RATE ESTIMATION FOR H.264/AVC SPATIAL RESOLUTION REDUCTION

Peter H. W. Wong, Robert T. W. Hung, Jack Y. B. Lee, S. C. Liew, C. S. Kim, Roland T. Chin†

Department of Information Engineering, The Chinese University of Hong Kong
Department of Computer Science, The Hong Kong University of Science and Technology†

ABSTRACT

The emergence of wireless networks has posed new challenges for multimedia content providers. In particular, many existing video contents have bit-rates and resolutions far exceeding the capability of wireless networks and mobile devices, and they need to be adapted before transmission. In this paper, we propose a technique using ρ-domain information to estimate the bit rate of down-sampled video in the H.264/AVC format. The bit rate is estimated in two stages. In the initialization stage, we use a low-complexity algorithm to estimate the ρ-domain parameters for half and quarter resolutions, respectively. In the refinement stage, the estimation error is feed-forwarded to improve the estimation accuracy. Experimental results show that the proposed technique can estimate the bit rate accurately at a wide range of down-sampling ratios.

I. INTRODUCTION

The emergence of wireless networks has posed many new challenges for multimedia content providers. In particular, many existing pre-encoded video contents may have bit-rates and resolutions far exceeding the capability of wireless networks and mobile devices, and they need to be adapted to the communication technology being used. This can be done through the use of a video transcoder which converts high resolution, high bit-rate videos to low resolution, low bit-rate versions of the videos.

A video transcoder can reduce the bit-rate using three techniques: re-quantization, spatial down-sampling and frame skipping [2-7]. As transcoding may need to be performed on-the-fly on streamed video, the ability to accurately predict the output bit-rate is essential. For re-quantization, He *et al.* in a pioneering work [1] introduced the ρ-domain approach for rate estimation, where ρ is defined to be the percentage of zeros among the quantized coefficients in the transformed domain. He *et al.* showed that ρ is linearly related to the number of bits to encode a frame and the bit rate of the video can be accurately estimated using ρ-domain information.

Relatively little effort has been made for the rate estimation of spatio-temporally down-sampled video. Existing works on spatial down-sampling focus on motion vectors re-estimation [3-4] and techniques to perform down-sampling in the DCT domain [5-6]. In this work we extend the ρ-domain rate estimation technique to video down-sampling within the framework of the emerging H.264/AVC video standard [8].

The proposed algorithm consists of two stages. In the initialization stage, we use a low-complexity algorithm to estimate the ρ values for resolutions with down-sampling ratio 0.5 and 0.25. In the refinement stage, the estimation error is feed-forwarded to improve the estimation accuracy. Experimental results show that the proposed technique can estimate the bit rate at a wide range of down-sampling ratios with a good accuracy.

II. RATE ESIMATION

Our experiments on spatial down-sampling reveal a surprising relation between bit-rate and sub-sampling ratio. For example, using the 'Foreman' test sequence originally encoded with the H.264/AVC JM61e codec, we decode the video, sub-sample it, and then re-encode it using the same quantization parameter QP for different sub-sampling ratios and plot the results in Fig. 1.

Much to our surprise, the results show that the video bit rate decreases near linearly with the sub-sampling ratio. The same results are observed for different video resolutions as well as different QP's on various test sequences. This motivates us to develop a low-complexity technique to estimate the proportionality constant (i.e., slop of the line) without actually carrying out the full down-sampling procedure.

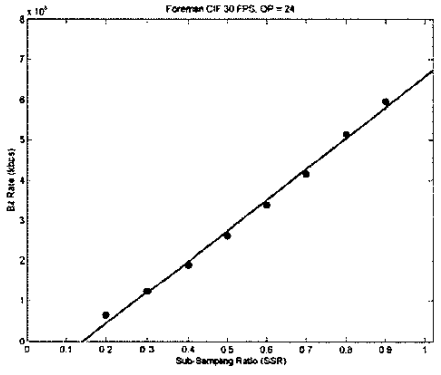

Fig. 1. Bit rate of down-sampled video at QP=24.

0-7803-8554-3/04/$20.00 ©2004 IEEE.

A. Rate Estimation Using ρ-Domain Information

The ρ value at the full resolution can be obtained directly from the quantized DCT coefficients in the H.264/AVC bit stream. We estimate the ρ values at lower resolutions by applying low pass filtering and sub-sampling in the DCT domain. The motion vectors of the sub-sampled video may not be the same as the full resolution counter-parts, and the residual signal can be affected by the motion re-estimation and the motion compensation after the sub-sampling. However, the motion re-estimation is computationally too expensive and thus we simply sub-sample the residual signal in the DCT compressed domain. Only luminance component is used to estimate the bit rate, since chrominance components usually need a substantially fewer number of bits to encode.

