Rate Prediction Model for Video Transcoding Applications

Pedro F. Correia[1,2], Vítor Mendes Silva[1,3], Pedro A. Amado Assunção[1,4]

[1]Institute of Telecommunications
Pólo II, FCTUC, Pinhal de Marrocos, 3030-290 Coimbra, Portugal
Phone: +351 239796236; Fax: +351 239796293

[2]Polytechnique Institute of Tomar-EST, [3]University of Coimbra-DEEUC, [4]Polytechnique Institute of Leiria-ESTG
{pcorreia, vitor, amado}@co.it.pt

Abstract - **This paper describes a model for rate prediction in video transcoding applications. The model is based on a log-linear approximation of rate-quantisation characteristics. It is shown that such a model is capable of accurately predicting the rate of transcoded pictures, given one R-Q point derived from the original coded signal. Its application in mobile multimedia communications where hard transcoding is required, is also discussed in the paper.**

I. INTRODUCTION

The forthcoming multimedia services over wireless mobile channels will certainly include a major component of coded video signals. The inherent high bandwidth of this type of signals is not always matched to the transmission channels or user terminals due to a number of reasons. For example, the time-varying characteristics of mobile channels, the diversity of client terminal capabilities, and user demand options available through interactivity, pose a great deal of different constraints. While bandwidth and channel errors are the most limiting factors on the network side, processing power, memory and display size are the scarcer resources at the receiver side. All these constraints together demand for specific coding parameters for matching mobile applications, which are obviously more strict than those used for fixed broadband applications.

Furthermore, many of the current multimedia architectures are client-server systems not only for distribution of pre-encoded media but also for real-time broadcasting [1,2]. In this case the coded signals are transmitted from the server through heterogeneous channels and it is very likely that many of these do not match the bandwidth requirements for real time transmission. In order to cope with different constraints, a single media stream can still be delivered to different clients or networks by inserting transcoding devices in the transmission path [3,4]. In general, a transcoded stream is adapted for having a lower bandwidth and/or a different coding format than those of the input [5].

In this paper we describe a rate prediction model for hard transcoding applications, where a coded video stream must be highly constrained in order to cope with low bandwidth channels, such as mobile. The hard transcoding problem arises whenever the media adaptation function demands a great deal of bandwidth reduction such that it is no longer possible to maintain the frame rate of the original coded signal during the whole session. In this case the transcoding algorithm is often forced to skip several frames in those video segments where the scene is more difficult to encode at very low bit rates, *i.e.* high spatial complexity and fast motion. Whenever skipping occurs in consecutive frames, the decoded signal quality becomes very poor, because of a greater inefficiency in the interpolation process at the decoder.

By using the proposed rate prediction model it is possible to avoid consecutive frame skipping by computing an estimate of the number of bits for few frames ahead of that being currently transcoded. The advantage of knowing the number of bits before actual transcoding takes place, is that a more intelligent algorithm can be used for frame skipping than the straightforward approach of discarding a frame whenever the output buffer level exceeds a predefined threshold.

The model proposed in this paper is based on the logarithmic behaviour of the rate-quantisation (R-Q) characteristics of video signals. The model is then used for estimating the number of bits of transcoded frames. The results show that a good prediction of the buffer occupancy can be obtained and therefore any situation of buffer overflow which would imply frame skipping.

II. RATE-QUANTISATION CHARACTERISTICS

In order the derive a rate prediction model we have studied the R-Q characteristics of several video signals using the MPEG-4 visual coding standard [6]. Although we have done several simulations with different video sequences, for the results presented throughout the paper we have used the Coastguard sequence, QCIF format, 30Hz.

