NONLINEAR PREDICTIVE RATE CONTROL FOR CONSTANT BIT RATE MPEG VIDEO CODERS

Yoo-Sok Saw, Peter M. Grant, John M. Hannah and Bernard Mulgrew
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

A nonlinear predictive approach has been employed in MPEG (Moving Picture Experts Group) video transmission in order to improve the rate control performance of the video encoder. A nonlinear prediction and quantisation technique has been applied to the video rate control which employs a transmission buffer for constant bit rate video transmission. A radial basis function (RBF) network has been adopted as a video rate estimator to predict the rate value of a picture in advance of encoding. The quantiser control surfaces based on nonlinear equations, which map both estimated and current buffer occupancies to a suitable quantisation step size, have also been used to achieve quicker responses to dramatic video rate variation. This scheme aims to adequately accommodate non-stationary video in the limited capacity of the buffer. Performance has been evaluated in comparison to the MPEG2 Test Model 5 (TM5) in terms of the buffer occupancy and picture quality.

1. INTRODUCTION

The rate control algorithm of TM5 is based on the previous history of video rate, global and local picture complexity measures. This technique is known to be inappropriate for non-stationary videos with frequent scene changes or rapid motion [1], since the statistical properties are changing accordingly. Therefore, for such video, a different approach is required. Recently, we have developed a feed-forward video rate control technique using scene change features [2, 3]. The main feature of the technique is that the predictive estimation of the video rate value is derived from a series of scene change features. The employed prediction technique is based on a one-step ahead linear prediction using previous video rate data in a heuristic way. In this paper a nonlinear estimation technique is applied in order to more effectively control dramatic scene changes. The RBF-network [4] was designed to estimate the video rate using the scene change features of the input video so that the quantisation step size can be adjusted in advance of encoding the picture. The scene change features are framewise variances and picture type information. The nonlinear quantiser control surface changes the quantisation step size depending on the estimated video rate and the current buffer occupancy. Three performance evaluation measures were used; number of coded bits per frame, buffer occupancy and peak signal-to-noise ratio (PSNR).

2. A RBF-NETWORK RATE ESTIMATOR

Before describing the detail of the RBF-network estimator, an example of performance of the linear predictive video rate control techniques described in [2] is shown in Figure 1.

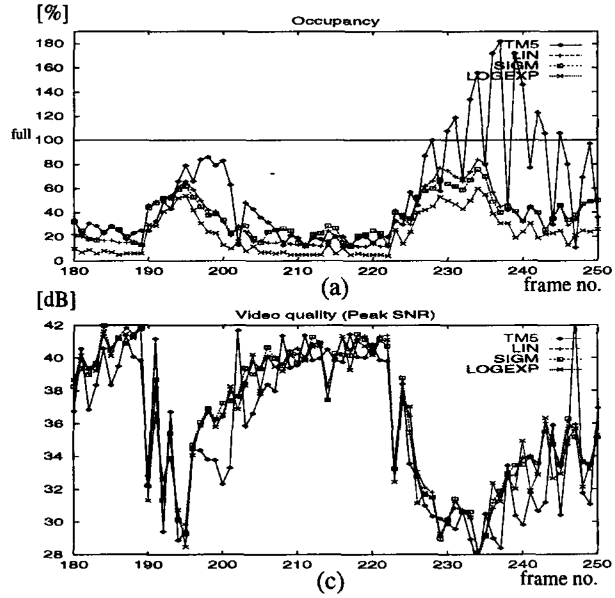

Figure 1. A performance example of linear predictive rate control techniques: (a) buffer occupancy, (b) Peak SNR.

TM5 represents the video rate control technique employed in the MPEG2 TM5. The other three methods (LIN, SIGM and LOGEXP) are based on the same linear predictive method [2] but a different quantisation control function is applied to each method. A linear function and a sigmoidal function are used for the methods, LIN and SIGM, respectively. For LOGEXP, a combination of logarithmic and exponential functions is employed, which is collectively named "unimodal" in later sections of this paper, instead of the linear or sigmoidal function. The results shown in Figure 1 are obtained from the MPEG2 encoding of "JFK" movie sequence at the 1024 kbits/s channel rate and the 30 frames/s frame rate. TM5 shows the worst performance often reaching the buffer full state, also showing wider variations in the PSNR. On the other hand, the other schemes exhibit far smaller variation in the occupancy and the PSNR alike. Particularly, LOGEXP shows the most stable occupancy profile with the very similar quality to LIN and SIGM.

