J. Indian Inst. Sci., July–Aug. 2006, **86**, 407–413

Low-complexity rate-control algorithm for hybrid video coding

V. V. GOHOKAR[1,]* AND S. P. KSHIRSAGAR[2]
[1]Electronics Engineering Department, Shri Sant Gajanan Maharaj College of Engineering, Shegaon 444 203, India.
[2]Conquest Integrated Technologies Pvt Ltd, Pune 415 001, India.
email: vvgohokar@rediffmail.com; Fax: 91-7265-252346.

Received on March 7, 2006; Revised on August 10, 2006.

Abstract

Design and operation of an encoder involves the optimization of many decisions to achieve the best possible trade-off between rate and distortion under the given constraint on delay and complexity. In this paper, a new rate-control algorithm is proposed by optimal bit allocation and macroblock mode selection. It is low-complexity, single-pass algorithm applicable in areas that are sensitive to delay and having very limited resources. Improvement in PSNR is obtained for video sequences with varying contents and bit rate. The algorithm does not violate the buffer constraints.

Keywords: Distortion, mpeg2, macroblock, motion estimation.

1. Introduction

Digital video communication can be found today in many applications such as video-on-demand, video conferencing, broadcasting and Internet. The basic communication problem may be posed as conveying source data with the highest possible fidelity within the available bit rate. Conveying the source data using the lowest bit rate possible while maintaining a specified reproduction quality decides its rate distortion performance.

International video coding standards that have been developed in the last decade offer very high compression ratios, allowing transmission, storage and manipulation of visual information in various environments. For achieving high compression ratios, hybrid video coding is used. Traditional still image compression methods based on DCT and quantisation are adopted to reduce spatial redundancy. Temporal compression is done by methods such as motion-compensated prediction and frame skipping.

A hybrid video encoder typically processes as follows. Each picture is split into blocks. The first picture of video sequence is typically coded in intra mode. For all remaining pictures of the sequence or group of pictures (GOP), inter picture coding is used. The encoding process for inter prediction consists of choosing motion data comprising the selected reference picture and motion vector (MV) to be applied. The motion and mode decision data, transmitted as side information, is used by encoder and decoder to generate identical inter-

*Author for correspondence.

prediction signal using motion compensation (MC). The residual of intra- or inter-prediction, which is the difference between the original block and its prediction is transformed by a frequency transform. The transformed coefficients are then scaled, quantised, entropy coded and transmitted together with the prediction side information. The encoder duplicates the decoder processing so that both will generate identical predictions for subsequent data. The quantised transformed coefficients are then inverse-transformed to duplicate the decoded prediction residual. The residual is then added to the prediction and the result of that addition may then be stored for the prediction of subsequent encoded pictures [1].

The design and operation of an encoder involves the optimization of many decisions to achieve the best possible tradeoff between rate and distortion, given the constraint on delay and complexity. The task is complicated by the fact that the various coding options show varying efficiency at different bit rates or levels of fidelity and scene contents. In order to obtain an optimal encoder, focus is given to image preprocessing, motion estimation, coding mode decision and rate control. Design of rate control algorithm should ensure that the video buffers do not overflow or underflow. If this condition is not satisfied video information can be lost and thus the reconstructed video at the decoder is seriously damaged [2].

Among the encoder modules, motion estimation (ME) module produces both motion vector and prediction residual for each macroblock. Most encoder implementations have usually been designed with a certain motion search strategy where a displacement vector that produces a minimal energy residual is selected to represent the MB motion. With such a strategy, the rate of the residual will be minimum but a large portion of the bit rate is consumed by coding transform coefficients. Coding the motion information also requires certain number of bits that becomes relatively substantial as the overall bit rate decreases.

Another important block in the encoding process is the determination of the coding mode for a given input source. As video sequence contains widely varying contents, the encoding performance can be improved if different strategies are permitted to code different regions. Existing hybrid coders utilize several modes of operations that are selected on the MB-by-MB basis [3, 4].

The method presented in this paper is a one-pass method in which the selection of field or frame mode at macroblock level is done by considering the activity of the scene. The very first step, discussed in Section 2, is the analysis of MPEG2 encoder at macroblock level. An algorithm using new method for macroblock mode selection is developed in Section 3. The quantisation parameter is modified depending on the macroblock type and the activity analysis. Test results using proposed algorithm are given in Section 4. Conclusions are presented in Section 5.

2. Mode analysis of MPEG2 standard

In MPEG2, the standard that uses TM5 for rate control, and motion estimation mode is decided by selecting the mode with the lowest residual energy. Given the motion estimate intra/inter decision is made next, based on the luminance prediction error. With a given picture-level mode, there are various macroblock modes that are associated with it. The relationship between the picture- and the macroblock-coding modes is as shown in Fig. 1.