In figure 1, we show the R-Q characteristics of frames 3, 30, 51, 72, 120 and 180, for quantisation parameters from 3 to 31 (the full range is 1-31). These R-Q characteristics have two important features that can be seen in figure 1. Firstly all the characteristics exhibit an exponential behaviour which is confirmed by comparing with the reference curve also shown in the figure. Secondly, we can observe a strong similarity between all curves, in the sense that their derivatives are quite similar and only some shifting parameter make the difference between each other. The reference curve shown in the figure is the function (2.1) with $k = 12$ and $m = -1$.

$$R = e^k Q^m \qquad (2.1)$$

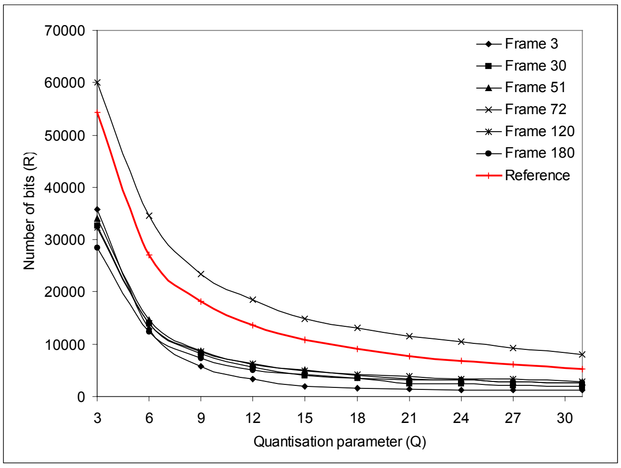

Figure 1. R-Q characteristics of Coastguard sequence

By using logarithmic scales for R and Q we can confirm the above statements. This is shown in figure 2, where the same curves of figure 1 are drawn as $\log R = f(\log Q)$. From this figure, it is quite obvious the log-linear characteristic of the R-Q curves, as well as the similarity between their derivatives. By drawing the straight line given by expression (2.2) and labelled in the figure as "reference", we could say that a log-linear model would make a reasonable approximation for R-Q curves. This is the reasoning behind the R-Q model described in the next section.

$$\log R = k + m \log Q \qquad (2.2)$$

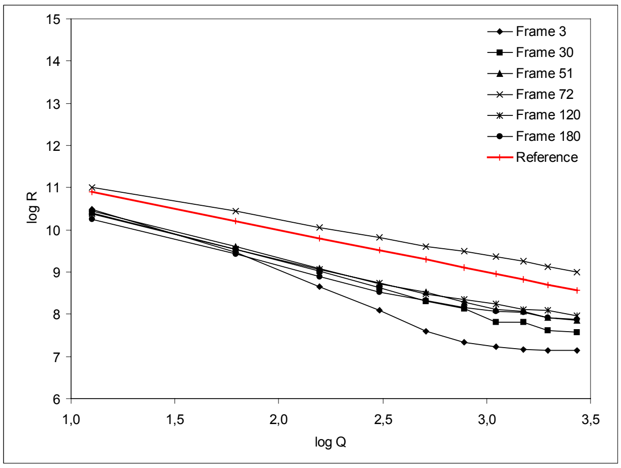

Figure 2. Log-linear approximation of the R-Q characteristics

III. R-Q MODEL

The R-Q model described in this section is based on a log-linear approximation as suggested by the results shown above. The proposed model is given by the following expression,

$$\log R = k + m \log Q \qquad (3.1)$$

where R is the number of bits for one frame, Q is the average quantisation parameter, and k, m are the model parameters. Given any two points R_1-Q_1 and R_2-Q_2, the parameter m can be computed as follows,

$$m = \frac{\log R_2 - \log R_1}{\log Q_2 - \log Q_1} \qquad (3.2)$$

while k can be obtained from equation (3.1) given m and one R-Q point. Note that, in transcoding applications at least one R-Q point per frame is always available beforehand, because the input signal is a compressed bitstream.

As it can be seen from equations (3.1) and (3.2) the model is a straight line approximation of the $\log R$-$\log Q$ characteristics on a frame basis. The parameter m defines the inclination of the curve while k is the intersection point with the vertical axis. Since the model is intended to be used for rate prediction, the parameters m and k must be known for computing the estimated rate.

From our simulations and from figure 2 we can observe that m does not exhibit strong variations along the sequence, being almost invariant in consecutive frames. Therefore, for future frames, we can predict m from the most recently transcoded frames by using some predefined prediction function. A simple method is to set a fixed m for all frames within a temporal window of limited size. In regard to parameter k, this can be easily computed from the R-Q point available from the input bitstream.

