RATE-DISTORTION OPTIMIZED DCT-DOMAIN VIDEO TRANSCODER FOR BIT-RATE REDUCTION OF MPEG VIDEOS

Wen-Nung Lie, Ming-Lun Tsai, and Tom C.-I. Lin

Department of Electrical Engineering
National Chung Cheng University, Chia-Yi, 621, Taiwan, ROC.
Phone : 886-5-2720411 ext. 33211, Fax : 886-5-2720862
E-mail : wnlie@ee.ccu.edu.tw

ABSTRACT

In this paper, we propose a rate-distortion optimized video transcoder which converts MPEG videos into a similar form at lower bit-rates. Our transcoder design is characterized of two features: 1) transcoding is performed in DCT domain and the MV information is re-used, and 2) the rate-distortion relationship is optimized in both the frame-level rate allocation and MB-level rate control, thus leading to performances even better than the direct encoding and re-encoding methods based on the well-known TM5. In the proposed algorithm, the Lagrangian multiplier plays not only its traditional role in MB-level optimization, but also a variable to be optimized in frame-level rate allocation. These two levels of optimization process are highly linked. Experiments show that the R-D optimization is effective in getting better video quality, even the drift errors are ignored. Several speedy schemes were developed to make our transcoder design suitable for real-time video transmission over heterogeneous networks.

I. INTRODUCTION

Traditional videos coded at constant bit rate (CBR) suffered from poor adaptativity to channel capacity variation. Transcoding is essential to video transmission over networks which in general faces varying channel bandwidths and require best-effort quality of service (QoS).

While a video sequence is transcoded to be at a lower bit rate, it will suffer from picture quality degradation. How to develop an efficient algorithm to make quality degradation as less as possible, or make efficient use of allowable bit rate, is still a challenging problem and needs insight discussions.

The basic design of a transcoder is the cascade of a decoder and an encoder. Two kinds of transcoders, pixel-domain and frequency-domain [1-3], have ever been developed. For pixel-domain transcoders, frame intensities are reconstructed, new residual DCT coefficients are calculated by re-using the remained motion vectors (MVs) and performing DCT transformation, and finally a new bit stream is constructed after re-quantization and VLC coding to fit the specified display format or channel bandwidth. The re-use of prior motion vectors clearly decreases the computational complexity. For frequency-domain transcoder design, not only motion vectors are re-used but also a pair of Inverse-DCT and DCT calculation is cancelled to speed up the transcoding process.

Rate allocation and rate control play important roles in networked video applications. In [5,6], the rate-ρ and distortion-ρ models (ρ stands for the percentage of zeros of DCT coefficients in a frame) were established and used to estimate the resulting bit rate and distortion. These estimated data were then utilized in a Lagrangian R-D optimization method. However, these ρ-based models may not be proper for a DCT-domain transcoder since some parameters in the models can not be obtained directly from the DCT coefficients. Besides, the ρ-based models are not accurate enough when applied in MB-level rate control.

Here, a DCT-domain transcoder is adopted. Figure 1 reveals the processing flow of our algorithm. Bit rate reduction was achieved by simply re-quantizing the de-quantized DCT coefficients. No explicit arrangements for drift-error elimination were performed. However, optimizations were considered in both the frame-level rate allocation and the MB-level rate control. As traditions, the Lagrangian multiplier method [4] was adopted for optimization in determining frame target bit rate and choosing a proper quantization parameter (QP) for each MB. Though drift errors were not compensated during the transcoding process, they were implicitly reduced via the optimization procedure and limited by the GOP (Group of Pictures) structure in the standard. Our rate control algorithm is not model-based, that means, try-and-error search on proper QP for each MB will be necessary. Though this has a higher accuracy and efficiency in rate control, it also leads to extensive computations. Our algorithm was accelerated to achieve a 50% saving in computation without sacrificing visual quality.

