SCENE ADAPTIVE VIDEO ENCODING FOR MPEG AND H.263+ VIDEO

W. A. C. Fernando[1] and C. N. Canagarajah[2]

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

Due to rapid advances in video products such as digital cameras, camcorders, storage devices (DVDs) and the explosion of the internet, the digital video "for every one" is now becoming a reality. The demand for digital video is also increasing in areas such as video conferencing, multimedia authoring systems, education and video-on-demand systems. The large channel bandwidth and memory requirements for the transmission and storage of image and video necessitate the use of video compression techniques. 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 [1,2,3]. To remove temporal correlation, motion compensation has been used extensively [4,5,6,7]. Most of the compression algorithms address many of these important issues relating to removal of these correlations [1,3,4].

MPEG-1/2 coding standards define three picture types (I-, P- and B-frames) and encode pictures with a fixed arrangement [4]. Although, encoding with a fixed arrangement picture type may reduce computational complexity, it does not help to change the temporal statistics. Video can be compressed 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.

The rest of the paper is organised as follows. Section 2 summarises the related work on this area. An overview to H.263+ video is given in section 3. The effects of scene transitions on video encoding are discussed in section 4. The proposed scheme is explained in section 5 and section 6 presents some experimental results with the proposed scheme and the conventional scheme. Finally, section 7 presents the summary for this paper.

2. RELATED WORK

The basic requirements for MPEG-1/2 video coding is a high compression ratio with good image quality and the support for a number of features such as random access, fast search, reverse play back, etc. To achieve a high compression ratio, the spatial and temporal correlation in natural video sequences must be removed as much as possible. In order to encode a sequence at a fixed bit rate, a bit number allocation strategy is used in TM5 [8]. Although this algorithm can guarantee a relatively good coding performance, it cannot avoid a decrease in visual quality at scene changes.

Wang [9] presented an algorithm to eliminate the impact of abrupt scene transitions on picture quality. The author proposed 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

[1] Telecommunications Program, School of Advanced Technologies, Asian Institute of Technology, P.O. Box 4, Klong Luang 12120, Thailand. Email – fernando@ait.ac.th

[2] Image Communications Group, University of Bristol, Merchant Ventures Building, Woodland Road, Bristol BS8 1UB, United Kingdom.

Contributed Paper

Manuscript received June 25, 2001

scheduled I-picture as a P-picture. However, this scheme does not handle gradual scene transitions and rapid motions. Luo *et al.* proposed another scheme on MPEG-2 target bit number allocation during scene changes [10]. 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.* [11] presented a framework for the forward bit rate control method for a real time MPEG-1/2 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.* [12] proposed motion analysis to adapt to scene content. Actual picture type was decided by examining the accumulation of motion measurements since the last reference frame. However this scheme is computationally most expensive to evaluate the motion.

An adaptive-size GOP is examined in [13]. The extreme case of allowing only one independent "I" reference frame for every GOP was considered. This type of reference frame is situated as the first frame of every new scene in order to prevent predictions from being made between two different scenes. This encoding scheme was used for the purpose of indexing video according to scene changes. Random-access indexing facilitates video editing and browsing by allowing a user to search through scenes rather than frame by frame for a particular segment. Fast and robust scene change detection was performed by using MPEG bit stream encoded parameters. The correlation measures developed were based on the prediction error power of "P" pictures, the average number of macroblocks using forward or backward prediction, and motion-vector inner products. Frames not classified as a scene change were encoded in a fixed periodic structure of picture types to reduce bit-rate variations for asynchronous transfer mode (ATM) network scheduling. This encoded technique only adapts to abrupt changes in content and ignores the changes occurring within a scene. An important conclusion from this work was the fact that infrequent "I" reference frames did not cause severe degradation in performance due to error propagation.

Yoneyama *et al.* [14] proposed an MPEG encoding algorithm with a scene adaptive dynamic length and dynamic sub-GOP length. In this scheme, the characteristics of the original picture are analysed mainly by using macroblock (MB) 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 a maximum one GOP size is required.