IV. RATE PREDICTION ALGORITHM

Based on the previous R-Q model we have implemented a rate prediction algorithm for transcoding applications within the MPEG-4 framework. A constant bit rate (CBR) transmission is assumed thus, an output buffer must be used. The proposed algorithm computes an estimate of the Q parameter and buffer occupancy, n frames ahead of the current one. If the buffer level exceeds a predefined threshold, then the corresponding frame might be labelled for possible skipping.

According to equation (3.2), after transcoding the current frame i, m_i is computed as follows:

$$m_i = \frac{\log R_i - \log \widehat{R}_i}{\log Q_i - \log \widehat{Q}_i} \qquad (4.1)$$

where $R_1 - Q_1$ and $\widehat{R}_i, \widehat{Q}_i$ are the $R-Q$ points obtained from the input bitstream and the current transcoded frame, respectively. Then we assume that m_i keeps the same value for those future frames that have a temporal distance from the current one between 1 and n, *i.e.,*

$$m_i = m_{i+1} = \cdots = m_{i+n} \qquad (4.2)$$

Following the same strategy as in [7], we have defined the target buffer fullness to 50% of its size, hence the target rate R_n for frame n is initially set to the ratio between the transmission rate R_B and the frame rate F,

$$R_n = \frac{R_B}{F} \qquad (4.3)$$

and then adjusted according to the difference between 50% of the buffer size and the current buffer level B_i, as given by equation (4.4).

$$R_n^{'} = R_n \frac{B_i + 2(B_{\max} - B_i)}{2B + (B_{\max} - B_i)} \qquad (4.4)$$

After computing $R_n^{'}$ a verification procedure takes place for checking buffer underflow or overflow and, if this is the case, a minimum or maximum level respectively, is set. Then, for frame n, k_n is computed as follows,

$$k_n = \log R_n^{'} - m_i \log Q_n \qquad (4.5)$$

After having computed m_n and k_n it is possible to determine an estimate for the average quantisation parameter Q_n to be used in frame n as given by,

$$Q_n = R_n^{1/m_n} e^{-k_n/m_i} \qquad (4.6)$$

with Q_n limited to $1 \le Q_n \le 31$. If in expression (4.6) $Q_n > 31$, then Q_n is set to 31 and a new target rate is computed for frame n given by equation (4.7) instead of equation (4.3).

$$R_n = e^{k_n} 31^{m_i} \qquad (4.7)$$

The expected buffer fullness after transcoding frame n is finally given by,

$$B_n = B_i + \sum_{j=i+1}^{i+n} R_j - n\frac{R_B}{F} \qquad (4.8)$$

where B_n and B_i are the buffer occupancies after frame n and i respectively, and R_j is the expected number of bits for frame j.

V. SIMULATION RESULTS

In order to test the accuracy of the proposed model we have simulated a transcoding scenario using a MPEG-4 coded bitstream as input. For this purpose the Coastguard sequence was encoded at 256 Kbps, QCIF format with frame rate of 30Hz. The reason for using this sequence is because it has a difficult scene to encode around frames 60-70, which makes the buffer fullness to increase and therefore it is a good test case for a rate prediction algorithm. It is also around these frames that standard encoders are forced to skip some frames in very low bit rate coding.

The buffer size was set to implement a delay of 500ms which corresponds to half of the transmission rate *e.g.* with R_B=256kbps, B_{max} = 128Kbits.

Figures 3, 4, 5 and 6 show the performance of the proposed rate prediction algorithm for the case of transcoding from 256 Kbps into 56Kbps. Each figure shows a comparison between the predicted buffer level, obtained through our model based algorithm, and the actual buffer level observed after transcoding each frame. In figure 3 the prediction is for the next frame while in figures 4, 5 and 6 the prediction is for 2, 3 and 4 frames ahead of the current one.

As it can be seen from these figures, the actual buffer level accurately follows the log-linear model though with smaller variations. For longer temporal distances, e.g., n=4, the rate prediction deviates more from the actual data as expected, because the model error accumulates over each predicted frame rate. Overall the results show that this model can be used with an acceptable level of confidence.