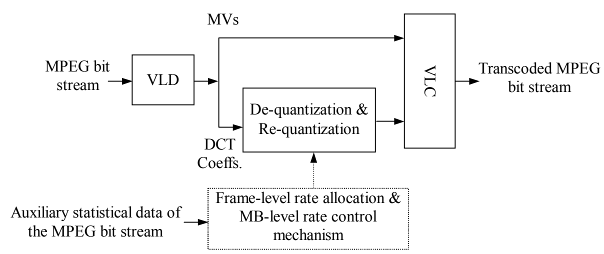


Fig.1 Processing flow of the proposed frequency-domain video transcoder

II. MACROBLOCK-LEVEL RATE CONTROL

Rate control in MB level includes the determination of QP for each MB in a frame such that the quantized DCT coefficients, after being VLC-coded, meet the frame target bit rate. The MB-level rate control can be considered as a QP assignment problem, with intent to optimize the transcoding error of a frame. For intra-coded (I) frames, the transcoding distortion all comes from the re-quantization errors. For inter-coded (P or B) frames, the transcoding errors include both the re-quantization error and the drift errors.

Let Q be a vector of quantization parameters QPs: $Q=\{q_1,q_2,\cdots,q_M \mid q_k \in \{qp_1,qp_2,\cdots,qp_N\}\}$, where M represents the number of MBs in a frame and N is the number of possible QPs that can be chosen. $D(Q)$ is the total frame distortion under Q. Here we use the MSE (mean square error) for the measure of distortion between the images before and after transcoding. Denote $d(q_k)$ as the k-th MB's distortion under quantization parameter q_k, $R(Q)$ as the total bit rate according to the adoption of Q, and $r(q_k)$ as the k-th macroblock's bit rate according to q_k. We can write $D(Q)$ and $R(Q)$ as below

$$D(Q)=\sum_{k=1}^{M} d(q_k) \quad (1)$$

$$R(Q)=\sum_{k=1}^{M} r(q_k) \quad (2)$$

Under the constraint of target bit rate R_c, the optimization problem can be written as:

$$\underset{Q=\{q_k|k=1\sim M\}}{Minimize}\{\sum_{k=1}^{M} d(q_k)\} \quad (3)$$

subject to $$\sum_{k=1}^{M} r(q_k)\le R_c \quad (4)$$

According to the Lagrangian multiplier method, Eqs.(3)&(4) can be converted to

$$\underset{Q=\{q_k|k=1\sim M\}}{Minimize}\{\sum_{k=1}^{M} d(q_k)+\lambda\cdot\sum_{k=1}^{M} r(q_k)\} \quad (5)$$

Grouping individual q_k^* for each sub-optimization problem is ready to form $Q^*=\{q_k^*|k=1\sim M\}$. By varying λ, we search for λ^* which minimizes

$$\varepsilon=\left|\sum_{Q^*(\lambda)} r(q_k^*)-R_c\right|. \quad (6)$$

$Q^*(\lambda^*)$ is thus what we purchase.

For each MB, we have to appraise candidates of q_k to minimize $d(q_k)+\lambda\cdot r(q_k)$. This clearly spends extensive computing time since in MPEG, each QP ranges from 1 to 31. Denote the QP value obtained from the transcoder input for the k-th MB to be q_k^0. Also denote $q_k^{\max}$ $(\ge q_k^0)$ as the least QP value which is expected to re-quantize all DCT coefficients in the k-th MB to zeros. Fortunately, the search range of q_k can be narrowed down to $q_k^0\sim q_k^{\max}$, rather than 1~31.

III. FRAME-LEVEL RATE ALLOCATION

How to determine R_c for each frame so that the coding bit rate meets the channel bandwidth at any time? Though our MB-level rate control is modeless, our rate allocation between frames is based on certain prediction models to prevent extensive computations.

A. *R*-λ and *D*-λ models

It was found that there exist regular relations between the Lagrangian multiplier λ^* and the resulting frame bit-rate $R(Q^*(\lambda^*))$. It is similar between λ^* and the resulting distortion $D(Q^*(\lambda^*))$. It is thus our goal to convert this kind of regular relations into prediction models and use them for frame-level rate allocation. Here, the Lagrangian multiplier λ can be considered as the position index on the R-D curve of a frame (Fig.2). Different point corresponds to different λ. Determination of the multiplier of a frame is thus to determine its rate and distortion. We call these relations the *R*-λ and *D*-λ models.