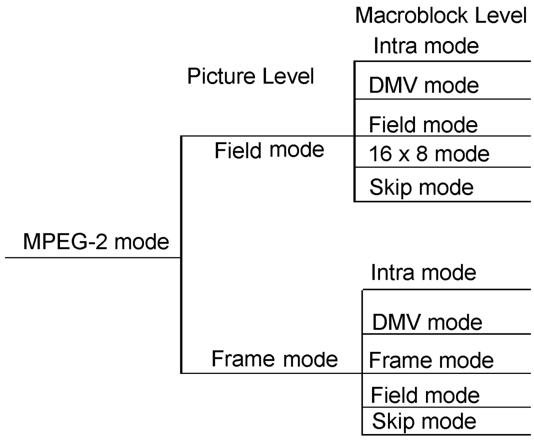

FIG. 1. Mode analysis of MPEG2.

MPEG2 video encoder may use either frame- or field-only coding. In frame-only coding, all the frames of a sequence are encoded as frame picture and in field-only coding, frame is encoded as two field pictures. All macroblocks within I picture are coded intra; however, macroblocks within a P picture are coded either as intra or non_intra. Macroblock within the B pictures can be independently selected either as intra, forward predicted, backward predicted or both forward or backward predicted (Interpolated). Macroblock header contains an element called macroblock_type that can flip these modes on/off. Skipped macroblock is the macroblock for which no data is encoded. In decoder, skipped macroblock is detected by information of Macroblock Address Increment (MBAI) that indicates the difference between macroblock address and the previous macroblock address increment. In skipped macroblock, the decoder has neither information on DCT coefficients nor on motion vector. Macroblock type regarded as skipped macroblock in P picture is macroblock type having a not coded or a NO MC. In B picture, macroblock is skipped when prediction type and motion vectors are the same compared to previous macroblock. There are no skipped macroblocks in I pictures [5].

The study of MPEG2 encoder for different video sequences shows the following observations. Three sequences, Susie, Tennis and BBC3, with 356×288 pixels and frame rate of 25 f/s were used for analysis.

1. The macroblock coding type and the number of bits of encoded macroblock are intimately related to each other.
2. Many coded macroblocks in intercoded frame have zero motion vector or quantised coefficients.
3. Quality of the reproduced picture can be improved if more number of bits is allocated to intracoded macroblock in P and B pictures.
4. Statistics of encoder shows that some macroblocks with low sum of absolute differences are not skipped.
5. Human visual system is more sensitive to distortion in smooth areas and can tolerate more distortion in highly textured areas.
6. Distortion is more visible in brighter areas than in dark areas.

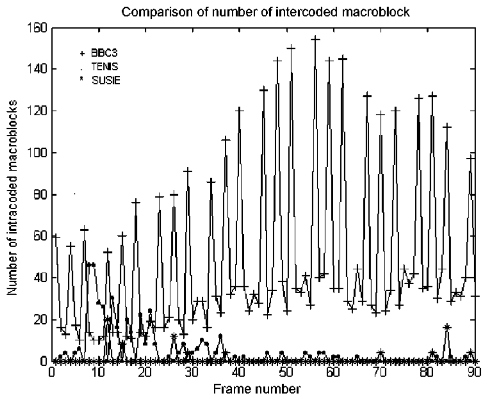

Fig. 2. Intracoded macroblocks in P and B frames.

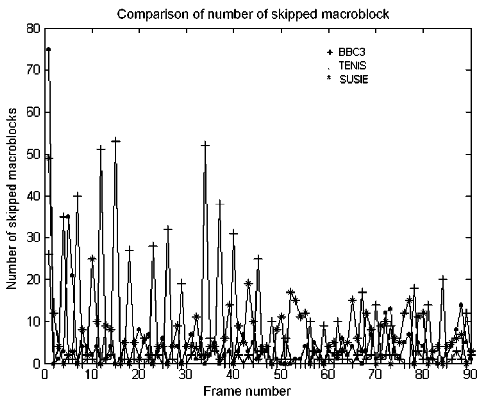

FIG. 3. Skipped macroblocks in P and B frames.

7. If the number of actual coded bits is larger than the allocated target bits reproduced, picture quality degrades and there is possibility of buffer overflow/underflow.
8. Buffer can be prevented from overflow or underflow if difference between target and actual number of bits is kept within a threshold.

For progressive sequences at picture level, frame-only mode is used, but at macroblock level, either field or frame mode is selected. The sequence or the part of sequence with high motion activities favor field-only coding, while sequences with low motion activities favor frame only coding at macroblock level. For selection of macroblock mode as field or frame, it is advantageous to take scene activity into account.