3. MPEG-1/2 AND H.263+ OVERVIEW

3.1 MPEG-1/2

MPEG-1/2 [15] video compression is used in many current and emerging products for digital television and broadcasting. MPEG-2 video is broken up into a hierarchy of layers to help with error handling, random search editing, and synchronisation. From the top level, the first layer is known as the video sequence layer. The second layer down is the group of pictures (GOP), which is composed of one or more groups of intra (I) frames and/or non-intra (P and/or B) frames. Each GOP is divided into sub-units called sub-GOP which contain B-frames and an I- or a P-frame. The third layer down is the picture layer itself, and the next layer beneath is called the slice layer. Each slice consists of MBs, which are 16x16 arrays of luminance pixels, or picture data elements, with 8x8 arrays of associated chrominance pixels. The MBs can further be divided into 8x8 blocks, for further processing. Figure 1 shows a typical MPEG-2 video sequence with GOP of 12 and sub-GOP size of 3.

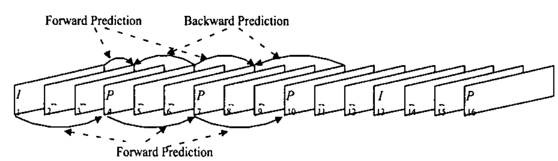


Figure 1: A Typical MPEG-2 Compressed Video Sequence

3.2 H.263+

The H.263 video codec is the ITU-recommended standard [8] for very low bit-rate video compression. Currently, H.263 encoding and decoding is the predominant technology used in videoconferencing across analogue telephone lines, optimised at rates below 64 kbits/s. The primary goal of the H.263 encoder is to represent information as compactly as possible without compromising video fidelity. In the encoder, all frames in a video sequence are categorised as either I-frames or P-frames. These frames are encoded in the same manner as in MPEG-1 video encoding. B-frames can also be used in H.263 coding as a PB-frame, which consists of two pictures coded as one unit. A PB-frame consists of one P-picture, which is predicted from the last decoded P-picture and one B-picture, which is predicted both from last decoded P-picture and the P-picture currently being decoded. This picture is called the B-picture, because parts of it may be bi-directionally predicted from the past and current P-pictures.

Figure 2: A typical picture arrangement in H.263 video coding

Recently, a development effort has been intended for the short term standardisation of enhancements of the H.263 video coding algorithm for real time telecommunication and related non-conversational services [16]. The ITU-T standard H.263+, also called H.263 version 2, is an extension of the original H.263 standard incorporating extra features. H.263+ is a layered scalable coding scheme. This means that the bit stream is organised in a hierarchical structure, divided into a number of levels, the base layer and one or more enhancement layers. The base layer, encoded at a low bit rate, is the minimum needed to decode the video sequence, and consists of an intra-coded picture and multiple inter-coded pictures. The enhancement layers, which require additional bandwidth for encoding, improve upon the spatial or temporal quality of the reconstructed base layer sequence. Thus, an enhancement layer can only be decoded if the layer immediately below it, known as the reference layer, has already been decoded.

4. EFFECT OF SCENE TRANSITIONS ON VIDEO ENCODING

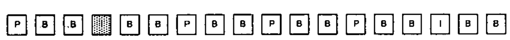


Figure 3 (a): Scene change occurred at an I-frame

Figure 3(b): Scene change occurred at a P-frame

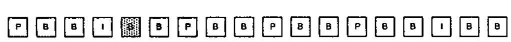


Figure 3(c): Scene change occurred at a B-frame (first B-frame)

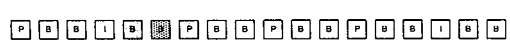
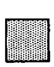
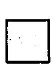


Figure 3(d): Scene change occurred at a B-frame (second B-frame)



First frame of the new scene



Immediately influenced frames due to a scene transition

Figure 3: Influence of a scene change to other frames

As long as 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 to decrease at the scene change frame, a decrease in 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 is clearly illustrated in Figure 3. 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-frame, two B-frames before the I-frame are affected due to the scene change as these B-frames cannot be backward predicted from this I-frame. When the scene change occurred at a P-frame two B-frames before the P-frame are immediately influenced as backward prediction from this P-frame is not possible. If the first frame of the second sequence is a B-frame, the other B-frame and the next P-frame is immediately 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 affected and it takes several frames to recover from this scene transition.