To establish *R*-λ and *D*-λ models for each frame, we performed pre-analyses and stored model parameters for later transcoding. Since parameters of both models depend on the image content, their acquiring in advance is surely helpful to transcoding performance. This pre-analysis strategy makes our proposed transcoder optimized, even disregarding the drift errors caused by simple re-quantization of the DCT coefficients.

For each MB, by varying λ, the corresponding $R(Q^*(\lambda))$ and $D(Q^*(\lambda))$ can be figured out. The collected data can be plotted in (R, λ) and (D, λ) planes, as shown in Fig.3, and fitted both by using the logarithmic models :

$$\begin{cases} R(\lambda)=a\cdot\ln(\lambda)+b \\ D(\lambda)=c\cdot\ln(\lambda)+d \end{cases}. \quad (7)$$

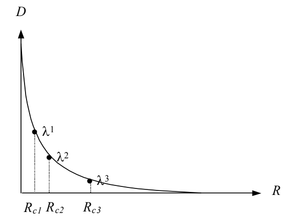
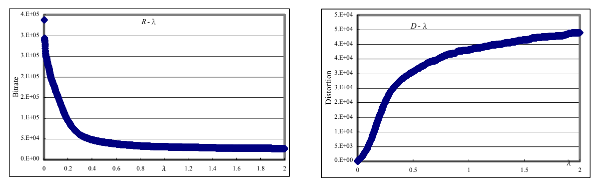


Fig.2 Rate-distortion curve with λ as the position index



(a) (b)

Fig.3 Plotting of *R*-λ and *D*-λ data sets

To solve the parameters *a, b, c,* and *d,* the popular least-square-error method can be applied. Since each frame has its own statistical characteristic, we have to calculate individual *R*-λ and *D*-λ model parameters for each frame. Besides, they occupy 16 bytes for storage requirement.

B. Frame-level rate allocation

The aim of frame-level rate allocation is to find out the optimal bit-rate for each frame to achieve the best overall video quality. This can be also considered as a resource allocation problem in GOP. The *R*-λ and *D*-λ models established above are used to complete the rate allocation process. For clarity, we describe the frame-level rate allocation and MB-level rate control in an integrated manner as follows.

Step 1: For each frame, *N* data points of the retrieved *R*-λ and *D*-λ models are interpolated. Denote them as $\{(\lambda_{i,1},R_{i,1}^F),(\lambda_{i,2},R_{i,2}^F),\ldots,(\lambda_{i,N},R_{i,N}^F)\}$ and $\{(\lambda_{i,1},D_{i,1}^F),(\lambda_{i,2},D_{i,2}^F),\ldots,(\lambda_{i,N},D_{i,N}^F)\}$, where $\lambda_{i,n}$, $R_{i,n}^F$, and $D_{i,n}^F$ represent the n-th data points of the i-th frame.

Step 2: Describe the frame-level rate allocation problem in terms of a constrained optimization problem:

$$\underset{\{\lambda_i\}}{Minimize}\{\sum_{i=1}^{N_{GOP}} D_i^F(\lambda_i)\} \quad (8)$$

subject to $$\sum_{i=1}^{N_{GOP}} R_i^F(\lambda_i)\le R_{GOP}, \quad (9)$$

where N_{GOP} is the GOP size, R_{GOP} is the target bit rate, and $D_i^F(\lambda_i)$ and $R_i^F(\lambda_i)$ are the distortion and bit-rate, respectively. Now, our goal is to determine λ for each frame so that the sum of induced distortions is minimized under the GOP bit-rate constraint. Hence, the frame-level and MB-level rate allocation are very similar. The major difference comes from the fact that $d(q_k)$ and $r(q_k)$ in Eqs.(3)&(4) are obtained from real trials, while $D_i^F(\lambda_i)$ and $R_i^F(\lambda_i)$ in Eqs.(8)&(9) are interpolated from the retrieved models. Similarly, we convert the constrained optimization problem into

$$\underset{\{\lambda_i\}}{Minimize}\{\sum_{i=1}^{N_{GOP}} D_i^F(\lambda_i) + \Lambda \cdot \sum_{i=1}^{N_{GOP}} R_i^F(\lambda_i)\}, \quad (10)$$

where Λ is the non-negative Lagrangian multiplier (in contrast to the λ in MB-level).

