FEED-FORWARD BUFFERING AND RATE CONTROL BASED ON SCENE CHANGE FEATURES FOR MPEG VIDEO CODER

Yoo-Sok Saw, Peter M. Grant, and John M. Hannah
Dept of Electrical Engineering, University of Edinburgh, Edinburgh, EH9 3JL, UK.
Tel: +44 131 6505655; fax: +44 131 650 6554
`e-mail: ys@ee.ed.ac.uk`

ABSTRACT

Video traffic management has been a challenging task in the fields of network management and multi-media communication. Transmission buffering is widely used to smooth bursty traffic and maintain a steady traffic level by adapting the incoming source traffic to the buffer. This paper describes an efficient adaptive buffering scheme which is based on feed-forward control to adaptively handle the non-stationary nature of bursty video traffic. The performance of a series of quantisation scale mapping curves is presented in terms of occupancy and video quality.

1 INTRODUCTION

A digital video compression algorithm such as MPEG generates a coded bit stream with variable rate. The term "traffic" is used to indicate the time-varying non-stationary nature of the video bit stream, since the video can alter from still to moving pictures.

In applications using transmission buffer, overflow and congestion may take place in bursty traffic or a low bit-rate CBR (constant bit rate) channel. While much research on traffic modelling for ATM (Asynchronous Transfer Mode) networks has been investigated [1], less research work has been carried out on buffering and rate control [2, 3, 4, 5, 6] in the video encoder. Thus, we concentrate on how effectively the buffer occupancy can be controlled within a given channel rate. We introduce features representing the amount of scene change which are used for calculating predictive buffer occupancy. ISO/IEC 11172 (MPEG1) and 13818 (MPEG2) are representative video compression standards to which the buffering scheme can ultimately be applied.

2 CONFIGURATION OF BUFFERING IN MPEG ENCODER

In an MPEG video encoder, quantisation step size can be adaptively controlled by buffer occupancy and picture details [7, 8]. When a buffer is used for traffic control, it is vital to use a reactive mechanism which controls the quantisation step size by feedback information on buffer occupancy. Exploitation of picture details can also be applied to control the picture quality. Occupancy-based reactive control becomes a dominant function, since maintaining a constant occupancy is the main objective. Figure 1 shows a diagram of the reactive buffering scheme.

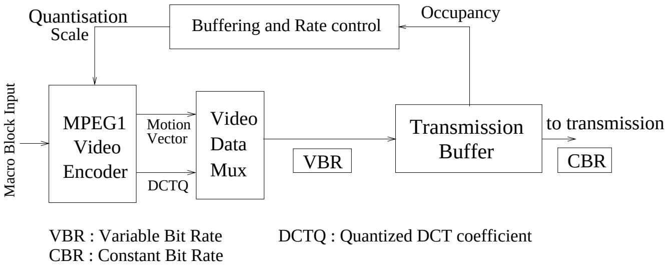


Figure 1: Reactive buffering scheme specified in MPEG

3 MAPPING QUANTISATION STEP SIZE

The target in rate control is to effectively map the buffer occupancy to the quantisation step size. Several different mapping curves have been investigated. They can be classified into linear [9, 10], piecewise linear [8, 3], and non-linear [11, 5, 6]. Two non-linear mapping curves, sigmoidal [11] and logarithmic [6] have been investigated.

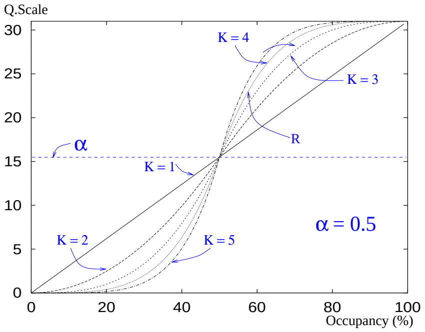


Figure 2: Sigmoidal quantisation scale mapping

The reference simulation software [9, 8] uses linear mapping i.e., $Qs(O) = O$ where $Qs(O)$ and O are the normalised quantisation scale and occupancy. In sigmoidal mapping, two nonlinear equations form a nonlinear mapping curve with the shape of a skewed S. K is a steepness factor and α is a control factor to determine the symmetry of each curve, e.g., if α is 0.5, the curve shows a symmetrical shape whose upper half and lower half form images. The combination of two curves forms a skewed sigmoidal mapping curve, Figure 2.