5. PROPOSED SCHEME

An ideal algorithm for scene adaptive video encoding is to identify all these scene changes 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.

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

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 $N_1 \mathrm{x} N_2$ 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 $n+p$ are 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 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.

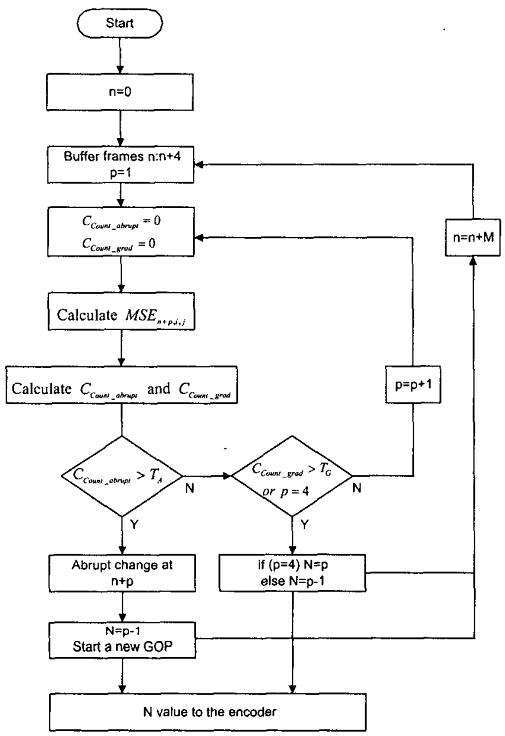


Figure 4: Flowchart of the proposed scheme

For all simulations, parameters are set as follows.

$N_1 = N_2 = 8$
$T^G_{Scene} = 2500$
$T^A_{Scene} = 5000$
$T_G = T * 25\%$
$T_A = T * 75\%$

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

If several MSEs exceed the pre-determined threshold (T^A_{Scene}), then an abrupt transition can be declared. All MSEs which exceed the threshold T^A_{Scene} are counted to a variable C_{Count_abrupt} and this count is 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^G_{Scene}, are counted to an another variable C_{Count_grad} 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. Therefore, 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 the value of p should be incremented by one and the same algorithm is applied until it finds a suitable value for N. In practice, the value of N cannot be too large as the encoder needs to buffer N number of frames before encoding them, which increases the delay at the decoder [17]. Therefore, in this proposed scheme the maximum value for N has been limited to four. Figure 4 illustrates the complete algorithm.

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. 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 (2) and (3). This procedure ensures an effective encoding of abrupt scene transitions.

$$M = M + F_M \tag{2}$$

$$R = R + R_G \tag{3}$$

where F_M is the remaining number of frames in the current GOP after the scene change frame (including the scene changed frame), and R_G is the number of bits assigned to a regular GOP of M frames.

As explained previously, during gradual transitions, N may be changed according to the scene contents. Since change in N is not very frequent, this may not create a problem.

5.1 Proposed scheme embedded in a MPEG-1 encoder

Figure 5 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. When a sub-GOP terminates, this block buffers the next four frames and processes these frames as explained previously and determines a suitable value for N. When it is estimated, value of N is directly input to the normal MPEG-1 encoder.

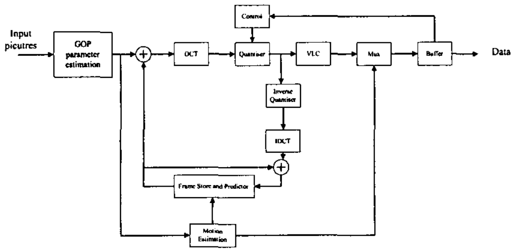


Figure 5: Adaptive frame selection algorithm embedded in a MPEG-1 encoder

6. RESULTS

This section presents some of the results which are used to demonstrate the effectiveness of the statistical feature based adaptive video encoder. Several video sequences are used for simulations. For the performance comparison, the PSNR of the luminance signal is considered to measure the objective quality of the encoded video. For all experiments the bit rate of the encoder has been selected as 1.152 Mbits/s. For MPEG-1, the results are compared against the conventional scheme, which used M=12 and N=3 GOP structure. An ordinary H.263+ encoder is used as the conventional scheme for the comparison of results in H.263+ encoding with scene transitions.