Step 3: By varying Λ from zero to infinity, find Λ^* (and associated $\{\lambda_i^*\}$) that minimizes

$$\varepsilon = \left| \sum_{i=1}^{N_{GOP}} R_i^F(\lambda_i^*) - R_{GOP} \right| . \quad (11)$$

$R_i^F(\lambda_i^*)$ forms the result of rate allocation for the *i*-th frame and is set as the target bit rate R_c for MB-level rate control.

Step 4: Let $\{\lambda_i^*\}$ obtained above be the initial Lagrangian multipliers for MB-level rate control.

Step 5: Complete the MB-level rate control for each frame in the GOP based on the initialized $\{\lambda_i^*\}$. Use the rate control result $Q^* = \{q_k^*\}$ to re-quantize DCT coefficients.

Basically, the setting of R_{GOP} was obtained from the channel estimator to reflect the channel bandwidth available. Since the *R*-λ and *D*-λ models are not absolutely precise, there would be an error between the target R_{GOP} and the actual bit-rate after MB-level rate control. Over-use or under-use of bit-rates in previous GOPs should be compensated in current GOP to prevent error accumulation.

In steps 3&4, a search of the Lagrangian multipliers Λ^* (for a GOP) and λ^* (for a frame) is necessary. Clearly, a brute-force search of them will spend lots of computing time. In this paper, a gradient-descent-like fast searching algorithm was introduced to speed up the transcoding process.

Our rate allocation and rate control algorithms have a need to pre-analyze the MPEG video sequences for *R*-λ and *D*-λ model parameters. Assuming a 4-byte floating number is adopted for each parameter, this overhead information would occupy a storage of 4*4*8*30=3840 bits per second. For an input MPEG video coded at 6144 kbps and 30 fps, it is equivalent to 0.062% of overheads, which can be ignorable.

IV. ACCELERATION IN SEARCHING PARAMETERS

We know that the search of optimal multipliers, both for a GOP or for a frame, is computationally intensive. This obviously slows down the transcoding process. Hence, our goal is to find a near-optimal multiplier efficiently, without losing much of the transcoding performance, i.e., the image quality.

To make a tradeoff between picture quality and time saving, we propose a gradient-descent-like method to find the optimized Lagrangian multipliers quickly and accurately. In the following, the search of the optimized Λ is given as an example and the search of the optimized λ adopts a similar strategy.

A. Gradient-descent-like procedure in searching Lagrangian multipliers

According to experiments, it was found that there exists a monotonically decreasing relationship between the multiplier Λ and the produced GOP bit rate. The gradient-descent-like method to find the optimal multiplier Λ can be illustrated in Fig.4, where the strategies of $\Lambda_{t+1} = 0.5(\Lambda_t + \Lambda_{t-1})$, $\Lambda_{t+1} = 0.75 \cdot \Lambda_t$, or $\Lambda_{t+1} = 1.5 \cdot \Lambda_t$ is adopted, according to the relative behavior between $B(\Lambda_t)$ and $B(\Lambda_{t-1})$, where $B(\Lambda_t)$ represents $\sum_{i=1}^{N_{GOP}} R_i^F(\lambda_i(\Lambda_t))$. Record the converged value of Λ and mark it as Λ^*. Get the corresponding $\{\lambda_i^*\}$.