The RBF-network video rate estimator aims to further improve the performance by applying its nonlinear predictive properties to non-stationary signals. The innovated MPEG2 encoder contains three additional rate control functions as shown in Figure 2: the scene change calculator, the rate estimator and the nonlinear quantiser control. The scene change calculator outputs the two variances, $var\_org(k)$ and $var\_dif(k)$, and the picture type information, $ptype(k)$, as inputs for the rate estimator. The predicted video rate, $\widehat{cbf}(k)$, is added to the current occupancy, $O(k-1,n)$, to form the predicted occupancy, $\widehat{O}(k,n)$, used by the the nonlinear quantisation control, which finally outputs the quantisation scale value, $Qs(k,n)$. $var\_org(k)$ and $var\_dif(k)$ represent the variance within an input picture and the variance between the input picture and the previous picture, respectively. $ptype(k)$ has a single integer for a particular picture type (I, P and B), thus it forms a cyclic time series as k increases such as 8,4,2,2,4,2,2,... for I,P,B,B,P,B,B....

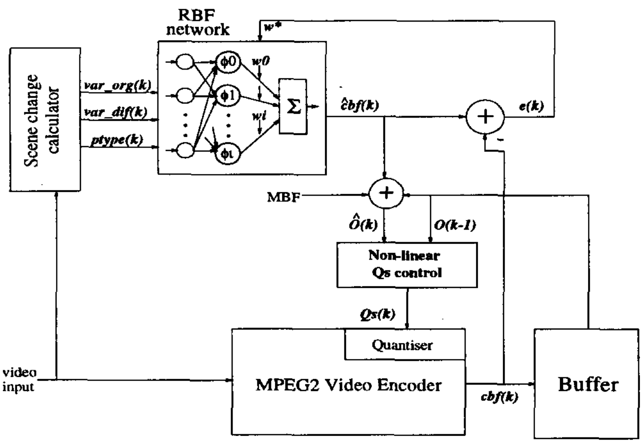

Figure 2. A RBF rate estimator-based MPEG2 video encoder.

A RBF network consists of centres with a radial basis function and linear weights, Figure 3, defined in the following equations:

$$\widehat{cbf}(\mathbf{x}) = \sum_{i=1}^{N} w_i \phi(||\mathbf{x}-\mathbf{x}_i||)$$

$$\phi(||\mathbf{x}-\mathbf{x}_i||) = \exp(-\frac{||\mathbf{x}-\mathbf{x}_i||^2}{2\sigma^2}) \tag{1}$$

where $\widehat{cbf}(\mathbf{x})$ is the output of the RBF network, w_i is the linear weight, $\mathbf{x}$ is an input vector containing scene change features, and $\mathbf{x}_i$ represents the selected centre. The radial function $\phi()$ is a Gaussian function. The Euclidian distance between the input and the centre ($||\mathbf{x}-\mathbf{x}_i||$) determines the output value of the RBF layer. σ^2 represents the variance of $\mathbf{x}$. The RBF centres may be selected by the orthogonal least square (OLS) algorithm [5]. The OLS algorithm selects representative RBF centres when supervised learning is used. However, in the case of the running MPEG2 encoder, supervised learning cannot be used properly since the nature of realistic video is non-stationary. A supervision of scene change features for a short period of time does not provide an entire insight into the whole properties of the non-stationary video. Thus the k-means clustering algorithm is used, which adaptively updates the RBF centres depending on variations in scene change features. The centres are updated as follows [6]:

$$\mathbf{x}_j(k) = \mathbf{x}_j(k-1) + g_c(cbf(k) - \mathbf{x}_j(k-1)) \tag{2}$$

where $\mathbf{x}_j$ is the jth centre and the constant g_c controls the learning rate. The linear weights, w_i, are optimised recursively in a least square sense (RLS) [7].

3. QUANTISATION CONTROL SURFACES BASED ON NONLINEAR EQUATIONS

The quantisation step size is the core parameter which controls the occupancy. The goal of the buffer-based rate control technique is to effectively map the occupancy to the quantisation step size specified in the MPEG2 standard. Several different control functions have been proposed. They can be classified into linear, piecewise linear and nonlinear [8, 2]. This paper focuses on the nonlinear control functions. The nonlinear quantiser control, as shown in Figure 4, uses both the current (b) and the predicted buffer occupancies (a) to select a quantiser control curve for the quantisation scale (c). It changes between linear and nonlinear curves depending on the predicted occupancy. If a dramatic change in the occupancy is predicted, then it changes the shape of the curve towards a more distorted one, otherwise, it selects a curve close to the linear function. The final quantisation scale value is determined by the current occupancy. In this paper two nonlinear mapping surfaces are examined, sigmoidal and unimodal as shown in Figure5.