6.1 Performance of the algorithm with varying M

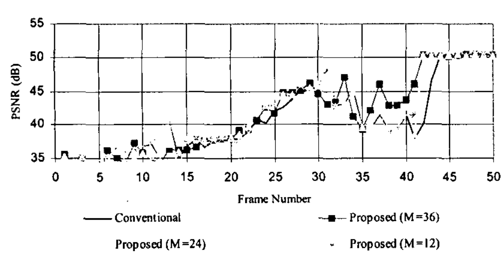


Figure 6: Comparison of the PSNR variation for MPEG-1 encoded video with different values of M

Consider a scene of length 50 frames, which has an abrupt scene change at the 41st frame and a rapid motion around frame number 25. Figure 6 presents the PSNR performance comparison of the proposed scheme with the conventional scheme for three different values of M. As expected, the PSNR of the proposed scheme (for all values of M) is well above the conventional scheme at the abrupt scene transition. For $M = 12$, the PSNR of both schemes are similar in every frame except at scene transitions frames. However, for other values of M, the PSNR is different for most of the frames as they use different N values.

Table 1 presents an average PSNR and number of bits for this sequence. The performances of the proposed scheme for all values of M are better than the conventional scheme. However, the highest average PSNR and the lowest number of bits are given for M=36. This is due to the fact that the intra-frame distance can be increased when correlation between frames is high. For all other simulations, the 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 they exceed the threshold (T_A), an I-frame is inserted and a new GOP is started as in the case of an abrupt transition.

6.2 Performance of the algorithm with standard sequences

In this section the performance of the proposed scheme is evaluated with *'akiyo'*, *'foreman'*, *'funf'* and *'tennis'* sequences. Furthermore a news clip is also used to evaluate the performance of the scheme.

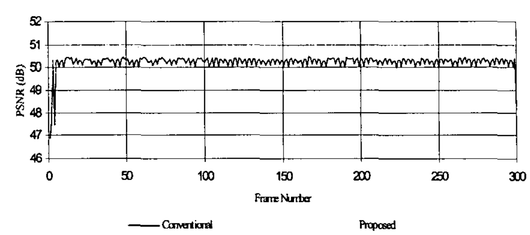


Figure 7: Comparison of the PSNR variation for MPEG-1 encoded *'akiyo'* sequence with the proposed scheme and the conventional scheme

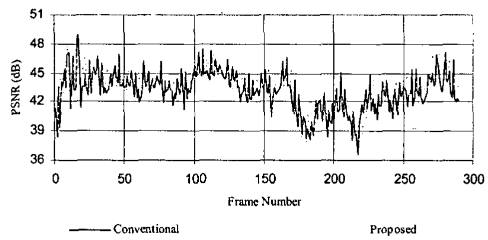


Figure 8: Comparison of the PSNR variation for MPEG-1 encoded *'foreman'* sequence with the proposed scheme and the conventional scheme

Figure 7 and Figure 8 depict the comparison of the PSNR variation for *'akiyo'*, and *'foreman'* sequences. Since *'akiyo'* is a head and shoulder type sequence, frames are highly correlated and hence both intra-frame and inter-frame distance can be increased. Therefore, the PSNR of the proposed scheme is continuously higher than the conventional scheme. The second sequence (*'foreman'*) has considerable motion and a camera pan. Thus, its PSNR has a large variation in both schemes. However, the proposed scheme mostly give higher PSNR compared to the conventional one as it used adaptive sub-GOP size.

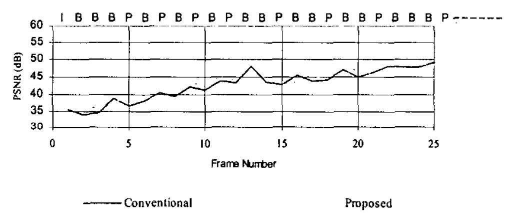


Figure 9: PSNR and the picture type variation for the first few frames of a movie clip

Figure 9 presents the PSNR variation for the first few frames for the movie clip considered. Furthermore, it also depicts how the picture type is changing for these frames. The PSNR of the conventional scheme is higher than the proposed scheme at frame number 13 where an I-frame is used in the conventional scheme. Generally, the PSNR of the proposed scheme is higher than the conventional scheme for other frames.

