INTELLIGENT VIDEO ENCODING FOR MPEG-1/2 VIDEO

H. Kodikara Arachchi, W.A.C. Fernando
Telecommunications Program, Asian Institute of Technology,
P.O. Box 4, Klong Luang,
Pathumthani 12120, Thailand.
Tel. +66-2-524-5746, Fax: +66-2-524-5730
E-mail: tcc008438@ait.ac.th, fernando@ait.ac.th

Abstract: Even though digital video coding techniques have been developed very rapidly, rate controlling is still an unsolved problem. This paper presents a new scene adaptive video encoding scheme for MPEG-1 and MPEG-2 video encoders. Degradation of picture quality at scene changes due to poor prediction of global complexity measure is addressed. The proposed scheme determines the picture types adaptively based on scene contents of each video frame. The effective placement of reference and inter-reference frames is studied at sudden scene changes and gradual scene changes respectively. Results show that the proposed scheme demonstrates a significant improvement in performance compared to TM5 [1].

1. Introduction

Digital video requires much higher data rates and transmission bandwidths. However the statistical analysis of video signals indicates that there is a strong correlation both between successive picture frames and within the picture elements themselves. Theoretically decorrelation of these signals can lead to bandwidth compression without significantly affecting the image resolution. Moreover, the insensitivity of the human visual system to loss of certain spatio-temporal visual information can be exploited for further reduction. Hence, subjectively lossy compression techniques can be used to reduce video bit rates while maintaining an acceptable image quality. The main objective of compression is to retain as little data as possible, just sufficient to produce the original images without causing unacceptable distortion in the images. To achieve a high compression ratio, the spatial and temporal correlation in natural video sequences must be removed. Transform coding has been proven to be an efficient means of removing spatial correlation of images [2,3]. To remove temporal correlation, motion compensation has been used extensively [2]. Most of the compression algorithms address many of these important issues relating to removal of these correlations [2,4].

Decorrelation of a video sequence however results in a variable rate data stream because statistical properties of pictures in a video sequence vary highly over the sequence [5]. Moreover different picture types with different compression ratios also increase the variance of the data rate. However, for constant bit rate channels (CBR) associated with most of the communication and storage systems require a constant rate data stream. Hence video encoders use a FIFO buffer with a feedback mechanism to regulate the encoded bit rate. When the coder detects a buffer overflow or underflow, the quantization step size is controlled accordingly to increase or decrease the output data rate. As a result the picture quality varies over the video sequence. Due to lack of temporal correlation, this problem is very much prominent at scene transitions.

MPEG-1/2 coding standards define three picture types (I-, P- and B-frames) and encode pictures with a fixed arrangement [6]. Although, encoding with a fixed arrangement picture type may reduce computational complexity, it does not help to change the temporal statistics. Video quality can be improved further by taking advantage of scenes where the statistics allow larger inter-frame distances. To this end, video scene transitions can be used to determine the picture types adaptively. This enables the effective placement of reference frames and inter-reference frames.

The scene transitions can happen anywhere in the fixed arrangement. In the event of an abrupt transition, the first frame of the new scene should be intra-coded in order to avoid severe coding errors. During a gradual scene transition, the distance between two reference frames (I or P) can be changed to improve the picture quality. During most of these gradual transitions, the temporal correlation tends to be reduced. This situation demands more frequent placement of predicted reference frames (P-frames) to uphold the required picture quality. When the video sequence contains rapid motions, this may also require frequent P-frames in order to improve picture quality. If frequent P-frames are not used during these occasions, the MPEG-1 encoder uses many intra-coded MBs to code these frames, as motion estimation becomes ineffective. This increases the bit rate. On the other hand, if the scene does not contain any rapid motions or gradual scene transitions, the inter-frame reference distance can be increased without affecting the picture quality. This is due to the strong correlation between frames. Therefore, this discussion clearly reveals that video can further be compressed very effectively by identifying scene transitions and rapid motions. In this paper we exploit scene transition boundaries and picture type is dynamically changed to improve the rate controlling of the encoder.