VI. APPLICATIONS

As already mentioned, the proposed model based algorithm is envisaged for hard transcoding applications where frame skipping is unavoidable. In this case, if only one single frame is skipped once in while, the decoder can always interpolate from the previous and next frames received. Still an acceptable quality is expected to be obtained, though very much dependent on the interpolation algorithm used at the receiver. However, most of the frame skipping methods are essentially based on the buffer occupancy, forcing the transcoding algorithm to skip one frame whenever the buffer level exceeds a predefined threshold. In general this methods cannot avoid consecutive frame skipping which makes the decoder task of interpolating the missing frames extremely difficult.

By using the proposed rate prediction model for future frames to be transcoded, it is possible to know in advance whether consecutive skipping is likely to occur. Then, if this is the case, the transcoder can decide beforehand which frames should be skipped in order to avoid consecutive skipping. Therefore, the decoded video signal is expected to be of higher quality if consecutive skipping is prevented at the transcoder. This is the subject of our future work.

This system finds useful applications in mobile multimedia services where the wireless network has much lower available bandwidth than that of the distribution network. In this case, media transcoding devices are necessary at the network interconnection nodes where the proposed work finds a useful application.

VII. CONCLUSIONS

In this paper we have described a rate prediction model for hard video transcoding systems where consecutive frame skipping is expected to occur. Based on the R-Q curves of MPEG-4 coded video we have derived a log-linear model for computing the rate and quantisation parameters of several frames ahead the transcoding algorithm. A buffer based algorithm was implemented for this purpose taking advantage from the fact that, in video transcoding one R-Q point is available for each frame beforehand. The simulation results show a good efficiency of the proposed model in predicting the transcoding buffer fullness and thus any skipping in future frames. The application of the proposed algorithm to improving transcoded video signals was also discussed in a mobile multimedia communications scenario.

VIII. REFERENCES

[1] G. Conklin, G. Greenbaum, K. Lillevold, A. Lippman, Y. Reznik, Video Coding for Streaming Media Delivery on the Internet, IEEE Transactions on Circuits and Systems for Video Technology, Vol 11, No 3, pp. 269-281, March 2001.

[2] Real Networks, Live Broadcast Distribution with RealSystem Server 8, RealSystem iQ White Paper, December 2000.

[3] P. A. A Assuncao and M. Ghanbari, Transcoding of Single-Layer MPEG Video Into Lower Rates, IEE Proceedings on Vision, Image and Signal Processing, Vol. 144, No 6, pp. 377-383, Dec. 1997.

[4] T. Warabino, S. Ota, D. Morikawa, and M. Ohashi, Video Transcoding Proxy for 3Gwireless Mobile Internet Access, IEEE Communications Magazine, pp. 66-71, October 2000.

[5] T. Shanableh and M. Ghanbari, Heterogeneous video transcoding to lower spatio-temporal resolutions and different encoding formats, IEEE Transactions on Multimedia, Vol. 2, No 2, pp. 101-110, June 2000.

[6] ISO/IEC 14496-2: Information Technology - Generic Coding of Audio-Visual Objects – Part 2:Visual, Vancouver, July 1999.

[7] H. Lee, T. Chiang, Y. Zhang, Scalable Rate Control for MPEG-4 Video, IEEE Transactions on Circuits and Systems for Video Technology, Vol 10, No 6, pp. 878-894, September 2000.

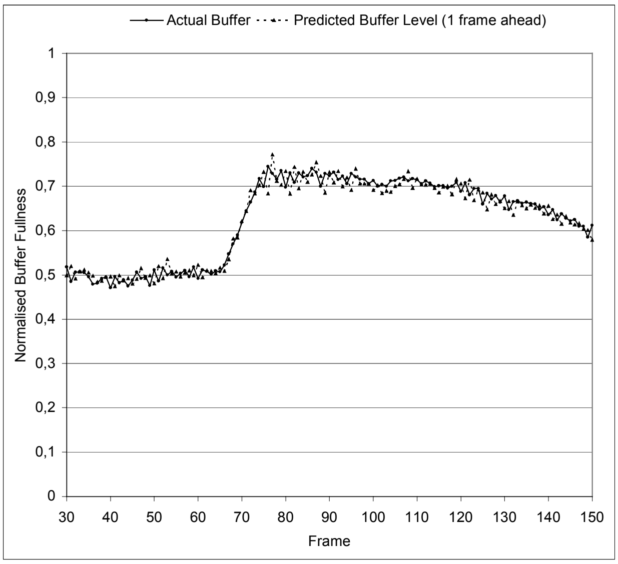

Figure 3. Comparison between the actual buffer level and predicted level 1 frame ahead.