To estimate the ρ-domain information for sub-sampling ratio SSR=0.5, each 8×8 luminance block is low passed, sub-sampled and then quantized to give a 4×4 quantized DCT block D_s as shown in Fig. 2. D_0, D_1, D_2 and D_3 are quantized 4×4 DCT blocks extracted from the original H.264/AVC bit stream.

Let X_1, X_2, X_3 and X_4 denote the spatial domain of D_0, D_1, D_2 and D_3 respectively. Following the symbols in [9], X_1, X_2, X_3 and X_4 can be written as

$$X_k = \left(C_i^T D_k' C_i + 2^5 E\right) >> 6 \qquad k = 0,1,2,3 \tag{1}$$

where

$$C_i = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 1/2 & -1/2 & -1 \\ 1 & -1 & -1 & 1 \\ 1/2 & -1 & 1 & -1/2 \end{pmatrix}, \tag{2}$$

$$D_k'(i,j) = D_k(i,j) B(Q_M, i, j) << Q_E, \tag{3}$$

$Q_M = QP \bmod 6$, $Q_E = \lfloor QP/6 \rfloor$, and E is a 4×4 matrix with all elements equal to 1. The reconstruction factor $B(Q_M,i,j) = S(Q_M,r)$ where $r = 0$ for $(i,j) \in \{(0,0), (0,2), (2,0), (2,2)\}$, $r = 1$ for $(i,j) \in \{(1,1), (1,3), (3,1), (3,3)\}$, and $r = 2$ otherwise. The matrix S is defined as

$$S = \begin{pmatrix} 10 & 11 & 13 & 14 & 16 & 18 \\ 16 & 18 & 20 & 23 & 25 & 29 \\ 13 & 14 & 16 & 18 & 20 & 23 \end{pmatrix}^T. \tag{4}$$

X_1, X_2, X_3 and X_4 are then grouped to form a 8×8 spatial block X, which is in turn low passed and sub-sampled to form a 4×4 spatial block Y:

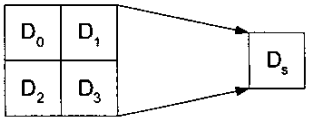

Fig. 2. Sub-sampling of 8x8 block for SSR=0.5.

$$X = P_0 X_0 P_0^T + P_0 X_1 P_1^T + P_1 X_2 P_0^T + P_1 X_3 P_1^T \tag{5}$$

where

$$P_0 = \begin{bmatrix} I_4 & 0 \end{bmatrix}^T, \quad P_1 = \begin{bmatrix} 0 & I_4 \end{bmatrix}^T, \tag{6}$$

and

$$Y = AFXF^T A^T. \tag{7}$$

Note F and A are the low pass filtering matrix and sub-sampling matrix. To reduce the computational complexity, we choose a simple low pass filtering matrix:

$$F = \begin{pmatrix} 1/2 & 1/2 & \cdots & 0 & 0 \\ 0 & 1/2 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1/2 & 1/2 \\ 0 & 0 & \cdots & 0 & 1/2 \end{pmatrix}, \tag{8}$$

$$A = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \end{pmatrix}. \tag{9}$$

The forward 4×4 DCT is applied to block Y to yield D_Y, which is quantized using the same QP of the original video to give the quantized block D_s. That is,

$$D_Y = C_f Y C_f^T, \tag{10}$$

$$D_s(i,j) = sign\{D_Y(i,j)\} \left[\left(|D_Y(i,j)| A(Q_M,i,j) + f 2^{15+Q_E} \right) >> (15+Q_E) \right], \tag{11}$$

$$C_f = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 2 & 1 & -1 & -2 \\ 1 & -1 & -1 & 1 \\ 1 & -2 & 2 & -1 \end{pmatrix}. \tag{12}$$

where f is the dead zone parameter, which is 1/3 for intra frames and 1/6 for inter frames. Also, the quantization factor $A(Q_M,i,j) = M(Q_M,r)$ is defined as the reconstruction factor B in (3). The matrix M is given by

$$M = \begin{pmatrix} 13107 & 11916 & 10082 & 9362 & 8192 & 7282 \\ 5243 & 4660 & 4194 & 3647 & 3355 & 2893 \\ 8066 & 7490 & 6554 & 5825 & 5243 & 4559 \end{pmatrix}^T \quad (13)$$

The total number of zero coefficients in frame n, $\rho(n,0.5)$, is obtained by summing up all the numbers of zero coefficients in all the quantized 4×4 block.