Rest of the paper is organized as follows. Section 2 presents works related to our method. Section 3 describes the proposed rate control scheme. Simulation results are presented in section 4. Finally section 5 presents the conclusion.

2. Related Works

Lee et el. [7] proposed a predictive model based algorithm which compensates for the distortion at sudden scene transitions. Picture target bit rate is adjusted in the direction to minimize the distortion in GOP using the rate-distortion (R-D) relations from each P-picture. Since the bit shortage could be occurred, proposed method extends the current GOP to the next.

Wang's [8] algorithm, in which first P-frame after a sudden scene transitions is replaced with an I-frame is more similar to our approach. The extra I-picture is balanced by coding the next scheduled I-picture as a P-picture. Since B-pictures are bi-directionally predicted, they can gain the picture quality from either reference frame which is coded well. Over consumption of bits is addressed by utilizing a binary search algorithm to find the best quantization parameter for the entire picture at the expense of the coding complexity.

Farin et el. [9] proposed to reschedule the GOP structure at sudden scene transitions. This is similar to the Wang's algorithm. However, they proposed to acquire bits by a strong reduction of the quality of the first B-frames after the scene transitions. The quality reduction of the B-pictures is not visible due to temporary masking property of human visual system.

3. Proposed Scheme

The proposed algorithm eliminates the impact of abrupt scene transitions on picture quality. This is done by coding the first scheduled picture in a new scene as an I-picture and the extra I-picture is balanced by coding the next scheduled I-picture as a P-picture. The rescheduled intra-coded picture will provide a good reference for the remaining pictures in the new scene. However, introducing an I-frame needs more bits than for a P-picture, which was scheduled before. Therefore, pictures after the rescheduled I-picture are coded with fewer bits, resulting in a poorer picture quality compared to the conventional scheme (TM5). Thus, the insertion of an additional intra-coded picture must be compensated by coding the next scheduled I-picture as a P-picture. Then the number of frames in the new GOP and the allocated number of bits are modified as in Equations (1) and (2). Equation (2) ensures that any over consumption or under consumption of bits in the previous GOP is compensated from the bit budget of the current GOP. This procedure ensures an effective encoding of abrupt scene transitions while keeping the coding complexity near to that of the TM5 algorithm.

$$N = N + F_N \qquad (1)$$

$$R = R_G + \sum_{i=0}^{m}(T_i - S_i) \qquad (2)$$

where F_N is the remaining number of frames in the current GOP after the scene change frame (including the frame where scene transition occurred), and R_G is the number of bits assigned to a regular GOP of N frames. T_i is the number bits allocated to the i^{th} frame of the current GOP and S_i is the actual number of bits generated by the encoder. m is the index of the last encoded frame in the current GOP before detecting the scene change. During gradual transitions, sub-GOP size (M) can be changed according to the scene contents. Abrupt and gradual scene transitions are detected using histogram techniques.

Figure 1 shows the picture type selection algorithm embedded in a MPEG-1 encoding scheme. The very first block determines the sub-GOP size and the position of an I-frame.

3.1. Scene Transition Detection

The effectiveness of the algorithm depends partly on efficiency of scene transition detection algorithm. Either compressed domain techniques [10,11] or uncompressed domain techniques can be used to detect the scene transitions and the type of the scene transition [11]. If compressed domain technique is used, the picture must be encoded first, and then apply the algorithm to check for the scene transition. For example, if the number of I-coded MBs in a P-coded picture is above a predefined threshold (say 60%) the picture can be considered to be the first P-picture of the new scene after a sudden transition. However to make this possible, the picture must be coded first. If scene transition is detected, then same picture must be coded as an I-picture in the new GOP structure again. Therefore one of the disadvantages of this technique is that some pictures must be encoded twice. Moreover, the picture should be partially decoded in order to get an accurate result [10]. In contrast, uncompressed domain techniques detect scene transitions before the picture is actually encoded. Hence, we used uncompressed domain luminance histogram technique to detect the scene transition.