Table 2 summarises the results of the proposed scheme and the conventional scheme for different short video sequences. The first four sequences are standard sequences and two of their PSNR distributions were presented earlier. The fifth sequence is a test sequence which has two abrupt transitions. The proposed scheme produces a considerable average PSNR gain (0.56 dB) for this sequence. The sixth sequence represents a news clip. Though the proposed scheme's average PSNR gain is small, bit rate reduction is significant for these sequences.

6.3 Proposed scheme with H.263+ video coding

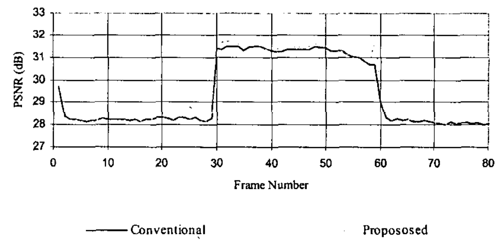


Figure 10: Comparison of the PSNR variation for H.263+ encoded video with the proposed scheme and the conventional scheme

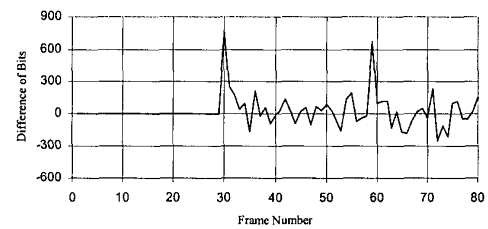


Figure 11: Difference of bits between the conventional and the proposed schemes

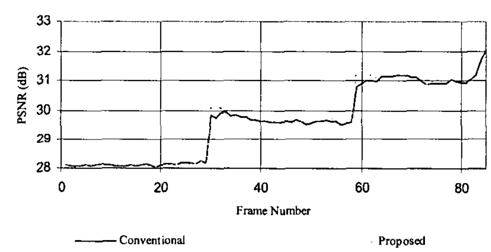


Figure 12: Comparison of the PSNR variation for H.263+ encoded video with the proposed scheme and the conventional scheme

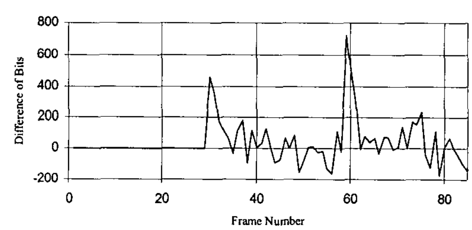


Figure 13: Difference of bits between the conventional and the proposed schemes

Figure 10 presents PSNR variation for a video sequence, which has two scene changes, with H.263+ video encoding. In H.263+ video coding abrupt scene transition is detected using the proposed scheme and an I-frame is inserted at the scene change. Figure 10 illustrates that the PSNR variation of the proposed scheme is higher than the conventional scheme at the scene transition and few frames later. The differences in bit rate for this sequence with the proposed and the conventional schemes are illustrated in Figure 11. This shows that the conventional scheme needs extra bits at the scene transitions compared to the proposed scheme. In the conventional scheme, most MBs are intra-coded at an abrupt scene transition. Thus, they need many overhead bits in the conventional scheme. Figure 12 and Figure 13 presents the average PSNR and the difference in bits for another sequence, which has two abrupt transitions. Similar behaviour can also be observed in this sequence.