It is observed that the number of intracoded and skipped macroblocks in a video sequence are related to the amount of motion present in the picture. For scenes having complex motion, the amount of intracoded and skipped macroblocks is more in P and B frames. The number of Intracoded macroblocks is zero for Susie sequence, which has least motion while for Tennis and BBC3 sequences it is more. All the sequences were encoded with the same buffer size and bit rate. Figures 2 and 3 show comparison of intracoded and skipped macroblocks for these sequences.

3. Single-pass rate-control algorithm

Several methods for applying R-D theory in MB mode selection are presented in the literature where the distortion-based approach of TM5 is modified by applying Rate Distortion theory. In general, the encoder performance is measured by bit rate R and distortion D introduced in the encoding process. In an R-D optimization framework, the goal is to minimize the distortion D, subject to rate constraint. This problem can be read as

$$\text{Min } \{D(R)\} \qquad \text{subject to } R < Rt$$

where Rt is the allowed rate. The above constrained problem can be solved using Lagrangian optimization. Improved coding efficiency is obtained at the cost of increase in coding time and complexity. For applications with very limited resources and low delay requirements this will not be helpful.

The coded type of macroblock and the number of bits of encoded macroblocks are closely related to each other. Therefore, they should be determined jointly in order to obtain better performance. In TM5, quantisation parameter is calculated by eqn (1).

$$mquant_j = Q_j \times N\_act_j \tag{1}$$

where Q_j is the modulation parameter, which indicates the contribution from the buffer fullness and N_act_j is the normalized activity indicating the contribution from the activity analysis on the j^{th} macroblock. The contribution to rate control from buffer fullness is relatively straightforward in MPEG2 where three virtual buffers are allocated to accommodate the encoding process for I, P, B frames, respectively. To encode an x-type frame, virtual buffer fullness is determined at macroblock level and is then used to determine the modulation parameter Q_j for the j^{th} macroblock. This operation can be described by eqn (2).

$$Q_j = (d_j \times 31 \times picture\_rate)/(2 \times bit\_rate), \tag{2}$$

where d_j is the virtual buffer fullness for the j^{th} macroblock inside the x-type frame. It is determined according to target bit rate allocated to the frame to be encoded, the number of bits generated by encoding the previous (j – 1) macroblocks inside the frame and the initial buffer fullness before the encoding of this frame starts [6, 7].

The information of the coded type of macroblock is not used in the calculation of the quantisatioin parameter. To improve the quality of reproduced picture more number of bits can be allocated to the intracoded macroblocks within the P and B pictures. Quantisation parameter for the intracoded macroblocks within the P and B type of picture can be reduced by a factor. The factor is calculated depending upon the difference between the target and the actual number of bits for that macroblock. All the macroblocks in I frame are intracoded. Some of them can be quantised with a higher quantisation parameter.

The algorithm is as follows:

Step1: Target bit allocation at picture level and calculation of reference quantisation parameter according to TM5.

Step2: Macroblock mode selection. For progressive video sequences, select macroblock mode as field mode if the activity of that macroblock > avg_act, where avg_act is the average activity of the previous frame.

Step3: For allocation of quantisation parameter to every macroblock, check the picture type. If it is I picture go to step 5.

Step 4: For P and B frames, check whether it is macroblock intra, if not go to step 4.

If it is Intra coded macroblock modify *mquant* using eqn (3).

$$mquant = mquant - factor. \tag{3}$$

Factor is obtained from lookup table. It is decided depending upon the difference between the target and the actual number of bits so as to take care of buffer overflow and underflow.

Step 5: Find the difference between the activity of the current and the previous macroblocks. If it is less than a threshold, increase *mquant* for that macroblock by 1.

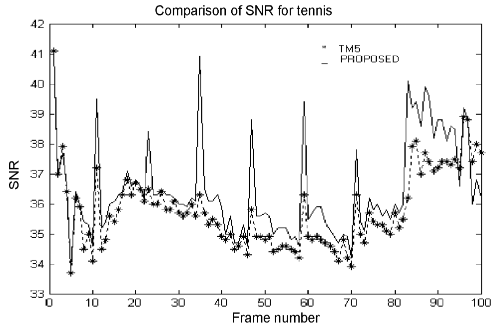

FIG. 4(a). Comparison of SNR for Tennis.

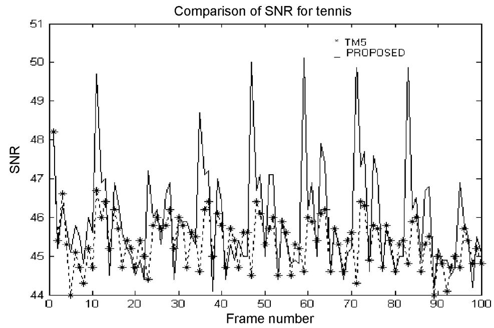

FIG. 4(b). Comparison of SNR for Susie.