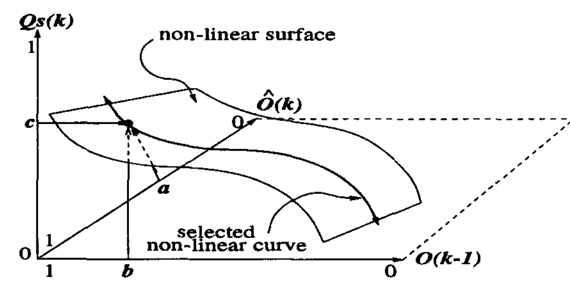

Figure 4. Nonlinear quantisation step size mapping.

The sigmoidal surface (SIGM) is formed by changing the steepness of a sigmoidal function. The unimodal surface (UNIM) consists of a combination of an exponential part and a logarithmic part. The quantisation scale for the macro block n, $Qs(k,n)$, is calculated as follows:

$$\begin{aligned} Qs(k,n) &= f(O(k-1,n), \widehat{O}(k)) \\ \widehat{O}(k) &= O(k-1,n) + \widehat{cbf}(k) - \mathrm{MBF} \end{aligned} \tag{3}$$

where MBF is the target video rate given by the mean value of bits per picture. $f()$ is one of the nonlinear mapping surfaces, and a value of $Qs(k)$ for the next macro block is determined for given $O(k-1)$ and $\widehat{O}(k)$. The two surfaces, shown in Figure 5, are expressed in equations of $f_S(O(k-1,n), \widehat{O}(k))$ and $f_U(O(k-1,n), \widehat{O}(k))$, which represent surfaces of SIGM and UNIM, respectively:

$$\begin{aligned} f_S(\bullet) &= \alpha(\frac{1}{\alpha}O(k-1))^{(T\widehat{O}(k)+1)} \\ &\times \ trunc(1+\alpha-O(k-1)) \\ &+ \ \left(1-(1-\alpha)\left(\frac{1}{1-\alpha}(1-O(k-1))\right)\right)^{(T\widehat{O}(k)+1)} \end{aligned}$$

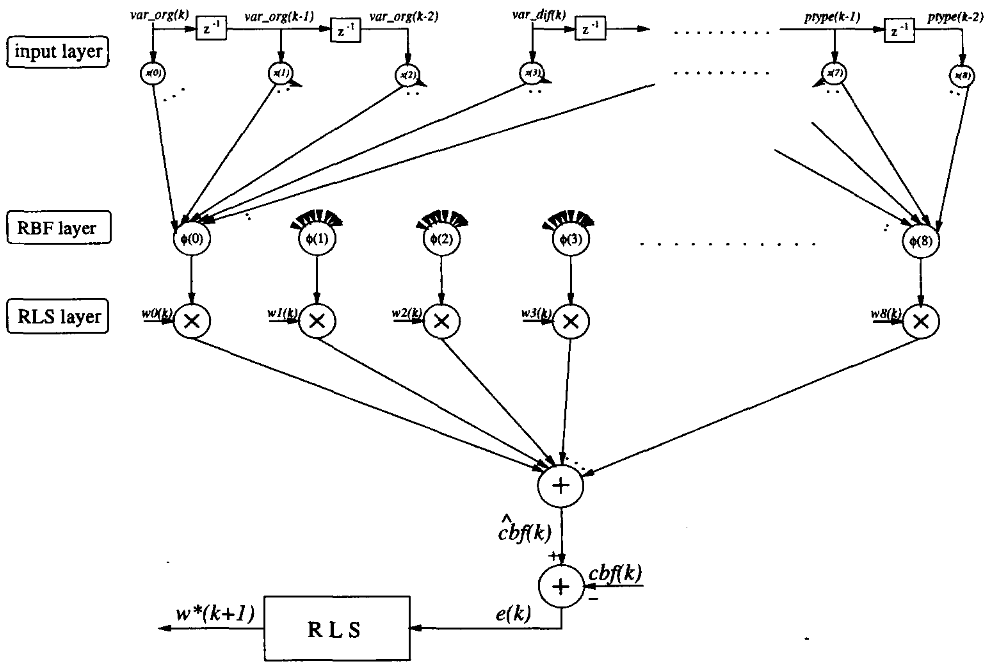

Figure 3. RBF predictor with 3 inputs and 9 taps.

$$\times \quad trunc\left(\frac{O(k-1)}{\alpha}\right) \tag{4}$$

$$f_U(\bullet) \quad = \quad O(k-1)^{C/(T\widehat{O}(k)+1)} \tag{5}$$

where *trunc* is a truncation function to output 0 or 1 depending on its input value.