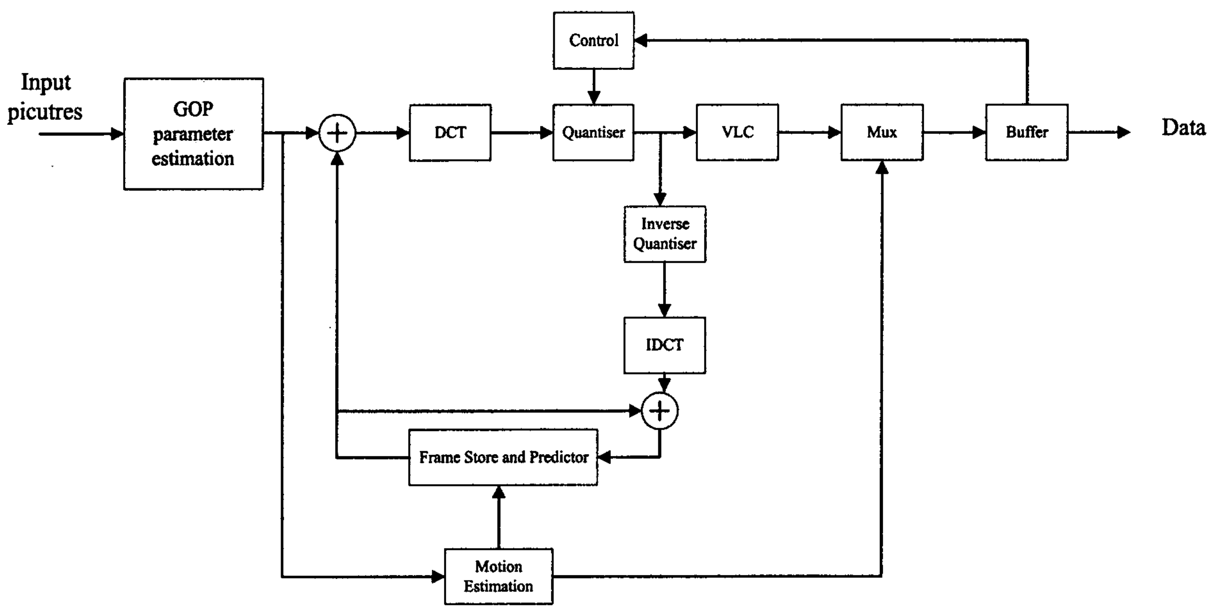

Figure 1. Adaptive frame selection algorithm embedded in a MPEG-1 encoder

4. Results

This section presents some of the results, which are used to demonstrate the effectiveness of the algorithm. Several video sequences are used for simulations. For the performance comparison, the Peak Signal-to-Noise Ratio (PSNR) of the luminance signal is considered as a measure of the objective quality of the encoded video. For all experiments the bit rate of the encoder has been selected as 1.152 Mbits/s. The results are compared against the conventional scheme (TM5), which used GOP size, N=12 and sub-GOP size, M=3 GOP structure.

Figure 2 shows an extract from the *Ragtime* sequence to demonstrate how the algorithm alters the GOP structure at a sudden scene transition. 15th GOP has terminated after nine pictures due to detection of scene transition. The 16th GOP, however, has more pictures in order to compensate for the remaining frames from the earlier premature GOP. Thereby the algorithm is capable of keeping the same number of I-pictures as that of the TM5 algorithm.

```
GOP #    Picture coding types (bitstream
         order / display order):
:        :
:        :
13       IBBPBBPBBPBB / BBIBBPBBPBBP
14       IBBPBBPBBPBB / BBIBBPBBPBBP
15       IBBPBBPBB / BBIBBPBBP
16       IBBPBBPBBPBBPBB / BBIBBPBBPBBPBBP
17       IBBPBBPBBPBB / BBIBBPBBPBBP
18       IBBPBBPBBPBB / BBIBBPBBPBBP
19       IBBPBBPBBPBB / BBIBBPBBPBBP
:        :
:        :
```

Figure 2. An extract from picture coding types map for *Ragtime* sequence

This algorithm was tested for a number of video sequences. Figure 3, Figure 4 and Figure 5 show the PSNR with respect to the encoding order of frames of the *Glas, Waterboy, Ragtime* sequences respectively encoded using both the TM5 and the proposed rate control scheme. Picture number 19 (21 in display order) of the *Glas* sequence shown in Figure 3 is the first picture to be encoded from the new scene. TM5 codes the new picture as a P-picture which results in a poor PSNR of 18.40dB. As high as 98% of the macroblocks of the frame are intra coded. Quality of an I-coding a picture with target bit allocation equal to that of a P-picture results in a lower quality because the coarse quantization.

