Key Techniques of Bit Rate Reduction for H.264 Streams

Peng Zhang, Qing-Ming Huang, and Wen Gao

Institute of Computing Technology, Chinese Academy of Science, Beijing, 100080, China
{peng.zhang, qmhuang, wgao}@jdl.ac.cn, http://www.jdl.ac.cn

Abstract. In previous techniques of bit rate reduction transcoding, reusing the mode of the input MB is widely adopted. However, directly re-using the mode of input MB will cause additional losses for H.264 bit-rate scaling. Variable-size block-matching based motion estimation makes H.264 much more efficient than other coding standards while increasing its complexity. As the target bit rate changes, the best mode of the MB may also change. In this paper we concerned two key techniques for H.264 bit rate reduction transcoding: Mode decision and rate control. A fast MB mode decision algorithm and an improved rate control algorithm depending on statistics of input streams is proposed.

1 Introduction

Bit-rate scaling is a key solution to media adaptation in hybrid networks. Cascaded decoder-encoder is a straightforward approach which can achieve the best quality at the cost of high complexity. In the literature, some papers proposed different techniques to get a trade-off between acceptable quality and complexity over the past few years in which the focus has been centered on two specific aspects, complexity and drift compensation [1-4]. Generally, two architectures, close-loop and open-loop, are widely in use. Because open-loop systems are relatively simple but subject to drift-error, close-loop systems are more common in practical. [1] proposed an close-loop transcoder which requires one DCT and one IDCT, while cascade decoder-encoder requires one DCT and two IDCT. This architecture achieved nearly the same quality as cascaded decoder-encoder with some arithmetic inaccuracy introduced. All of these transcoding schemes are based on the assumption of fixed MB size for motion compensation. This assumption is effective to many coding standards, such as MPEG-2, MPEG-4 and H.263.

Variable-size block-matching (VSBM) based motion compensation is a key feature of H.264/AVS [5] which ensures its high coding efficiency. [6] proposed a Lagragian mode decision technique by looping all available mode to select one with minimized rate-distortion cost. It is a computationally heavy procedure. However, directly reusing the mode of the input macroblock (MB) will cause severe quality losses because the best mode varies when its QP changes. Table-1

K. Aizawa, Y. Nakamura, and S. Satoh (Eds.): PCM 2004, LNCS 3332, pp. 985–992, 2004.

Table 1. Mode variations for different QP values, FOREMAN, CIF, 60 frames (ROW: QP = 32; COL: QP = 28)

	SKIP	P-16X16	P-16X8	P-8X16	P-8X8	I-4X4	I-16X16
SKIP	4222	658	80	80	7	10	67
INTER-16X16	1933	3155	452	470	97	56	96
INTER-16X8	347	783	696	200	148	27	33
INTER-8X16	525	1006	234	921	151	36	34
INTER-8X8	148	624	535	597	1291	64	10
INTRA-4X4	26	107	79	80	51	2071	394
INTRA-16X16	158	53	14	10	3	101	820

demonstrates the variation of mode at different QP, where the column is the mode distribution at the QP 32 and the row is at QP 28. From Table-1, we can see that the best mode may change as the QP value changes. In general, in the case of bit rate reduction transcoding for H.264 streams, the first step is to decide a proper QP for the target rate, then decision of MB mode for MC must be made. In this paper, two key techniques, rate control and mode decision, are concerned.

This paper is organized as follows: in section II, a mode classification approach is introduced and simplified rate-distortion optimization based on this mode classification is proposed. In section III, we propose a new rate control algorithm which fully utilizes the information from the input stream. Section IV are experimental result of our transcoding scheme and two comparison transcoders.

2 Mode Classification for Bit-Rate Scaling Transcoding

As we know, MPEG-2 [7] also supports variable block size motion compensation. Macroblcoks of P frame can be coded in INTRA, INTER-16x16 or SKIP mode, and INTER-16x8 mode for interlaced videos. The small mode set makes it unnecessary to re-select modes in transcoding. H.263 [8] specifies INTER-8x8 coding mode. H.263+ adds some improvements in compression efficiency for the INTRA macroblock mode. Motion compensation in MPEG-4 [9] is based on 16x16 blocks and support INTER-8x8 mode. Also MPEG-4 includes alternate scan patterns for horizontally and vertically predicted INTRA blocks. However, the mode sets of these coding standards are rather small, and the reuse of the previous mode in transcoding has little effect on the quality.

