encoded Video Streams," in *Visual Communications and Image Processing*, Proc. SPIE 4671, pp.164-171, San Jose, CA., Jan. 2002.
- [2] Huifang Sun, Wilson Kwok, and Joel W. Zdepski, "Architectures for MPEG Compressed Bitstream Scaling", *IEEE Trans. on Circuits and Systems for Video Technology*, Vol. 6, No. 2, April 1996.
- [3] B. Sostawa, T. Dannemann, and J. Speidel, "DSP-Based Transcoding of Digital Video Signals with MPEG-2 Format," *IEEE Trans. On Consumer Electronics*, vol. 46, No. 2, May 2000, pp. 358 – 362.
- [4] K.-D. Seo, S.-H. Lee, J.-K. Kim, and J.-S. Koh, "Rate Control Algorithm For Fast Bit-Rate Conversion Transcoding," *IEEE Trans. on Consumer Electronics*, Vol. 46, pp. 1128-1136, Nov. 2000.
- [5] P.N. Tudor and O.H. Werner, "Real-Time Transcoding of MPEG-2 Video Bit Streams," *IEE international Broadcasting Convention*, Amsterdam, pp. 286-301, 1997.
- [6] ISO/IEC-JTC1/SC29/WG11, MPEG93/457, Test Model 5, April 1993.
- [7] Weiping Li, Jens-Rainer Ohm, Mihaela van der Schaar, Hong Jiang, Shipeng Li, "MPEG-4 Video Verification Model V18.0", ISO/IEC JTC1/SC29/WG11 N3908, Pisa, January 2001.
- [8] Rong Xie, Jilin Liu, and Xingguo Wang, "Efficient MPEG-2 to MPEG-4 Compressed Video Transcoding," in *Visual Communications and Image Processing*, Proc. SPIE 4671, pp.192-201, San Jose, CA., Jan. 2002.
- [9] Sun Yu and Ishfaq Ahmad, "A New Rate Control Algorithm for MPEG-4 Video Coding," in *Visual Communications and Image Processing*, Proc. SPIE 4671, pp.698-709, San Jose, CA., Jan. 2002.



(a) PSNR with Frame Skipping

(b) Buffer Fullness With Frame Skipping

(c) Scene Change at frame 42.

**Figure 3:** Transcoding experimental results.

**Table 1:** Transcoding experimental results without frame skipping.

Sequence	Conversion (Bps)	Actual Bitrate (Kbps)			PSNR (dB)		
		TM5	VM18	FDRC	TM5	VM18	FDRC
Container (QCIF)	1500K~1024K	1024.77	1020.15	1024.44	28.27	28.66	28.69
Mobile (QCIF)	1500K~1024K	1024.77	1020.15	1024.44	28.27	28.66	28.69
Stefan (CIF)	3000K~1024K	1026.29	1142.11	1023.44	29.39	28.48	29.80
Mother_Daughter (CIF)	3000K~512K	513.07	538.55	512.98	38.83	38.71	39.18
Table (CIF)	3000K~384K	384.90	467.07	383.85	29.11	29.15	29.26
News (CIF)	3000K~512K	513.82	531.47	511.11	35.78	35.32	36.09
Flower (CCIR601)	6000K~3500K	3515.49	3945.09	3502.61	27.11	25.93	27.47
Cheer (CCIR601)	6000K~5000K	3991.34	5159.06	3992.68	27.80	27.77	28.01

**Table 2:** Transcoding experimental results with frame skipping.

Sequence	Conversion (Bps)	Actual Bitrate (Kbps)		PSNR (dB)		Skipped Frames	
		VM18	FDRC	VM18	FDRC	VM18	FDRC
Foreman (QCIF)	1500K~84K	84.18	84.54	27.97	28.22	33	0
Stefan (QCIF)	1500K~128K	141.68	128.13	23.19	24.33	49	1
News (CIF)	3000K~256K	254.56	256.03	31.47	31.77	25	0
Mobile (CIF)	3000K~640K	641.92	638.64	22.81	23.54	34	20
Mother_Daughter (CIF)	3000K~256K	256.97	256.96	35.03	35.58	15	9