In logarithmic / exponential mapping the quantisation scale maps to a set of logarithmic and exponential curves. A value of quantisation scale is selected based on mean occupancy O^- for a specific period of time and the current occupancy, i.e.,

$$Qs(O^-, O) = \alpha(O^-)\log_{10}(\beta(O^-)O + 1) \qquad (1)$$

$$Qs(O^-, O) = \rho(O^-)\exp(\mu(O^-)O) - \mu(O^-) \qquad (2)$$

where O is the current occupancy, O^- is the mean occupancy, and $\alpha(O^-)$, $\beta(O^-)$, $\rho(O^-)$, and $\mu(O^-)$ are the coefficients which form the logarithmic and exponential curves. If the O^- value does not remain in a predefined occupancy range, a more nonlinear curve is selected. Otherwise, if O^- stays in the range, the curve selected becomes close to a straight line, Figure 3.

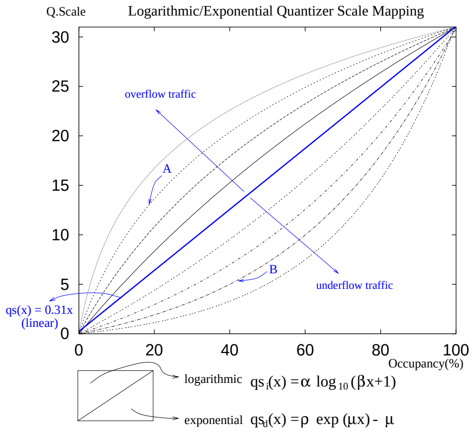


Figure 3: Logarithmic quantisation scale mapping

4 SCENE CHANGE FEATURES

This buffering scheme takes advantage of the frame delays for re-ordering picture types since MPEG encodes a P or I picture first, instead of having B pictures preceding them. Hence, if a scene change occurs in a B picture, the encoder realises this in advance of encoding the P or I picture. This can effectively be used for advance adjustment of quantisation scale. Figure 4 shows a detailed block diagram of feed-forward buffering combined with the MPEG main profile encoder. NQS (non-linear quantisation scale mapping) selects a non-linear curve to estimate future occupancy. The SCF (scene change function) and MVF (motion vector function) calculate the ratio of the variance of a difference picture frame to the variance of an input frame and the mean value of the motion vector function in a slice, respectively. Output values of NQS, SCF, and MVF go to the Q.scale control block to determine the best quantisation scale value for a slice. No quantisation scale adaptation is performed for a macro block or smaller scale of block since we assume that macro blocks have high correlation.

5 PREDICTIVE BUFFER OCCUPANCY

Framewise variances form an appropriate measure for scene change. It is an extension of the intra / inter decision for a macro-block specified in ISO11172-2 [7]. Two variances, *var_org* and *var_dif* are defined. *var_org* is the variance of input picture and *var_dif* is the variance of the difference picture between current input picture and previous input picture. Using these two variances, the graph is divided into four regions, Figure 5.

"A" represents the area with no dramatic scene changes and no subsequent abrupt changes in traffic since the *var_dif* is small. "B" is the area with higher *var_dif*, however, there would be no dramatic traffic level change at this picture since *var_org* is smaller than *var_dif*. The DCT encodes macro blocks in intra mode and its performance is considered to be better than encoding them in inter mode. Areas "C", and "D" may cause more dramatic increases in traffic. "C" and "D" are separated by the partition $var\_dif = var\_org$. Estimating the required number of coded bits ($C^p_{bits}(k)$ and $C^i_{bits}(k)$) of kth input frame is thus based on *var_dif*, *var_org*, and previous statistics of coded bits per frame, i.e.,

$$C^p_{bits}(k) = \frac{var\_org}{var\_dif} \times C_{bits}(P(k-n)) \qquad (3)$$

$$C^i_{bits}(k) = \frac{var\_dif}{var\_org} \times C_{bits}(I(k-n)) \qquad (4)$$

where $C_{bits}(P(k-n))$ and $C_{bits}(I(k-n))$ are the number of bits of the previously encoded I or P picture.