H.264 supports variable prediction mode for motion estimation. For I frames, macroblocks can be coded in INTRA-16x16 or INTRA-4x4 modes, and furthermore, there are 9 prediction directions for INTRA-4x4 mode. For P frames, mode of target macroblock must be selected from SKIP, INTER-16x16, INTER-16x8, INTER-8x16 and INTER-8x8 mode respectively besides INTRA-16x16 and INTRA-4x4 modes. In case the INTER-8x8 macroblock mode is chosen, each 8x8 block can be further partitioned into blocks of 8x8, 8x4, 4x8 or 4x4 luminance samples. B frames are similar to P frames. Also, H.264 support multi-frame motion-compensated prediction. When bit-rate scaling, the best mode and best reference picture of target macroblock under the new rate may differ with those under its original rate. To maintain the quality, we classified the modes into different levels.

We know that the Lagrangian coder control will assign more INTER-8x8 or INTRA-4x4 mode to active regions and SKIP or INTER-16x16 to static backgrounds. That means active MB' mode level is high while static MB's mode level is low. So, we ordered the modes from low level to high level as SKIP, INTER-16x16, INTER-16x8, INTER-8x8 and INTRA-4x4. In the case of QP increases, the possible mode set follows the rule that the output mode order will not higher than the input mode order. For example, if the input mode is INTER-16x16, when the QP increases, the new best mode will be selected from modes whose level is lower than INTER-16x16. So, we classified all these modes into five classes for bit-rate scaling transcoding: 1.{SKIP, INTER-16x16};2.{INTER-16x8, INTER-8x16};3.{INTER-8x8};4.{INTRA-4x4};5.{INTRA-16x16}.

Based on this classification, we propose our mode decision algorithm. This mode decision algorithm decide the possible mode options for RDO which effectively restricts the Lagrangian method loops in a small range. Detailed algorithm is like the following pseudo-codes:

```
Switch( input macroblock mode)
{
case SKIP:
        Target mode set = {SKIP, INTER-16x16};
case INTER-16x16:
        Target mode set = {SKIP, INTER-16x16};
case INTER-16x8:
        Target mode set = {SKIP, INTER-16x16, INTER-16x8};
case INTER-8x16:
        Target mode set = {SKIP, INTER-16x16, INTER-8x16};
case INTER-8x8:
        Target mode set = {SKIP, INTER-16x16, INTER-16x8,
                           INTER-8x16,INTER-8x8};
case INTRA-4x4:
        Target mode set = {INTRA-4x4};
case INTRA-16x16:
        Target mode set = {INTRA-4x4, INTRA-16x16, SKIP};
}
```

After mode decision, different motion compensation scheme is used for different modes. If one MB's mode is decided as INTER-16x16, MV can be reused and traditional transcoding techniques can be directly adopted. Otherwise, if target modes includes INTER-16x8, INTER-8x16 or INTER-8x8, motion vectors refinement is needed. Because in most sequences, SKIP and INTER-16x16 modes possess a large percent, the above algorithm can efficiently reduce the

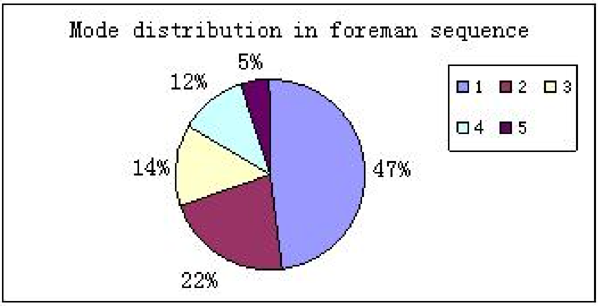


Fig. 1. 1: INTER-16x16 and SKIP mode, 2: INTER-16x8 and INTER-8x16 mode; 3: INTER-8x8 mode, 4: INTRA-4x4, 5: INTRA-16x16 mode

computation complex while maintaining the acceptable quality. Furthermore, to keep the quality of areas with high activity, in case the input macroblock mode is INTRA-4x4, the target macroblock mode is INTRA-4x4.