It is computationally expensive to obtain D_Y. To reduce the computational complexity, we combine (1), (5), (7) and (10) to obtain the D_Y:

$$D_Y = \begin{pmatrix} B_0 D_0' B_0^T + B_0 D_1' B_1^T \\ + B_1 D_2' B_0^T + B_1 D_3' B_1^T \end{pmatrix} >> 10 \quad (14)$$

where

$$B_0 = 4 \cdot C_f A F P_0 C_i^T = \begin{pmatrix} 8 & 0 & 0 & 0 \\ 12 & 3 & 0 & -1 \\ 0 & 6 & 0 & -2 \\ -4 & 9 & 0 & -3 \end{pmatrix} \quad (15)$$

and

$$B_1 = 4 \cdot C_f A F P_1 C_i^T = \begin{pmatrix} 8 & 0 & 0 & 0 \\ -12 & 3 & 0 & -1 \\ 0 & -6 & 0 & 2 \\ 4 & 9 & 0 & -3 \end{pmatrix}. \quad (16)$$

It is worthy to point out that we can further reduce the computational complexity since there are many zeros in matrices B_0 and B_1.

To estimate the ρ value for SSR=0.25, the DC coefficients of 16×16 contiguous macro blocks are extracted to form a 4×4 block. . Then, it is processed by the forward DCT and the quantization to make the block D_s, as shown in Fig. 3. The number of zero coefficients $\rho(n,0.25)$ is obtained by counting zero coefficients in all the quantized 4×4 blocks.

Armed with the estimated number of zero coefficients, we can then use this ρ-domain information to estimate the output bit rate. Specifically, the estimated number of bits $R_1(n,SSR)$ to encode frame n with sub-sampling ratio SSR is given by

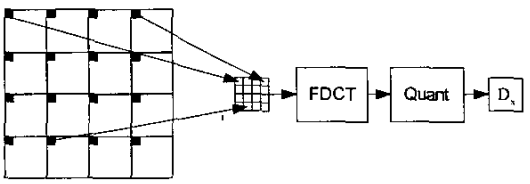

Fig. 3. Sub-sampling of 16x16 block for SSR=0.25.

$$R_1(n,SSR) = \frac{R_0(n,1)}{1-\rho(n,1)} \cdot \begin{bmatrix} 4(\rho(n,0.25) - \rho(n,0.5)) \cdot SSR \\ + \rho(n,0.5) - 2\rho(n,0.25) + 1 \end{bmatrix}, \quad (18)$$

where $R_0(n,1)$ is the actual number of bits used to encode frame n at the full resolution. Both $R_0(n,1)$ and $\rho(n,1)$ can be obtained directly from the original H.264/AVC bit stream.

B. Rate Estimation Refinement

The previous technique enables us to estimate the output bit rate without actually carrying out the down-sampling process. However, our experiments show that the estimation becomes less accurate for very small sub-sampling ratios (e.g. SSR=0.2). To overcome this problem, we propose a prediction error feed-forward technique to compensate for the estimation errors.

Specifically, we adjust the estimated bit rate by adding the estimation error from the previous frame to the current frame:

$$R_2(n,SSR) = R_1(n,SSR) + (R_0(n-1,SSR) - R_1(n-1,SSR)) \quad (19)$$

As the estimation error is fairly consistent at small SSR values, this technique can effectively reduce the estimation error to negligible levels.

III. RESULTS AND DISCUSSIONS

The performance of the proposed rate estimation algorithm is evaluated on various video sequences. Due to the limited space, only parts of the results on the 'Foreman' CIF (352×288, 30 frames/s) sequence are provided in this paper. The simulation is run for 10 seconds of video (300 frames) with QP={24, 36} and SSR={0.9, 0.8, 0.7, 0.6, 0.5, 0.4, 0.3, 0.2}.

Table 1 compares the total actual and estimated numbers of bits for the 300 frames. Although the estimation error still increases with smaller SSR, the maximum estimation error is only 2.4% (QP=36, SSR=0.2).

	QP=24		QP=36	
SSR	Actual	Estimated	Actual	Estimated
0.9	5969984	5981533	936760	935559
0.8	5158872	5168940	793440	792755
0.7	4166768	4176522	653112	652587
0.6	3384816	3392202	534504	533538
0.5	2620168	2627358	399608	398678
0.4	1895000	1899478	305856	304534
0.3	1240456	1241081	210552	208645
0.2	644808	640185	123216	120322

Table 1. Total number of bits for 300 frames.