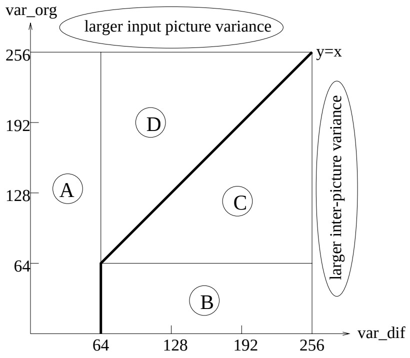


Figure 5: Plot for framewise inter / intra decision

As the pictures in both "C" and "D" may generate bursty traffic, it is necessary to choose either $C^p_{bits}(k)$ or $C^i_{bits}(k)$ when the value of var_dif/var_org is greater

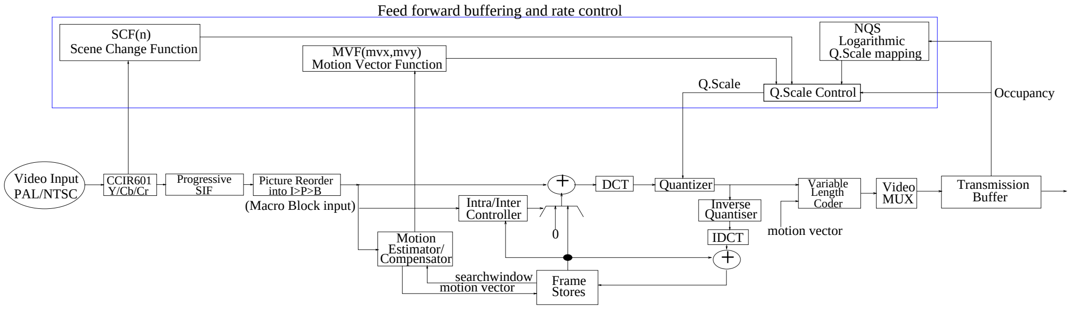


Figure 4: Feed-forward buffering scheme for MPEG

than 1. Once $C_{bits}(k)$ is determined, current buffer occupancy $O(k)$ is checked, i.e.,

$$O(k) = \text{int}\left(\frac{D_T \times FR_*^{-1} - O_{ms}(k)}{FR_*^{-1}}\right). \tag{5}$$

where the term $D_T \times FR_*^{-1}$ represents the maximum delay in ms and $O_{ms}(k)$ stands for current occupancy in ms. Therefore, $O(k)$ represents current buffer capacity in the number of frames. D_T is the delay target representing the maximum tolerance of coding delay in the number of frames, and FR_*^{-1} is the reciprocal of frame rate. If $O(k) > 1$, i.e., if the remaining capacity of the buffer is able to accept bits of a complete frame, the following equation applies with each $C^p_{bits}(k)$ or $C^i_{bits}(k)$.

$$O_F(k) \quad = \quad \frac{O(k) + \frac{1}{2}C^p_{bits}(k)}{D_T \times \text{MBF}} \tag{6}$$

$O_F(k)$ represents the predictive buffer occupancy and is used for selecting quantisation scale for the kth frame, $QSF(k)$, function,

$$QSF(k) = qs(T_b, O_F) \tag{7}$$

$QSF(k)$ is a function of the predictive occupancy $(O_F(k))$ and the traffic balance (T_b) defined as follows.

$$T_b = \frac{C_{bits}(\Delta k)}{L \times MBF} \tag{8}$$

where L is the number of frames in a set of pictures, e.g., for BBP or BBI L becomes 3, and MBF is the mean bits per frame. The function qs is one of the logarithmic / exponential mapping curves described in the previous section. Thus, the quantisation scale for a whole input frame is determined by short-term traffic history and future occupancy.