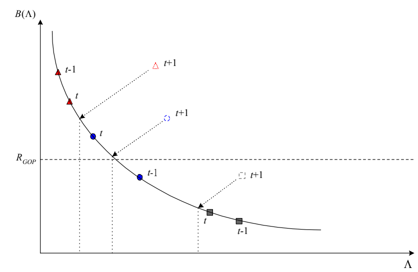


Fig.4 Gradient-decent-like search for optimized Λ

As for the search of λ^* for each frame in rate control process, this algorithm was similarly adopted. Fortunately, the set of parameters $\{\lambda_i^*\}$ found in the rate allocation process for each GOP also represents the Lagrangian multipliers and hence can be used as the initial values for λ^*'s. This obviously speeds up the rate control process for each frame.

B. Speedup in searching the quantization parameters

In our MB-level rate control process, the distortion and produced bit rate have to be figured out for each possible QP value from q_k^0 to q_k^{max} so that the optimization process can be conducted. On one hand, real quantization and VLC coding for a specific q_k were performed on DCT coefficients to get the $r(q_k)$ and $d(q_k)$. On the other hand, a prediction model was adopted to reduce the number of evaluated q_k's. It means that the search range of q_k can be further narrowed down provided that it still covers the optimum q_k^*. Based on this strategy, it was found that the difference between q_{k-1}^* and q_k^* is mostly around 0~3. Taking advantage of this property, a full search was done for the first MB in a frame and a narrowed-down search range $q_k \in (q_{k-1}^*\text{-}3) \sim (q_{k-1}^*\text{+}3)$ was adopted for the following MBs. This process speedsup the MB-level rate control efficiently, without loss of perceivable quality degradation. We will show the experimental results in next section.

V. EXPERIMENTAL RESULTS

The experiments were focused on the transcoding performances and the acceleration issue. Four kinds of methods were compared. (1) Direct encoding (the raw video data were encoded into the target bit-rate, along with the TM5 rate control), (2) Re-encoding (the input MPEG video was decoded into the pixel domain and then re-encoded with the TM5 rate control), (3) TM5-based transcoding (a DCT-domain video transcoder based on re-quantization strategy. TM5 rate control mechanism was adopted), and (4) Proposed transcoding algorithm. To compare the transcoding performances, the videos were at first encoded at 30 frames/sec, 6144 kbps (CIF format), with the TM5 rate control mechanism adopted. The bit-rate is then reduced to be 1024, 2048, and 3072 kbps.

Table I lists the statistics of average PSNR and transcoding accuracy in bit-rate. In the table, "Loss" represents the difference in PSNR for the proposed algorithm with respect to the considered method and "output precision" represents the difference between the actual output bit-rate and target bit-rate, in terms of the proportion with respect to the target bit-rate. It can be found that there is a gain of about 2.82 dB ~ 5.6 dB, 1.38 dB ~ 3.57 dB and 0.9 dB ~ 1.92 dB for the "Flower", "Table_tennis", and "Foreman" sequence, respectively. In view of the output bit-rate accuracy (shown in Table II.), the proposed algorithm can always keep it within 1% of precision (0.49% ~ -0.81%) with respect to the target bit-rate.

Our output bit-rate accuracy is worse than the "Re-encode" and "Encode", but sill within an acceptable tolerance (<0.81%). The inaccuracy mostly comes from mis-match of the R-λ and D-λ models in frame-level rate allocation. Experiments proved that the proposed models are accurate enough.

Our proposed transcoder even outperforms the "Re-encode" and "Encode" designs, in spite of ignoring the drift errors in reconstructing P and B frames. This seldom phenomenon reveals the importance of rate allocation/control designs over the compensation of drift errors in a transcoder. Since the proposed R-λ and D-λ models for each frame is accurate enough and a full optimization procedure is conducted for each MB, the overall video quality is better than algorithms that are somehow heuristic in nature or fully model-based.

For acceleration of MB-level rate control, our speedup is approximately 50%~60% of the full search with only less than 0.15 dB video quality degradation.