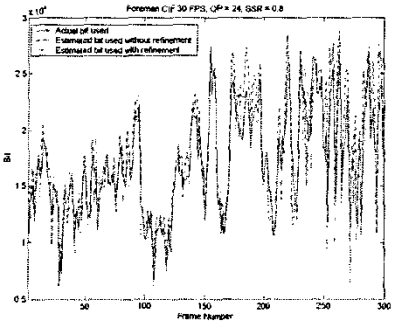

Fig. 4. Rate estimation results for QP=24, SSR=0.8.

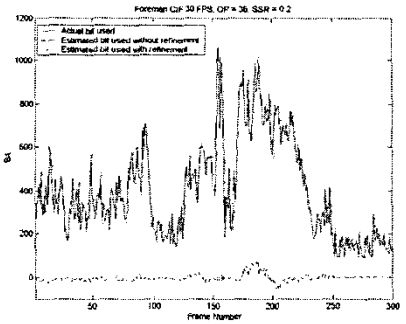

Fig. 5. Rate estimation results for QP=36, SSR=0.2.

Figs. 4 and 5 further illustrate the estimation performance at large (0.8) and small (0.2) SSR's respectively. With QP=24 and SSR=0.8 in Fig. 4, the initial estimation is very close to the actual number of bits used. The improvement achievable by feed-forwarding the prediction error is negligible.

However with QP=36 and SSR=0.2 in Fig. 5, the initial estimation error is substantial because the video is now compressed at a very low bit rate. But despite the estimation error, the predicted bit rate curve still mimics the shape of the actual bit rate curve. Thus after applying the feed-forward technique presented in Section II-B, the prediction errors are effectively reduced to negligible levels. Note that while the feed-forward technique is very effective, the initial rate estimation using ρ-domain information is also important. For example, if we skip the initial rate estimation process and simply use the previous frame's bit rate as the bit rate for the current frame, the estimation accuracy will drop substantially.

IV. CONCLUSIONS

In this paper, we proposed a new technique to estimate the output bit rate of down-sampled video. The technique comprises (a) low-complexity ρ-domain information estimation; (b) bit rate estimation from the ρ-domain information; and (c) prediction error feed forward scheme to achieve accurate rate estimation. The authors are now extending this work to frame skipping, ultimately to develop a model relating visual quality to all three transcoding techniques, *i.e.* re-quantization, spatial down-sampling, and frame skipping.

ACKNOWLEDGMENT

The authors want to express their gratitude to the anonymous reviewers for their insightful comments and suggestions in improving this paper. This research is funded in part by a Direct Grant, and Earmarked Grant (CUHK 4209/01E, CUHK 4328/02E) from the HKSAR Research Grant Council and in part by the Area of Excellence Scheme, established under the University Grants Council of the Hong Kong Special Administrative Region, China (Project No. AoE/E01/99).

REFERENCES

[1] Z. He and S. K. Mitra, "A unified rate-distortion analysis framework for transform coding" *IEEE Trans. Circuits Syst. Video Technol.* vol. 11, pp. 1221-1236, Dec. 2001.

[2] A. Vetro, C. Christopoulos, and H. Sun, "Video transcoding architectures and techniques: an overview", *IEEE Signal Proc. Magazine*, vol. 20, pp. 18-29, Mar. 2003.

[3] T. Shanableh and M. Ghanbari, "Heterogeneous video transcoding to lower spatio-temporal resolutions and different encoding formats," *IEEE Trans. Multimedia*, vol. 2, pp. 101-110, June 2000.

[4] Y. Q. Liang, L. P. Chau and Y. P. Tan, "Arbitrary downsizing video transcoding using fast motion vector reestimation," *IEEE Signal Proc. Letters*, vol. 9, pp. 352-355, Nov. 2002.

[5] P. A. A. Assunção and M. Ghanbari, "A frequency-domain video transcoder for dynamic bit-rate reduction of MPEG-2 bit streams," *IEEE Trans. Circuits Sys. Video Technol.*, vol. 8, pp. 953-967, Dec. 1998.

[6] S. Liu and A. C. Bovik, "Local bandwidth constrained fast inverse motion compensation for DCT-domain video transcoding," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 12, pp. 309-319, May 2002.

[7] J. W. Lee, A. Vetro, Y. Wang and Y. S. Ho, "Bit allocation for MPEG-4 video coding with spatio-temporal tradeoffs," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 13, pp. 488-502, June 2003.

[8] T. Wiegand, G. J. Sullivan, G. Bjøntegaard and A. Luthra, "Overview of the H.264/AVC video coding standard," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 13, pp. 560-576, July 2003.

[9] H. S. Malvar, A. Hallapuro, M. Karczewicz and L. Kerofsky, "Low-complexity transform and quantization in H.264/AVC," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 13, pp. 598-603, July 2003.