Figure 1 illustrates the computational complexity participation in the case of Table-1 as QP equals to 28. From figure 1, we can see that about 65% macroblocks are INTRA or INTER-16x16 modes which doesn't require RDO and motion estimation and only 22% macroblocks' mode can be selected from INTER-16x8, INTER-16x16 or INTER-8x16, INTER-16x16, and only 14% macroblocks require looping all INTER modes. So, the computational complexity is greatly reduced by this method. Also, this classification can be helpful in INTRA-Refresh technique for error-resilient transcoding to convert macroblocks with high activity into INTRA modes.

3 Rate Control Algorithm for H.264 Transcoding

Video encoding rate control is important to achieve consistent video quality and has been the interest of research in recent years [10-13]. As we know, the key role of R-Q model is to determine the QP before encoding one frame by available channel bandwidth, output buffer fullness, and picture and motion complexity. There are two problems in rate control: the first one is target frame bits allocation, and the second is the estimation of quantization parameters. [12] proposed and improved method named "MAD ratio" for target frame bits allocation. And in [14], the author proposed a two-pass QP prediction algorithm based on TM5. In video transcoding, the statistics such as number of bits and QP can be easily derived from the stream. We can compute the complexity of the picture and target bit allocation based on these statistics. [15] proposed a bit allocation algorithm based on the fact that the ratio of the complexity between input and output pictures keeps constant. However, this algorithm needs the information of the whole GOP, and is not fit for the sequences in which the content changes severely. In this paper, the target frame bit number depends on the actual coded bits number of the input frame and ratio of target rate and input rate, and then QP can be predicted based on the first-order R-Q model. This rate control method can be adopted in frame level, object level and macroblock level.

In our rate control algorithm, the fluid traffic model is applied the same as that in [6]. Let $R_i(j)$, $B_i(j)$, $V_i(j)$ and $b_i(j)$ denote the instant available bit rate, total bits for the rest pictures, the occupancy of the virtual buffer and the actual generated bits in j^{th} picture respectively, where i and j means j^{th} picture in i^{th} GOP to be coded. And N_i denotes the total number of pictures in i^{th} GOP, f is the frame rate.

$$\begin{cases} \frac{R_i(j)}{f} \times N_i - V_i(j) & j = 1 \\ B_i(j-1) + \frac{R_i(j)-R_i(j)}{f} \times (N_i - j + 1) - b_i(j-1) & j = 2, 3, .., N \end{cases} \tag{1}$$

$$\begin{array}{l} V_i(1) = \begin{cases} 0 & j = 1 \\ V_{i-1}(N_{i-1}) & other \end{cases} \\ V_i(j) = V_i(j-1) + b_i(j-1) - \frac{R_i(j-1)}{f} \quad j = 2, 3, .., N \end{array} \tag{2}$$

The determination of target bits for current P frame is composed of four steps as the following:

Step 1: Compute the target bits, this step can be divided into 2 or 3 sub-steps depends on whether frame level or basic unit level rate control is applied.

Sub-step 1: Determine target buffer level for current P picture

$$\begin{array}{l} S_i(2) = V_i(2) \\ S_i(j+1) = S_i(j) - \frac{S_i(2)}{Np(i)-1} \quad j = 2, 3, .., N_i \end{array} \tag{3}$$

Sub-step 2 Compute the target for current P picture The target bits allocated for the j^th P picture is determined based on the target buffer level, the frame rate, the available channel bandwidth, and the actual buffer occupancy as follows:

$$T_{buf} = \frac{R_i}{f} + \gamma \times (S_i(j) - V_i(j)) \tag{4}$$

Meanwhile, the remaining bits are also computed as:

$$T_r = \frac{B_{org}(j)}{N_j \times (R_{org}/f) - \sum_{k=0}^{j-1} b_{org}(k)} \times B_i(j) \tag{5}$$

Where $b_{org}(j)$ and $R_{org}(j)$ are actual input bits of the j^{th} picture and the channel bandwidth of the input stream. The final target bit T is a weighted combination of T_{buf} and T_r

$$T = \beta \times T_r + (1 - \beta) \times T_{buf} \tag{6}$$

Where β is a weighting factor and its typical value is 0.5.

