OPEN ACCESS

University of BRISTOL

Link to published version (if available):
10.1109/ISCAS.2001.921054

Link to publication record in Explore Bristol Research
PDF-document

University of Bristol - Explore Bristol Research

General rights

SCENE ADAPTIVE VIDEO ENCODING FOR MPEG AND H263+ VIDEO

W.A.C. Fernando, C.N. Canagarajah, D. R. Bull

Image Communications Group, Centre for Communications Research, University of Bristol,
Merchant Ventures Building, Woodland Road, Bristol BS8 1UB, United Kingdom.
Voice - +44-117-954-5198, Fax - +44-117-954-5206
Email - W.A.C.Fernando@bristol.ac.uk

ABSTRACT

This paper presents a new scene adaptive video encoding scheme for MPEG and H263+ video encoders. The proposed scheme determines the picture types adaptively based on statistical features of each video frame. Results show that the proposed scheme demonstrates a significant improvement in performance compared to existing schemes.

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 image resolution. Moreover, the intensity of the human visual system to loss of certain spatio-temporal visual information can be exploited for further reduction. Hence, subjectively loosy compression techniques can be used to reduce video bit rates while maintaining an acceptable image quality.

The MPEG-1/2 standards are two remarkable milestones whose overwhelming impact, in defining the next generations of digital multimedia, products, equipments and services is already being felt. These coding standards define three picture types (I, P and B frames) and encode pictures with a fix arrangement. Though, encoding with a fixed arrangement picture types may reduces computational complexity, it does not help to change the temporal statistics. Video can be further compressed by taking advantage of scenes where the statistics allow larger intra and inter-frame distances. To this end, video scene change transitions can be used to determine the picture types adaptively. This ensures that reference frames and inter-reference frames are placed very effectively.

Scene transitions can be divided into abrupt transitions and gradual transitions. Abrupt transitions are defined as the instantaneous scene transitions. Gradual transitions are defined as the transitions where it involves several frames to complete the transitions. Gradual transitions include camera movements such as panning, tilting, zooming and video editing special effects such as fade-in, fade-out, dissolving, wiping. Therefore, these scene transitions can happen at anywhere in the fixed arrangement. In the event of an abrupt transition, first frame of the new scene can 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 selected to improve the picture quality. During most of these gradual transitions, 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 a video sequence has rapid motions, this may also requires frequent P-frames in order to improve picture quality. If frequent P-frames are not used during these occasions, MPEG encoder uses many intra-coded macroblocks (MBs) to code these frames as motion estimation becomes ineffective. This increases bit rate. On the other hand, if the scene does not contain any rapid motions or gradual scene transitions, inter-frame reference distance can be increased without affecting the picture quality. This is due to the strong correlation between frames. This discussion reveals that video can be further compressed very effectively by changing intra/inter frame distance adaptively according to scene transitions and motion characteristics.

In this paper a statistical feature based technique is proposed to evaluate picture types adaptively in MPEG-1/2 and H263+ video encoding. Rest of the paper is organised as follows. Section 2 summarises the related work on this area. Effects of scene transitions on video encoding are discussed in section 3. Proposed scheme is explained in section 4 and section 5 presents some experimental results with the proposed scheme and the conventional scheme. Finally, section 6 presents the conclusions.

2. RELATED WORK

In order to encode a sequence at a fixed bit rate, a bit number allocation strategy is used in TM5 [1]. Although this algorithm can guarantee a relatively good coding performance, it cannot avoid visual quality decrease at scene changes. To eliminate this problem, Luo et al. proposed a scheme based on MPEG-2 target bit number allocation during scene changes [2]. In this scheme, target bits of frames which are affected by an abrupt transition, are increased in order to improve the coding performance. Kim et al. [3] presented a frame work for the forward bit rate control method for a real time MPEG video encoder. They used bit rate estimation and control algorithms based on the linear relationship between the actually generated bit count and the code count. Lan et al. [4] proposed motion analysis based

scheme for adaptive video coding. Actual picture type was decided by examining the accumulation of motion measurements since the last reference frame. However the computational cost to evaluate the motion analysis is very high in this scheme. Yoneyama *et al.* [5] proposed an MPEG encoding algorithm with a scene adaptive dynamic length and dynamic sub Group of Picture (GOP) length. In this scheme, the characteristics of original picture are analysed mainly by using macroblock activity information, and the appropriate GOP length and sub-GOP length values are determined. However, it is not possible to apply this algorithm to hardware implementation with limited frame memories since a pre-load of original pictures with maximum one GOP size is required.