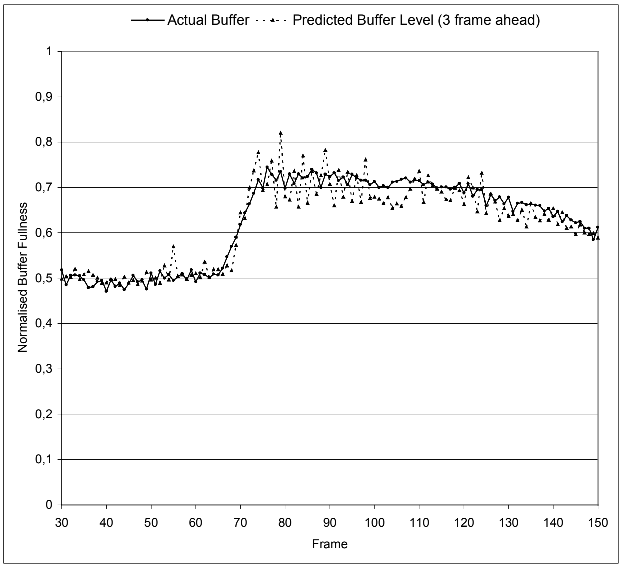

Figure 4. Comparison between the actual buffer level and predicted level 3 frames ahead.

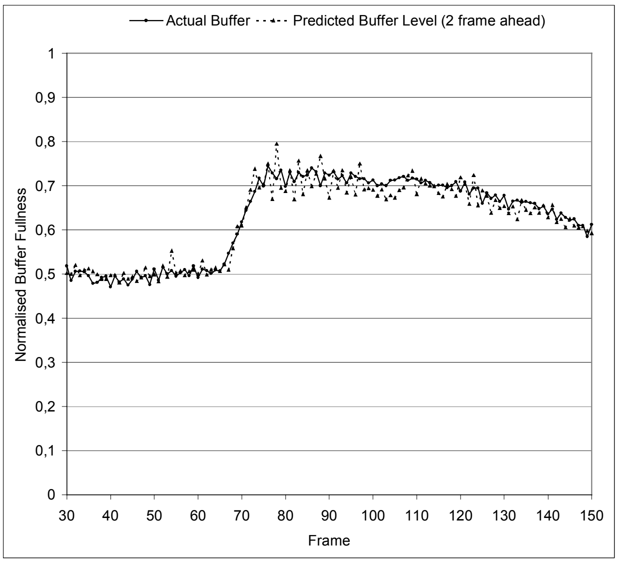

Figure 5. Comparison between the actual buffer level and predicted level 2 frames ahead.

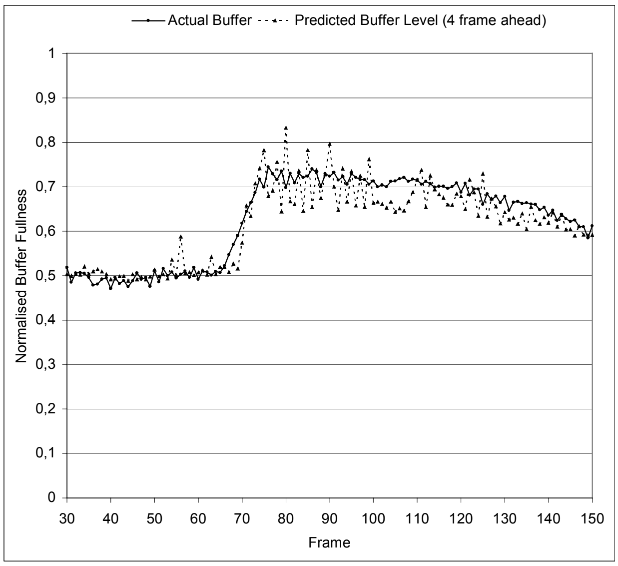

Figure 6. Comparison between the actual buffer level and predicted level 4 frames ahead.