Our proposed improvement here: In [6], T_r is computed as follows:

$$T_r = \frac{W_{p,i}(j-1)}{W_{p,i}(j-1) \times N_{p,r} + W_{b,i}(j-1) \times N_{b,r}} \times B_i(j) \tag{7}$$

Where $W_{p,i}(j)$ and $W_{b,i}(j)$ are the complexities of previous P picture and previous B picture respectively. This weighted factor will not be accurate when the content of stream varies severely. In (5), we use ratio of the actual input bits of current picture and the remaining bits of the input GOP as the weighted factor. This factor represents the attributes of the current picture to the GOP which will be much more accurate. In case of basic unit level rate control, the sub-step 3 is to compute the target bits for each unit as follows:

$$b_l = \frac{(b_{l,org} - c_l) \times QP_{l,org}}{\sum_{k=l}^{N_{unit}} (b_{k,org} - c_k) \times QP_{l,org}} \times T_l \tag{8}$$

Where b_l is the target bits for the l^{th} unit, $b_{l,org}$, c_l and $QP_{l,org}$ are the actual bits, header bits and QP of the l^{th} unit in the input stream respectively. In this step, we use a first-order R-Q model similar to [10], in which the complexity can be computed based on bits and QP like follows:

$$X = (R - c) \times QP \tag{9}$$

Where X denotes the complexity, R denotes the total bits of the unit, and constant C is the header bits.

Step 2: After computing the target bit, QP can be computed as follows:

In case of frame level rate control:

$$Q_{step} = \frac{(T_{org} - h_{org}) \times Q_{step,org}}{T - h_{org}} \tag{10}$$

In case of basic unit level rate control:

$$Q_{step,l} = \frac{(b_{l,org} - h_{l,org}) \times Q_{step,l,org}}{b_l - h_{l,org}} \tag{11}$$

In this step, we use the header bits of input picture or basic unit as the prediction of header bits for the current picture or basic unit.

Step 3: Perform RDO for all MBs in the current basic unit and code them by H.264.

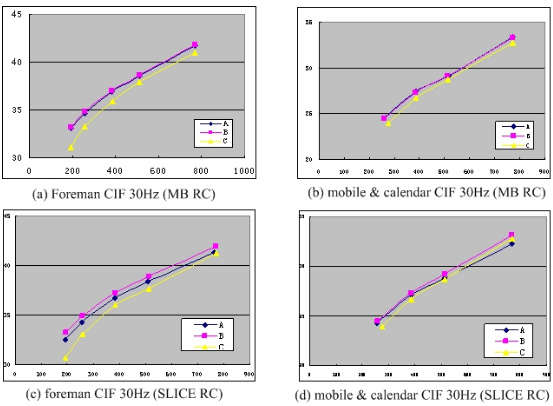
Step 4: Update the parameters

4 Experiment Results

Our experiments are based on JM7.6 codec. For simplicity, we implement a cascaded decoder-encoder transcoder with our mode-decision algorithm and rate-control algorithm. To compare the experiments results, two cascaded decoder-encoder also implemented where one of them fully decode the input stream and then re-encodes the reconstructed signals into the stream of target bit rate with RDO, and the other reuses the MB mode of input stream without RDO. We applied two types of input streams: the first one is CIF, 30Hz, 1024kbps, and

Table 2. Bit-Rate Scaling Results for Video Transcoding

Transcoding		Cascaded Decode-Encoder with RDO		Cascaded Decoder-Encoder without RDO	
Bit rate	PSNR	Bit rate	PSNR	Bit rate	PSNR
Foreman, CIF, 30Hz, 1024kbps, MB level rate control					
192.15	33.08	192.58	33.21	192.72	31.05
256.23	34.59	256.61	34.79	256.97	33.24
384.48	36.9	384.71	37.02	385.14	35.95
512.76	38.51	513	38.66	513.15	37.94
768.63	41.69	769.2	41.86	769.33	41.01