3. EFFECT OF SCENE CHANGES IN VIDEO ENCODING

According to MPEG standards, pictures are divided into a GOP structure and oganisation of these pictures are flexible. Each GOP contains an I-picture, a number of P-pictures and optionally a few B-frames in between two anchor frames (I or P). A GOP structure is specified by number of pictures in it (M) and the distance between two anchor frames (N). Figure 1 illustrates a typical GOP structure with M=12 and N=3.
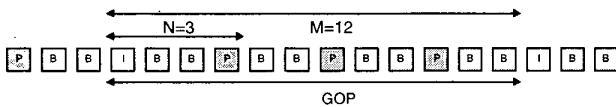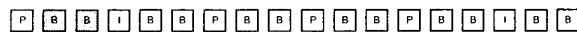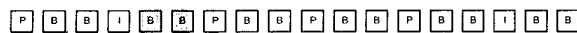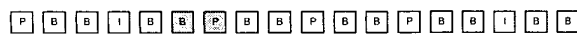
Figure 1: A typical GOP structure (M=12,N=3)

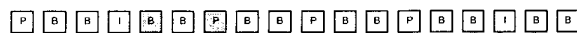### 3.1 Effect of abrupt transitions on video encoding

(a): Scene change at an I-frame

(b): Scene change at a P-frame

(c): Scene change at a B-frame (first B-frame)

(d): Scene change at a B-frame
(second B-frame)

▢ First frame of the new scene

▨ Influenced frames due to scene transition

Figure 2: Influence of a scene change to other frames

If the first frame after a scene change is not an I-frame, the scene change will have several influences on the MPEG

encoder. It may cause the quality decrease at the scene change frame, decrease the quality of the corresponding frames that use the scene change frame as reference frame and may cause buffer underflow at the decoder. The influence of a scene change are clearly illustrated in Figure 2. The first frame of the new scene can be an I-frame, a P-frame or a B-frame. If the first frame of the new scene is an I/P-frame, two B-frames before the I/P-frame are affected due to the scene change as these B-frames cannot be backward predicted from this I/P-frame. If the first frame of the second frame is a B-frame, the other B-frame and the next P-frame is affected as the P-frame cannot be predicted from the previous anchor frame. In H.263+ video encoding, frames after the abrupt scene change are effected and it takes several frames to recover from this scene transition.

Effect of gradual scene transitions on video coding are dependent on the type of scene change. For example during fading, frames do not have a very good correlation.

4. PROPOSED SCHEME

An ideal algorithm for scene adaptive video encoding is to identify all scene changes individually and decide frame types adaptively. However, this increases the computational cost especially for gradual scene change detection. As an alternative to this, a statistical feature based scheme is proposed to encode video adaptively. This scheme uses only mean and variance of blocks and therefore it is computationally inexpensive. This scheme can also deal with local motions effectively, which is also a vital factor in video encoding.

$$MSE_{n+p,i,j} = \left(m_{n+p,i,j} - m_{n,i,j}\right)^2 + \left|\sigma^2_{n+p,i,j} - \sigma^2_{n,i,j}\right| \qquad (1)$$

where, n - frame number, p - an integer, i, j - location of the block within the image.

In the proposed scheme, each video frame is divided into $N_1 \times N_2$ blocks. If frame $n+p$ can be motion compensated from frame n with an acceptable error, it indicates that both frames must have similar statistics. Thus, in this proposed scheme, statistical features for each block of frame $n+p$ is compared against corresponding blocks in frame n. Both mean and variance of each block are considered for this purpose. These two statistical features are combined to evaluate MSE between corresponding blocks in frame $n+p$ and frame n as shown in Equation (1). These MSE values are used to detect abrupt and gradual (or rapid motion) transitions in the sequence with two thresholds T^A_{Scene} and T^G_{Scene}. If most MSEs are very large, then correlation between frame $n+p$ and frame n are very low and this situation indicate that there should be an abrupt transition between these two frames. When MSEs are very