6.4 Proposed scheme with different scene transitions

The proposed scheme and conventional scheme are tested with different scene transitions: fade-in, fade-out, wipe, dissolve, camera motions and abrupt transitions. At the start of a fade-in scene transition, frames are not correlated as the transition tries to start from a solid colour to a second scene. Therefore, reducing inter-frame distance at the start of the transition may help the encoder to code it effectively. However, at the end of the transition correlation between frames is not that bad compared to the initial frames. Therefore, inter-frame distance may be increased. Figure 14 depicts how a number of intra-coded MBs are changing for a 25 frame fade-in operation. The number of intra-coded MBs in conventional scheme is high at the initial part of the transition and gradually the number reduces towards the end of the operation. In the proposed scheme a similar behaviour can be observed. But, the number of intra coded MBs is less compared to in the conventional scheme. This is due to the fact that the inter-frame distance is changing adaptively in the proposed scheme. For example, at the start of the operation $N=2$ and at the end of transition $N=4$. Similar behaviour can be observed during fade-out. The only difference is that the solid colour is now at the end of the transition. During dissolving the middle frames have a low correlation. But at the start and at the end of the transition, the correlation between frames is high. In wiping, the correlation of images is high as the transition only happens along the boundary and a high value of inter-frame distance would be desirable for this. Scene content is also changing during a camera movement. For a fast camera movement, frames are not correlated and the inter-reference frame distance may needs to be reduced.

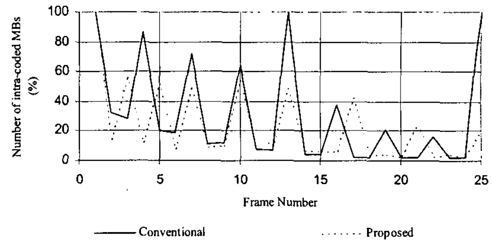


Figure 14: Behaviour of intra-coded MBs during a fade-in scene transition

Table 3 illustrates the average PSNR and the average bit rate with the proposed scheme and the conventional scheme for video sequences which has the above scene transitions. Results show that the proposed scheme performs better than the conventional scheme for all scene transitions. During camera motion there is a considerable PSNR gain but there is no significant difference in the average number of bits. This is due to the use of small sub-GOP size as the correlation between frames is low during this operation. Abrupt transition also provides considerable PSNR gain. During other scene transitions, there is no significant PSNR gain but the average number of bits has been reduced.

Table 4 illustrates the summarised results for some movie clips. Five movie clips are considered in this experiment. Results show that bit rate can be decreased significantly (in average 18%) with the proposed scheme for these sequences. In general, the proposed scheme has a higher average PSNR and a lower bit rate compared to the conventional scheme for all sequences considered.

The proposed scheme can identify sub-GOP size adaptively. Results show that video sequences can be further compressed without affecting the picture quality. During gradual and rapid motions sub-GOP size can be set very effectively to increase the picture quality and reduce the bit rate. Furthermore, the proposed adaptive picture placement scheme is also able to place intra-frames more effectively at an abrupt transition. Since the intra-reference frame distance has been increased to 36, video can be further compressed as the number of intra-coded frames is less compared to the conventional scheme. This is more useful when scene content does not have significant local motion. When the scene does not have any significant motion, frames are highly correlated and the intra-frame distance can be increased. However, if the scene has large motions, the placement of intra-coded frames may be needed more regularly. Since the algorithm compares the MSEs of each frame with the last intra-coded frame, an intra-frame can be placed though an abrupt scene transitions does not actually occur.

Although 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. Most of these flops are required to calculate MSEs. However, the computational cost is very low compared to the motion analysis based scheme [12].

7. CONCLUSIONS

In this chapter a scene adaptive video encoding scheme was proposed based on statistical features of each image. Experimental results showed that using the proposed scheme, further compression over the conventional fixed arrangement of picture types is possible while maintaining similar picture quality for MPEG-1 video. For H.263+ video coding intra-frames can be placed at abrupt scene transitions and this increases the PSNR and reduces the bit rate slightly. Furthermore, results also suggest that the objective qualities of the influenced frames at scene changes 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.

REFERENCES

1. 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.

2. Pratt, W.K., Kane, J. and Andrews, H.C. 'Hadamard Transform Image Coding,' Proceedings of IEEE, Vol. 57, pp. 58-68, 1969.

3. 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.

4. Shie, M.C., Fang, W.H., Hung, K.J. and Lai, F., 'Fast, Robust Block Motion Estimation Using Simulated Annealing,' IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences, Vol. E83-A, No.1, pp.121-127, 2000.