6 LOCAL ADAPTATION OF THE QUANTISATION SCALE

Parts of a picture may have different amount of visual information from the other parts of the picture. The value of $QSF(k)$ can be adaptively changed at each picture slice or horizontal stripe. In order to adapt $QSF(k)$ to the quantisation scale for each slice, motion vector values are used. The motion vector function, MVF_D is defined as follows:

$$MVF_D(s)$$

$$= \frac{\left|\sum_{k=0}^{MBN-1} mvx(k)\right| + \left|\sum_{k=0}^{MBN-1} mvy(k)\right|}{MBN \times (MVX_{max} + MVY_{max})} \tag{9}$$

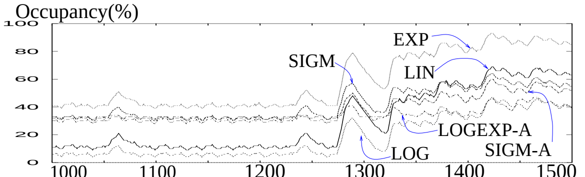
where s is slice number and MBN is the number of macro blocks in a slice, hence, $mvx(k)$ and $mvy(k)$ become motion vectors of macro block k. MVX_{max} and MVY_{max} are the maximum values of motion vectors for a specified motion search range. $MVF_D(s)$ is a directional motion vector function. As $MVF_D(s)$ ranges from 0 to 1, in order to make $MVF_D(s)$ a scaling factor of $QSF(k)$ it is multiplied by 2, and applied to $QSF(k)$, i.e.,

$$QSS(s) = 2. \times MVF_D(s) \times QSF(k) \tag{10}$$

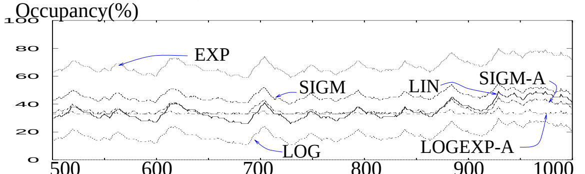
where $QSS(s)$ is the quantisation scale for the slice s.

7 SIMULATION RESULTS

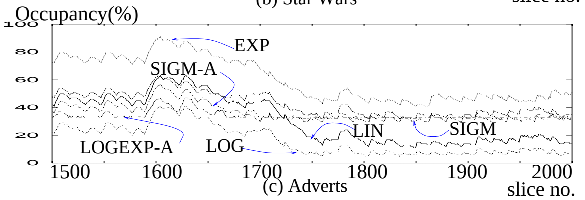
Six buffering schemes have been simulated over three video sequences "MAFBSU", "Star Wars", and "Adverts". "MAFBSU" is a cascaded video sequence of "Miss America", "Football", and "Susie". "Star Wars" and "Adverts" are taken from the movie "Star Wars" and television advertisements, respectively. The buffering schemes can be classified into adaptive and non-adaptive. The adaptive scheme has been applied to sigmoidal and logarithmic / exponential mapping curves, which adaptively change the steepness of the curves according to traffic balance. The non-adaptive method uses a single curve. The linear mapping scheme (LIN) takes a straight line as a mapping curve, as a performance benchmark. In the logarithmic mapping (LOG) and the exponential mapping (EXP), the curves A and B of Figure 3 are used, respectively. Sigmoidal mapping (SIGM) uses the curve R of Figure 2. Adaptive logarithmic / exponential mapping (LOGEXP-A) forms a combined scheme composed of LOG and EXP with adaptation. Adaptive SIGM (SIGM-A) is an adaptive scheme of SIGM.