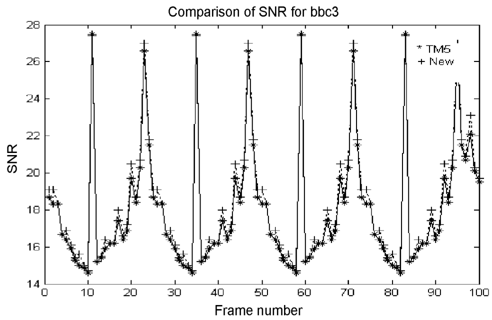

FIG. 4(c). Comparison of SNR for BBC3.

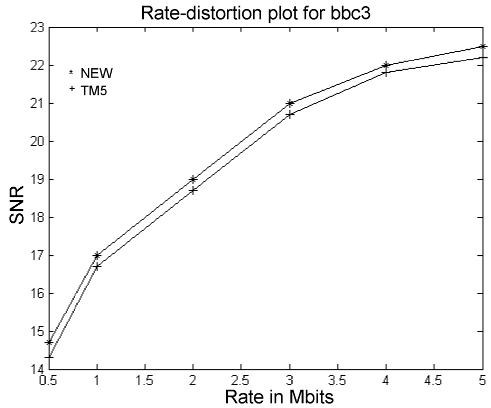

FIG. 5. R-D plot for BBC3.

4. Test results

To verify the effectiveness of the proposed algorithm, it is incorporated into the software MPEG-2 encoder. While the algorithm is designed at macroblock level, its performance assessment is carried out at frame level. This is mainly because the R-D performance of any video compression system is always evaluated at frame level, i.e. peak singal-to-ratio (PSNR) calculation is done at frame level rather than at macroblock level. The algorithm is compared for performance with the results of TM5 rate control scheme. The simulation was done for four sequences BBC3, Susie, and Tennis. In all experiments, the number of encoded frames was 100. Each video sequence was coded with a fixed GOP structure of 12 with distance between P frames equaling 3. Buffer size was set equal to 40*16 kbits. Simulation was done for bit rates of 1, 1.5, 2, 3 and 5 Mb/s. Figure 4 shows the PSNR performance comparison for TM5 and the proposed mode selection rate control algorithm for Tennis, Susie and BBC3 sequences at 1.5 Mb/s. Figure 5 gives R-D plot for BBC3 sequence. The results indicate that significant PSNR gains can be obtained with the proposed algorithm. The algorithm was tested under various buffer capacities to ensure that it would not violate buffer constraints. The buffers do not overflow or underflow for variation in size from 20*16 to 80*16 kbits.

5. Conclusion

For selection of macroblock mode as field or frame it is advantageous to take scene activity into account. Many macroblocks within P and B pictures are intracoded. To improve picture quality more bits can be allocated to these macroblocks. All the macroblocks within I picture are intracoded. To control the output bit rate, some of the intracoded macroblock are coarsely quantised. The proposed algorithm offers improvement in PSNR for various video sequences. The algorithm does not violate buffer constraints under small, medium and large buffer capacities.

References

1. Gary Sullivan, Video compression–from concepts to the H.264/AVC standard, *Proc. IEEE*, **93**, 19–30 (2005).
2. Thomas Sikora, Trends and perspectives in image and video coding, *Proc. IEEE*, **93**, 6–12 (2005).
3. Hyungjoon Kim, Nejat Kamaci and Yucel Altunbasak, Low-complexity rate-distortion optimal macroblock mode selection and motion estimation for MPEG-like video coders, *IEEE Trans. Circuits Systems Video Technol.*, **15**, 823–832 (2005).
4. MPEG-2, Test Model 5(TM5) Doc., Test Model Editing Committee, ISO/IEC JTC1/SC29/WG11/93-225b (1993).
5. Xi Min Zhang, Antony Vetro, Huifang Sun and Yun Q. Shi, Adaptive field/frame selection for high compression coding, *SPIE Conf. on Image and Video Communications*, pp. 583–593 (2003).
6. Christos Grecos, and Jianmin Jiang, Online improvements of the rate-distortion performance in MPEG-2 rate control, *IEEE Trans. Circuits Systems Video Technol.*, **13**, 519–527 (2003).
7. Sung-Hoon Hong, Sang-Jo Yoo and Si Woong Lee, Rate control of MPEG video for consistent picture quality, *IEEE Trans. Broadcasting*, **49**, 1–13 (2003).

Printed and published by Dr S. Venkadesan, Executive Editor, *Journal of the Indian Institute of Science*, Bangalore 560 012; Printed at Ravi Graphics, Bangalore 560 044.