VI. CONCLUSIONS

We propose a rate-distortion optimized video transcoder, derived from a cascade of decoder and encoder to convert MPEG video sequences into lower bit-rates that adapt to network bandwidth. At the MB-level rate control, the R-D optimization is based on a search on all possible combinations of QPs. At the frame-level rate allocation, the R-D optimization is based on the R-λ and D-λ models whose parameters can be pre-analyzed, stored, and later used in the transcoding process. Since the Lagrangian multiplier in our MB-level optimization process is also the variable to be optimized in the frame-level rate allocation process, these two optimization process can be highly linked. Our acceleration schemes made real-time transcoding achievable over heterogeneous networks. A variety of services and applications, such as Video on Demand (VoD), would benefit from the proposed transcoder design in this paper.

ACKNOWLEDGEMENT

This work was supported by NSC, Taiwan, ROC., under the grant of NSC91-2213-E-194-012.

REFERENCES

[1] Pedro A. A. Assuncao and Mohammed Ghanbari, "A Frequency-Domain Video Transcoder for Dynamic Bit-Rate Reduction of MPEG-2 Bit Streams," *IEEE Trans. on Circuits and Systems for Video Technology*, Vol. 8, No. 8, pp. 953~967, Dec. 1998.

[2] Limin Wang, Ajay Luthra and Bob Eifrig, "Rate Control for MPEG Transcoders," *IEEE Trans. on Circuits and Systems for Video Technology*, Vol. 11, No. 2, pp. 222~234, Feb. 2001.

[3] Yair Shoham and Allen Gersho, "Efficient Bit Allocation for an Arbitrary Set of Quantizers," *IEEE Trans. on Acoustics, Speech, and Signal Processing*, Vol. 36, No. 9, pp. 1445~1453, Sep. 1988.

[4] H. Everett, "Generalized Lagrange Multiplier Method for Solving Problems of Optimum Allocation of Resources," *Operation Res*., vol. 11, pp. 399-417, 1963.

[5] Zhihai He and Sanjit K. Mitra, "Optimum Bit Allocation and Accurate Rate Control for Video Coding via ρ-Domain Source Modeling," *IEEE Trans. on Circuits and Systems for Video Technology*, Vol. 12, No. 10, pp. 840~849, Oct. 2002.

[6] Zhihai He and Sanjit K. Mitra, "A Linear Source Model and a Unified Rate Control Algorithm for DCT Video Coding," *IEEE Trans. on Circuits and Systems for Video Technology*, Vol. 12, No. 11, pp. 970~982, Nov. 2002.

Table I Average PSNR (unit in dB)

Output bitrate	Test sequence	Proposed Transcoder	TM5 Transcoder		Re-encode		Encode	
			PSNR	Loss	PSNR	Loss	PSNR	Loss
1024 kbps	Flower	35.23	31.41	+3.82	30.28	+4.95	30.30	+4.93
	Table tennis	36.23	33.98	+2.25	33.59	+2.64	33.65	+2.58
	Foreman	37.01	35.96	+1.05	36.12	+0.90	36.08	+0.93
2048 kbps	Flower	41.83	36.94	+4.89	38.34	+3.49	39.01	+2.82
	Table tennis	41.79	38.94	+2.86	39.68	+2.11	40.41	+1.38
	Foreman	40.54	38.95	+1.59	39.30	+1.24	39.24	+1.30
3072 kbps	Flower	46.87	41.39	+5.48	41.26	+5.60	41.38	+5.49
	Table tennis	46.31	42.75	+3.57	43.21	+3.10	43.47	+2.85
	Foreman	43.08	41.18	+1.89	41.23	+1.84	41.16	+1.92

Table II Transcoding accuracy

Output precision	Test sequence	Proposed Transcoder	TM5 Transcoder	Re-encode	Encode
1024 kbps	Flower	0.49%	0.79%	0.26%	0.27%
	Table tennis	0.5%	1.3%	0.39%	0.4%
	Foreman	0.54%	1.18%	0.23%	0.24%
2048 kbps	Flower	0.74%	0.44%	0.11%	0.12%
	Table tennis	0.59%	0.92%	0.2%	0.2%
	Foreman	0.52%	0.57%	0.07%	0.07%
3072 kbps	Flower	0.52%	0.35%	0.09%	0.09%
	Table tennis	0.51%	0.67%	0.11%	0.12%
	Foreman	-0.81%	0.38%	0.01%	0.02%