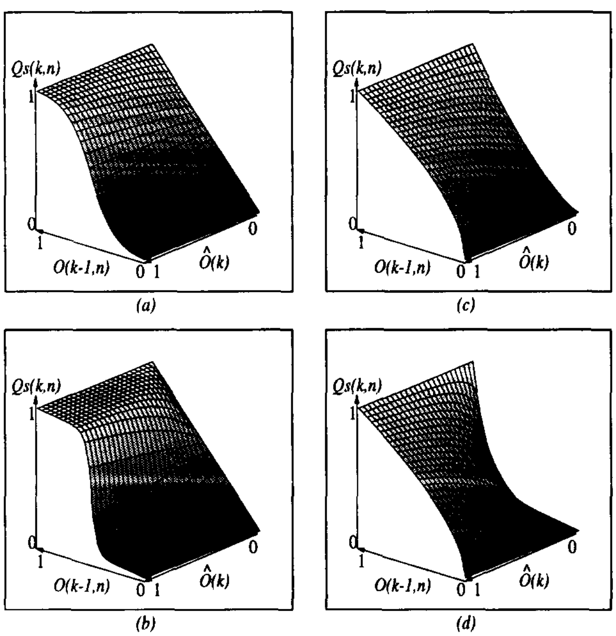

Figure 5. Sigmoidal surface: (a) T_{max} = 3, (b) T_{max} = 13. Unimodal surface: (c) T_{max} = 3, (d) T_{max} = 13.

The torsion factor, T, represents the shape distortion of the control surfaces, ranging from 1 to T_{max} which represents its maximum value, varying with channel rates, as shown below :

$$T_{max} = \frac{A}{channel\_rate} \tag{6}$$

where A is a constant. When the *channel_rate* is high, the expanded channel capacity can handle the video rate fluctuation, hence, a small T_{max} can be used. For a lower channel rate, a higher value is assigned to provide the surface with a larger torsion. The constants, α and C, are balancing factors forming the surfaces in a balanced or an unbalanced shape. Figure 5 shows two extreme cases of T_{max} for specific values 3 and 13. The surfaces with a larger T_{max} value exhibit more torsion.

4. SIMULATION RESULTS

Two video sequences, "Starwars" and "JFK", were used in simulations to give frequent scene changes and nonstationary input video data to the encoder. The sequence we used contains 300 frames captured from parts with rapid motion and dramatic scene changes. "JFK" has more dramatic scene changes: transitions between coloured and monochrome scenes and rapid zooming. The video encoder is set to operate at a channel rate at 1024 kbits/s and a frame rate at 30 frames/s. It has a buffer with the size of twice of MBF. For each value of *ptype*(), the integers 10, 8 and 6 are assigned to I, P and B pictures, respectively. We first assessed the performance of nonlinear surfaces depending on the values of T_{max}, as shown in Table. 1. NFVR in the middle column stands for normalised fluctuation of the video rate which represents the total amount of $cbf(k)$ fluctuation:

$$\mathrm{NFVR} = \frac{\sigma}{1+\sigma}, \qquad \sigma^2 = \mathrm{E}\left[\left(\frac{cbf(k)}{\mathrm{MBF}} - 1\right)^2\right] \tag{7}$$

where $\frac{cbf(k)}{\mathrm{MBF}}$ represents instantaneous fluctuation. Both surfaces show better rate control performance with reduced variance as T_{max} increases. While SIGM exhibits less fluctuations in video rate, UNIM appears superior in terms of mean PSNR with the standard deviation (std. dev.) close to SIGM.

Starwars		Occupancy(%)		coded bits / frame (bits)		PSNR (dB)	
T_{max}		mean	std.dev.	NFVR	std.dev.	mean	std.dev.
TM5		41	10.78	0.285	13704	33.70	1.96
5	SIGM	51	0.51	0.030	1060	31.84	2.36
	UNIM	26	4.89	0.122	4728	33.91	2.54
7	SIGM	51	0.39	0.027	958	31.65	2.35
	UNIM	18	4.58	0.117	4516	33.89	2.55
9	SIGM	51	0.33	0.023	802	31.57	2.35
	UNIM	13	4.16	0.111	4274	33.87	2.55
11	SIGM	51	0.26	0.022	782	31.39	2.34
	UNIM	10	3.69	0.106	4043	33.84	2.55
13	SIGM	51	0.24	0.020	701	31.33	2.35
	UNIM	7	3.43	0.099	3755	33.82	2.55

Table 1. Effect of changing the torsion factor.