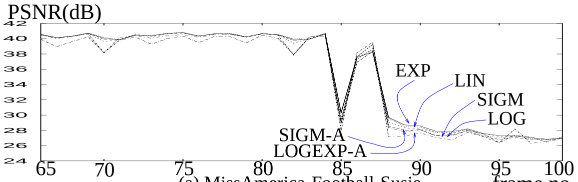
(a) MissAmerica-Football-Susie



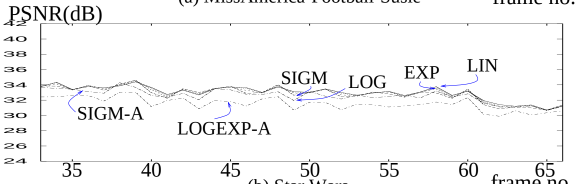
(b) Star Wars



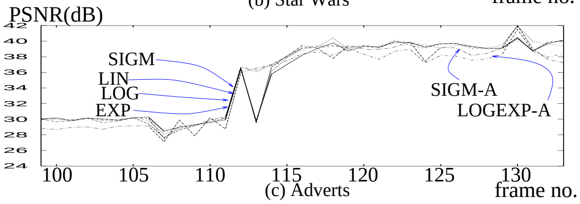
(c) Adverts



(a) MissAmerica-Football-Susie



(b) Star Wars



(c) Adverts

Figure 6: Occupancy and PSNR for 6 buffering schemes

Figure 6 shows the simulation results of the six methods both in terms of occupancy and peak SNR for these video sequences. In (a) there is a major scene change at slice 1300. The performance of LOG and LIN are lower than SIGM. EXP shows medium performance between SIGM and LOG. However, since it keeps lower quantisation scale values at low occupancy than SIGM and LOG, it generates more coded bits at low and initial occupancy. SIGM-A and LOGEXP-A show better performance since they adaptively change the mapping curves depending on the previous occupancy. SIGM-A shows steadier fluctuation in occupancy, however, at dramatic changes LOGEXP-A performs better since it can more quickly responds to the occupancy increase. Quality degradation, by controlling the quantisation scale, is seen in Figure 6 to be inversely proportional to the occupancy. The larger quantisation step sizes clearly lower the picture quality from the encoder, as shown in the PSNR plots.

8 CONCLUSION

In this paper we have investigated a feed-forward buffering scheme for video traffic using scene change features. Framewise variances and motion vectors in a slice have been effectively exploited to provide scene change information to control quantisation scale value. The performance of several different quantisation mapping curves (linear, sigmoidal, logarithmic, and exponential) has been evaluated. The adaptive scheme using combined logarithmic / exponential curves shows superior performance in terms of buffer occupancy.

9 ACKNOWLEDGEMENTS

Yoo-Sok Saw acknowledges the support of *The British Council* and *GoldStar Information and Communications Ltd., Korea.*

References

[1] V. S. Frost and B. Melamed, "Traffic modeling for telecommunications networks," *IEEE Communications Magazine*, vol. 32, pp. 70–81, Mar. 1994.

[2] C.-T. Chen and A. Wong, "A self-governing rate buffer control strategy for pseudoconstant bit rate video coding," *IEEE Transactions on Image Processing*, vol. 2, no. 1, pp. 50–59, 1993.

[3] J.-P. Leduc, "Bit-rate control for digital tv and hdtv codecs," *Signal Processing Image Communication*, vol. 6, pp. 25–45, 1994.

[4] L.-W. Lee, J.-F. Wang, J.-Y. Lee, and C. C. Chen, "On the error distribution and scene change for the bit rate control of mpeg," *IEEE Transactions on Consumer Electronics*, vol. 39, no. 3, pp. 545–554, 1993.

[5] J. Zdepski and D. Raychaudhuri, "Statistically based buffer control policies for constant rate transmission of compressed digital video," *IEEE Transactions on Com*, vol. 39, pp. 947 – 957, 1991.

[6] Y.-S. Saw, P. M. Grant, and J. M. Hannah, "Reduced storage transmission buffer designs for mpeg video coder," in *International Conference on Image Processing and its applications '95*, pp. 608 – 612, IEE, July 4 - 6, 1995, Edinburgh, UK.

[7] ISO/IEC JTC1/SC29, *Information Technology - Coding of moving picture and associated audio for digital storage media at up to about 1.5 Mbit/s - Part 2 : Video*, 1 ed., 1993.

[8] ISO/IEC JTC1/SC29/WG11, *ISO/IEC DIS 13818-2: Information Technology - Generic Coding of Moving Pictures and Associated Audio Information: Video*, 1994.

[9] ITU-T Study Group XV, *Recommendation H.261 : Line Transmission of Non-telephone Signals Video codec for audiovisual services at p × 64 Kbps*, 1993.

[10] ISO-IEC/JTC1/SC29/WG11, *Test Model 1*, 1 ed., 1992.

[11] Y. Kim and H. Kim, "A new buffer control strategy for image data compression," *IEEE Trans. on Consumer Electronics*, vol. 40, pp. 932 – 937, Nov. 1994.