In contrast, the proposed scheme converts this P frame to an I-frame by terminating the current GOP. As a result, the PSNR is improved by more than 6.5dB. Furthermore the introduction of a proper reference frame has improved the quality of the remaining frames of the GOP. This significant improvement at scene boundaries has accounted for the improvement of the average PSNR over the entire video sequences. Sudden dip in PSNR of *Waterboy* and *Ragtime* sequences shown in Figure 4 and 5 is due to picture after the rescheduled I-pictures belong to the previous scenes at a scene boundary.

Our experiments show that the objective quality can be improved significantly, while keeping the same bit rates or further reducing the bit rates. Some of the experimental results are shown in Table 1 and Table 2.

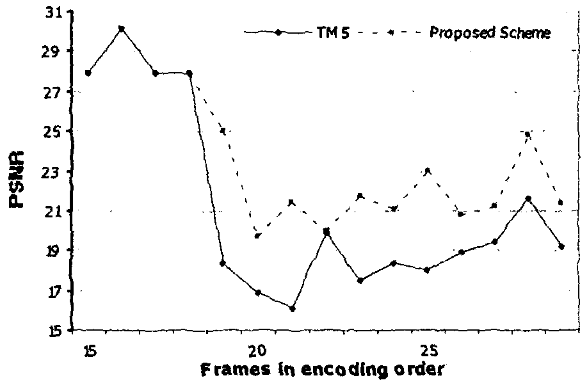

Figure 3. PSNR for *Glas* Sequence

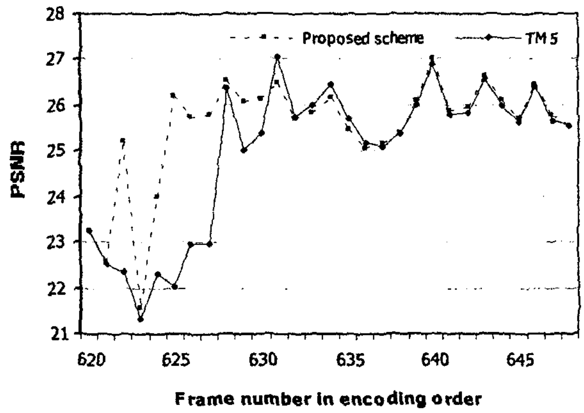

Figure 4. PSNR for *Waterboy* Sequence

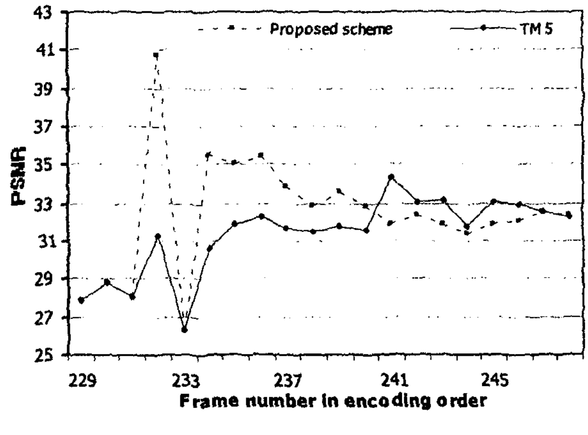

Figure 5. PSNR for *Ragtime* Sequence

Proposed modification affects basically the GOP immediately after the scene transition. Table 1 shows the percentage of I-coded macroblocks in the first P-picture after the scene transition when it is coded using TM5 rate control algorithm and the average improvement of PSNR over the GOP following the rescheduled I-frame. As high as 98% of the macroblocks are intra coded due to lack of proper reference (in *Glas* sequence). The reason for low PSNR at scene changes is due to coarsely quantizing these intra coded MBs in order to reduce the data rate generated by I-coded MBs. However, after introducing an I-picture in place of the P-picture can improve the average PSNR significantly.