II-250

small, both frames are highly correlated. If MSEs are in between these two extremes, then it can be assumed that either a gradual transition or a rapid motion is expected within these two frames.
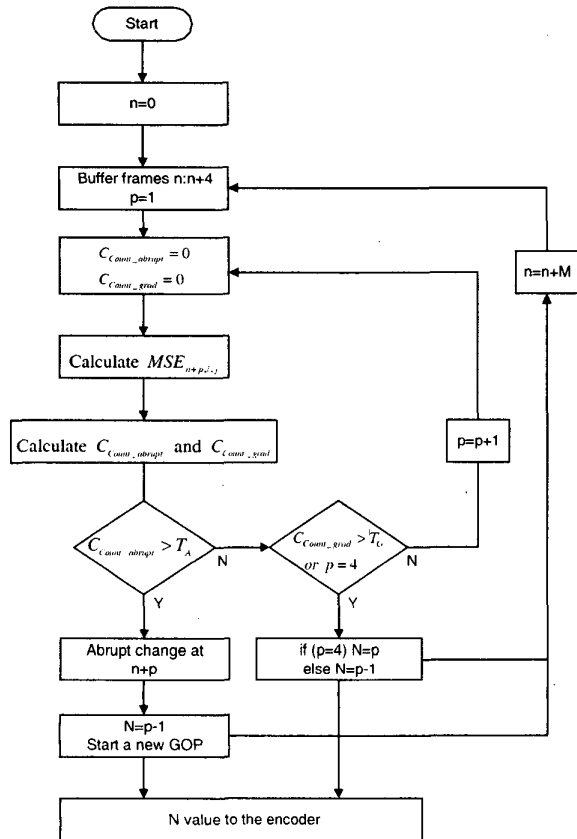

Figure 3: Flowchart of the proposed scheme

If several block MSEs exceed the pre-determined threshold (T_{Scene}^{A}), then an abrupt transition can be declared. All MSEs, which exceed the threshold T_{Scene}^{A} are counted and compared against another threshold (T_A) to identify the abrupt transition precisely. If an abrupt scene transition is identified, then current GOP is terminated and new GOP is started by coding the first scheduled picture in a new scene as an I-picture and the extra I-picture is further balanced by coding the next scheduled I-picture as a P-picture. If any abrupt transitions are not identified, then MSEs, which exceed the threshold T_{Scene}^{G} , are counted and compared against T_G . If it satisfies, then most of the blocks in frame $n + p$ and frame n are not highly correlated. Therefore, it can be assumed that that frame $n + p$ cannot be motion compensated by frame n . Thus, sub-GOP size (N) should be set to $N = p - 1$. If counted MSEs are less than the threshold T_G , it can be

assumed that frame $n + p$ can be motion compensated by frame n effectively. Then, value of p should be incremented by one and the same algorithm is applied until it finds a suitable value for N . In practice, value of N cannot be too large as encoder needs to buffer N number of frames before encoding them, which increases the delay at the decoder. Therefore, in this proposed scheme the maximum value for N is limited to four. Figure 3 illustrates the complete algorithm.

Value of M has been selected as 36. However, statistical features (MSEs) are monitored continuously against the statistical features of the last I-frame and if it exceeds the threshold (T_A), an I-frame is inserted and a new GOP is started as in the case of an abrupt transition. For all simulations, parameters are set as follows. $N_1 = N_2 = 8$, $T_{Scene}^{G} = 2500$, $T_{Scene}^{A} = 5000$, $T_G = \frac{T}{4}$, $T_A = \frac{3T}{4}$

(where T - Total number of blocks in a frame)

5. RESULTS

This section presents some results to demonstrate the effectiveness of the statistical feature based adaptive video encoder. Several QCIF video sequences are used for simulations. PSNR of the luminance signal is considered to measure the objective quality of encoded video. For MPEG-1, results are compared against the conventional scheme with M=12 and N=3 GOP structure. Ordinary H.263+ encoder is used as the conventional scheme for comparison of results in H.263+ encoding with scene transitions.
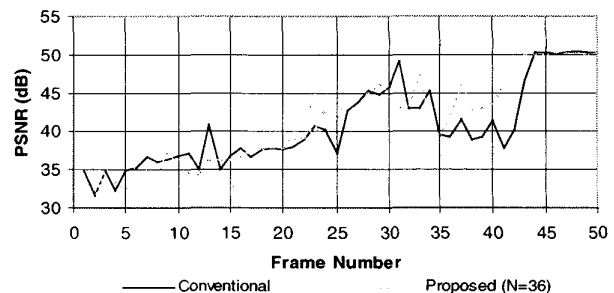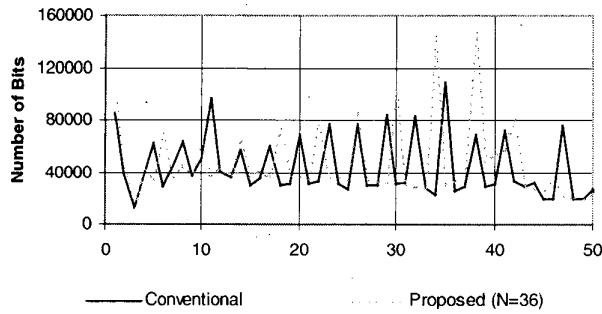