(a) Foreman CIF 30Hz (MB RC) (b) mobile & calendar CIF 30Hz (MB RC)

(c) foreman CIF 30Hz (SLICE RC) (d) mobile & calendar CIF 30Hz (SLICE RC)

Fig. 2. R-D curves for MB and SLICE level rate control for selected sequences A: Our transcoding scheme, B: Cascaded Decoder-Encoder with RDO, C: Cascaded Decoder-Encoder without RDO

the other is QCIF, 15Hz, 384kbps. In this experiment, for simplicity, B frames are not considered. The input streams are transcoded into different bit rates.

Table 2 shows results for a set of test sequences and test conditions seleceted to represent a bit-rate scaling transcoding application. The result of our transcoding scheme is close to that of cascaded decoder-encoder with RDO, and re-using mode without RDO have about 1dB loses. These results also prove our rate control algorithm effective.

R-D curves for MB and SLICE level rate control for selected sequences are plotted. From fig.2 we can see that the losses of cascaded decoder-encoder without RDO increases as long as the target bit-rate decreases, while the results of our transcoding scheme performs rather well at low bit-rate.

5 Conclusion

Experimental results show that RDO is necessary in H.264 transcoding. Our fast mode decision algorithm notably reduces the complexity of RDO. Also experiments show that our rate control algorithm is simple but effective especially when transcoding the input streams into low bit-rate streams. MB level rate control performs better than frame level and slice level rate control.

Acknowledgement. This research has been partially supported by NSFC under contract No. 60333020, "863" project under contract No. 2002AA118010, "973" project under contract No. 2001cca03300 and by Bairen project of CAS.

References

1. P. Assunao and M. Ghanbari, "Post-processing of MPEG-2 coded video for transmission at lower bit-rates," in Proc. IEEE Int. Conf. Acoustics, Speech and Signal Processing, Atlanta, CA, 1996, pp. 1998-2001.
2. Huifang Sun, Wilson Kwok and Joel W. Zdepski, "Architectures for MPEG compressed Bitstream Scaling," IEEE Trans. CSVT, Vol. 6, No. 2, April 1996
3. Jeongnam Youn, Ming-Ting Sun, and Jun Xin, "Video Transcoder Architectures for Bit Rate Scaling of H.263 Bit Streams", Lujun Yuan, Huifang Sun and Wen Gao, "MPEG TRANSCODING FOR DVD RECORDING", ICICS PCM, Dec. 2003.
4. S. F. Chang and D. G. Messerschmidt, "Manipulation and composing of MC-DCT compressed video," IEEE J. Select. Areas Commun., vol. 13, pp. 1-11, Jan. 1995
5. "Draft ITU-T recommendation and final draft standard of joint video specification (ITU-T Rec. H.264/ISO/IEC 14496-10 AVC" in Joint Video Team (JVT) of ISO/IEC MPEG and ITU-T VCEG, JVT G050, 2003
6. ISO/IEC TC JTC 1/SC 29 N5821
7. ISO/IEC 13818-2 (MPEG-2), 1994
8. ITU-T Recommendation H.263, 1995
9. "MPEG-4 Video Verification Model 18.0 (VM-18)," ISO/IEC JTC1/SC29/WG11, Doc. MPEG-N3908, 2001
10. Test Model 5, www.mpeg.org/MPEG/MSSG/tm5
11. T. Chiang and Y.-Q. Zhang, "A New Rate Control Scheme using a New Rate-Distortion Model," IEEE Trans. CSVT, pp. 246-250, Feb. 1997
12. Mingqiang Jiang, Xiaoquan Yi, and Nam Ling, "Improved Frame-Layer Rate Control for H.264 Using MAD Ratio,"
13. Z. Li, F. Pan, K. P. Lim, G. Feng, X. Lin and S. Rahardja, "Adaptive basic unit layer rate control for JVT," JVT-G012-r1, 7th Meeting, Pattaya II, Thailand, Mar. 2003 14. Siwei Ma, Wen Gao, Feng Wu, and Yan Lu, "Rate Control fro H.26L,"