Table 1. Picture level performances

Sequence	First P-picture after the scene change[2]	% of I-coded MBs when coded with TM5	Average gain of PSNR over the GOP with the proposed scheme[3]
Glas	19	98.0%	2.4973
	115	98.0%	1.6298
Waterboy	622	87.9%	1.0109
Ragtime	232	86.9%	0.8705
	271	83.8%	1.2556
Terrain	250	78.5%	1.9671
	499	64.9%	0.8358
	748	69.9%	1.0491

According to the Table 2 the average improvement of PSNR over the entire video sequence ranges from 0.05 to 0.18 dB. The average gain is high for a sequence with short scenes (e.g. *Glas* sequence) because the quality is improved at a larger number of scene transitions. However, for sequence with longer scenes, the average gain is small regardless of remarkable improvement at scene transitions. The research is being continued to improve the performance of this intelligent rate control scheme.

Table 2. Comparison and Improvement of PNSR

SEQUENCE	AVERAGE PSNR		AVERAGE IMPROVEMENT OVER TM5
	TM5	Proposed scheme	
Glas Sequence	25.28	25.47	0.1892
Terrain Scene	24.52	24.59	0.0630
Waterboy	25.59	25.64	0.0493
ragtime	28.26	28.34	0.0823

[2] Frame number is in encoding order

[3] Improvement of PSNR is averaged over the GOP

5. Conclusion

Experimental results showed that using the proposed scheme, further compression over the conventional fixed arrangement of picture types is possible while maintaining better picture quality for MPEG-1 video. Furthermore, results also suggest that the objective qualities of the influenced frames at scene transitions were significantly improved with less influence on other frames. Since the adaptive encoder is able to take advantage of the scenes, where long inter-reference frame distances are allowed, video can further be compressed without the degradation of picture quality. This algorithm was tested for many video sequences and proved that the average PSNT significantly be improved consuming the same or lesser number of bits. Therefore this algorithm is capable of controlling the bit rate of the MPEG encoder intelligently achieving a better picture quality. Future works is required to extend this algorithm with violent motions and gradual scene transitions.

References

[1] ISO/IEC/JTC1/SC29/WGII, "Test Model 5", Draft, April 1993.

[2] Andrews, H.C. and Pratt, W.K. "Fourier Transform Coding of Images", *Proceedings of Hawaii International Conference of Systems and Science*, pp.677-678, 1968.

[3] Pratt, W.K., Kane, J. and Andrews, H.C. "Hadamard Transform Image Coding", *Proceedings of IEEE*, Vol. 57, pp. 58-68, 1969.

[4] Pratt, W.K., Chen, W.H. and Welch, R. "Slant Transform Image Coding", *IEEE Transactions of Communications*, Vol. Com-22, pp. 1075-1093, August 1974.

[5] ISO/MPEG II (DIS 13818-2), "Generic coding of moving pictures and associated audio", March 1994.

[6] Sikora, T., "MPEG Digital Video-Coding Standards", *IEEE Signal Processing Magazine*, pp.82-100, September 1997.

[7] M. Lee, S. Kwon, and J. Kim, "A scene adaptive bit rate control method in MPEG video coding", *Visual Comm. Image Processing*, 1997.

[8] Wang, L. "Rate control for MPEG video coding", *Signal Processing: Image Communication*, 15, 6, pp. 493-511, 2000.

[9] D. Farin, N. Mache, P.H.N. de With, "A Software-based High-Quality MPEG-2 Encoder Employing Scene Change Detection and Adaptive Quantization", *International Conference on Consumer Electronics (ICCE)*, June 2001.

[10] W.A.C. Fernando, C.N. Canagarajah, D.R. Bull, "Sudden Scene Change Detection in MPEG-2 Video Sequences," *Proceedings of IEEE International Workshop on Multimedia Signal Processing (MMSP)*, pp. 259-264, 1999.

[11] Alberto Del Bimbo, *Visual Information Retrieval*, Morgan Kaufmann Publishers, Inc., San Francisco, California, 1999.