Two rate control schemes were evaluated in comparison to TM5; a linear rate estimator optimised with the recursive least square (RLS) algorithm, which has no RBF layer and the RBF-network estimator shown in Figure 3. Both schemes employed UNIM surface for better video quality. For the nonlinear quantisation mapping surfaces, T_{max} is set to 7. Figure 6 shows profiles of the three schemes for frames 180 to 250 where dramatic scene changes occur. Table 2 summarises the performance for all 300 frames.

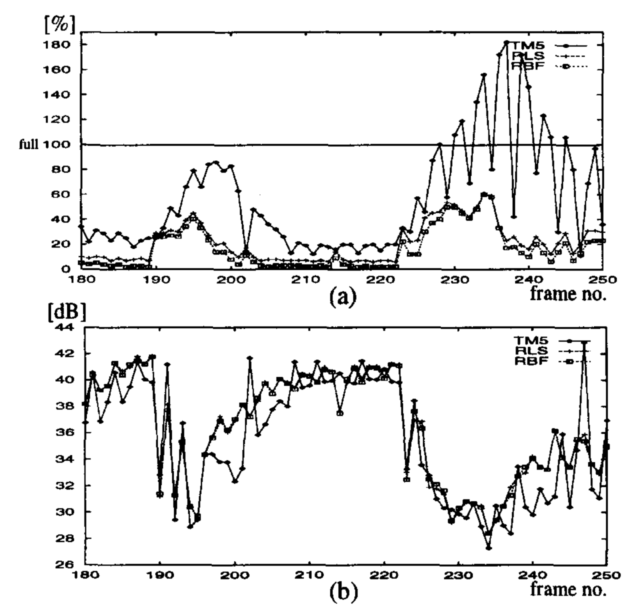

Figure 6. Performance of rate control algorithms ("JFK" sequence): (a) buffer occupancy, (b) PSNR.

JFK	Occupancy(%)		coded bits / frame (bits)		PSNR (dB)	
	mean(max.)	std.dev.	NFVR	std.dev.	mean	std.dev.
TM5	39 (172)	26.57	0.378	20517	36.08	3.34
RLS	17 (76)	10.33	0.188	7687	36.49	3.83
RBF	11 (61)	10.07	0.169	6661	36.48	3.83

Table 2. Mean and standard deviation of performance measures.

TM5 exhibits inferior control capability to the two other schemes in terms of both occupancy. Although the std.dev. in PSNR appeared smaller for TM5 than for RLS and RBF, the average PSNR of TM5 is slightly lower than the two others. RBF appeared to be capable of keeping the occupancy lower with a smaller std.dev. than RLS, without quality degradation. Note that the NFVR and the std. dev. of coded bits/frame are considerably smaller than those of RLS, and that the performance is better than Figure 1.

5. CONCLUSION

The MPEG2 video rate control technique, which is based on a nonlinear predictor and quantisation control, has been investigated for a constant bit rate transmission. The RBF network rate estimator appeared to improve the rate control performance in terms of video rate and video quality, when it is used in combination with the nonlinear quantisation technique employing the unimodal function. This signifies that the nonlinear predictive technique may substantially enhance the performance of the rate control mechanism when processing non-stationary video.

6. ACKNOWLEDGEMENTS

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

REFERENCES

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

[2] Y. Saw, P. M. Grant, and J. M. Hannah, "Feed-forward buffering and rate control based on scene change parameters for MPEG video coder," in *EUSIPCO'96*, vol. II, pp. 727–730, EURASIP, 10 - 13, September, 1996.

[3] Y. Saw, P. M. Grant, J. M. Hannah, and B. Mulgrew, "Radial basis function video rate estimator for constant bit rate MPEG coders," *IEE Electronics Letters*, vol. 32, no. 21, pp. 1969–1971, 1996.

[4] S. Haykin, *Neural networks : a comprehensive foundation.* Macmillan, 1994.

[5] E. S. Chng, S. Chen, and B. Mulgrew, "Efficient computational schemes for the orthogonal least square algorithm," *IEEE Trans. Signal Processing*, vol. 43, no. 1, pp. 373 - 376, 1995.

[6] S. Chen, B. Mulgrew, and S. McLaughlin, "Adaptive Bayesian equalizer with decision feedback," *IEEE Trans. Signal Processing*, vol. 43, no. 5, pp. 1937 – 1946, 1995.

[7] S. Haykin, *Adaptive Filter Theory.* Prentice-Hall, 1991.

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