5. Chen, P.Y. and Jou, J.M., 'Fast-search Motion Estimation Method and Its VLSI Architecture,' IEEE Transactions on Circuits and Systems II: Analog and Digital Signal Processing, Vol.46, No.9, pp.1233-1240, 1999.

6. Sung, W.R., Kang, E.K. and Choi, J.S., 'Adaptive Motion Estimation Technique for Motion Compensated Inter-frame Interpolation,' IEEE Transactions on Consumer Electronics, Vol.45, No.3, pp.753-761, 1999.

7. Choi, S.Y. and Chae, S.I., 'Hierarchical Motion Estimation in Hadamard Transform Domain,' Electronics Letters, Vol.35, No.25, pp.2187-2188, 1999.

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

9. Wang, L. 'Rate Control for MPEG Video Coding,' Signal Processing: Image Communications 15, pp. 493-511, 2000.

10. 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.

11. 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.

12. 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.

13. Liu, H.C. and Zick, G., 'Automatic Determination of Scene Changes in MPEG compressed video,' Proceedings of IEEE Symposium on Circuits and Systems, Vol. 1, pp.764-767, 1995.

14. 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.

15. Sikora, T., 'MPEG Digital Video-Coding Standards,' IEEE Signal Processing Magazine, pp.82-100, September 1997.

16. Draft ITU-T Recommendation H.263+, 'Video Coding for Very Low Bit Rate Communication,' September 1997.

17. Whitaker, J., 'DTV: The Revolution in Electronic Engineering,' McGraw-Hill Publishers, 1998.

BIOGRAPHIES

Anil Fernando is currently an assistant professor at Asian Institute of Technology (AIT). He received the B.Sc. Engineering degree (First class) in Electronic and Telecommunications Engineering from the University of Moratuwa in 1995 and the MEng degree (Distinction) in Telecommunications from AIT in 1997. He has completed his Ph.D. at the Department of Electrical and Electronic Engineering, University of Bristol. His current research interests include scene change detection in uncompressed and compressed video, video editing in compressed video, intelligent video encoding, COFDM for wireless channels, channel coding and modulation schemes for satellite channels.

Nishan Canagarajah is currently a Senior Lecturer in Signal Processing at University of Bristol. He has BA (Hons) and Ph.D. in DSP techniques for speech enhancement, both from the University of Cambridge. He is a committee member of the IEE Professional Group E5, member of the virtual centre of excellence in digital broadcasting and multimedia technology and an associate editor of the IEE Electronics and Communication Journal. He is also an editor of a book on Mobile Multimedia Technology. His research interests include image and video coding, non-linear filtering techniques and the application of signal processing to audio and medical electronics.

M	Average PSNR (dB)		Average number of bits/frame	
	Conventional	Proposed	Conventional	Proposed
12	40.82	41.29	45167	44769
24	-	41.34	-	44751
36	-	41.42	-	44741

Table 1: Average PSNR and number of bits

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
test sequence	46.91	47.47	32567	32425
news clip	49.35	49.39	86004	82311

Table 2: Summary of the results (with shorter sequences)

where *'test sequence'* contains two abrupt scene transitions which were made by combining *'akiyo'* and *'foreman'* sequences.

Sequence	Average PSNR		Average number of bits	
	Conventional (12,3)	Proposed	Conventional (12,3)	Proposed
abrupt change	46.9	47.5	32567	32425
wipe	49.5	49.7	86348	82640
fade-out	42.3	42.4	43597	41262
fade-in	42.7	42.7	61041	52728
dissolve	39.0	39.4	46093	45854
camera motion	43.1	43.9	46594	46100

Table 3: Summary of the results with the proposed scheme for different scene transitions

Movie clip	Number of frames	Average PSNR		Average number of bits	
		Conventional (12,3)	Proposed	Conventional (12,3)	Proposed
Independence day	100 000	41.18	41.21	54837	44931
Xmen	30 000	40.68	40.76	49045	42196
The Patriot	30 000	41.02	41.08	52324	44372
Titanic	40 000	41.24	41.31	53678	44215
Saving Private Ryan	40 000	41.01	41.06	50987	41728

Table 4: Summary of the results (with movie clips)