Figure 4: Comparison of the PSNR variation for MPEG-1 encoded video with the proposed scheme (N=36) and the conventional scheme

Consider a scene of length 50 frames, which has an abrupt scene change at 41st frame. Figure 4 presents the PSNR comparison of the proposed scheme with the conventional scheme. This clearly illustrates that the proposed scheme has a large PSNR gain around the scene change compared to the conventional scheme. Variation of number of bits required to code the frames in the proposed scheme are shown in Figure 5. Since an I-frame is inserted at the scene change it demands higher bit rate at this frame. Table 1 summarises results with

II-251

the proposed scheme and the conventional scheme. First four sequences are standard sequences. Fifth sequence is a test sequence which has two abrupt transitions. Proposed scheme produces a considerable average PSNR gain (0.56 dB) for this sequence. Sixth and seventh sequences represent a news clip and a movie clip respectively. Though the proposed scheme's average PSNR gain is small, bit rate reduction is significant for both sequences. In general, proposed scheme has higher average PSNR and lower bit rate compared to the conventional scheme for all seven sequences considered.
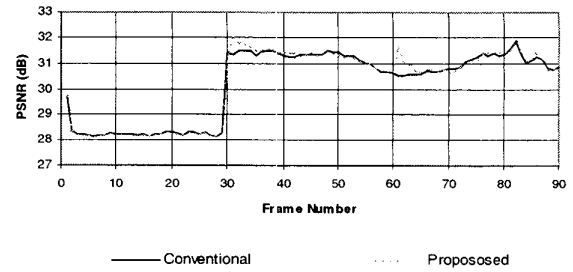

Conventional   Proposed (N=36)

Figure 5: Comparison of the bit rate variation for MPEG-1 encoded video with the proposed scheme (N=36) and the conventional scheme

Sequence	Average PSNR		Average number of bits	
	Conventional (12,3)	Proposed	Conventional (12,3)	Proposed
akiyo	50.07	50.86	27155	24193
foreman	43.17	43.38	46594	46099
funfair	35.76	36.65	45796	44181
tennis	37.50	37.57	46115	46030
abrupt change	46.91	47.47	32567	32425
news clip	49.35	49.39	86004	82311
movie clip	41.18	41.21	54837	44931

Table 1: Summary of the results

Motion analysis based scheme could achieve up to 13.9% (depending on video contents) savings in bits [4]. But picture quality sometimes worse than the conventional scheme. However proposed scheme can achieve up to 22.2% savings in bits while maintaining similar or better picture quality.

Figure 6 presents PSNR variation for a video sequence, which has two scene changes with H263+ video encoding. In H263+ video coding abrupt scene transition is detected using the proposed scheme and an I-frame is inserted at the scene change. Figure 6 illustrates that PSNR variation of the proposed scheme is higher than the conventional scheme. Average number of bits for this sequence with the proposed and conventional are 5736 bits and 5768 bits respectively.

Conventional   Propososed

Figure 6: Comparison of PSNR with the proposed scheme and the conventional scheme (H.263+)

6. CONCLUSIONS

In this paper a scene adaptive video encoding scheme is proposed based on statistical features. Results show that using the proposed scheme further compression over the conventional fixed arrangement of picture types is possible while maintaining similar or better picture quality. Though the proposed scheme performs better than the conventional scheme, it needs 6.4 flops/pixel to determine the sub-GOP size and the GOP size. However, this computational cost is very low compared to the motion analysis based scheme [4].

ACKNOWLEDGEMENTS

REFERENCES

1. ISO/IEC/JTC1/SC29/WGII, 'Test Model 5,' Draft, April 1993.

2. Luo, L., Zou, C. and He, Z. 'A new Algorithm on MPEG-2 Target Bit-Number Allocation of Scene Changes,' IEEE Transactions on Circuits and Systems for Video Technology, Vol. 7, No 5, pp.815-819, 1997.

3. Kim, T. Roh, B. and Kim, J. 'An Accurate Bit-rate Control for Real Time MPEG Video Encoder,' Signal Processing: Image Communications 15, pp.479-492, 2000.

4. Lan, A.Y., Nguyen, A.G. and Hwang, J.N. 'Scene-Content-Dependent Reference Frame Placement for MPEG Video Coding,' IEEE Transactions on Circuits and Systems for Video Technology, Vol. 9, No 3, pp.478-489, 1999.

5. Yoneyama, A., Nakajima, Y., Yanagihara, H. and Sugano, M., ' MPEG Encoding Algorithm with Scene Adaptive Dynamic GOP Structure,' IEEE Third International Workshop on Multimedia Signal Processing, pp. 297-302, 1999.