

AN UNEQUAL ERROR PROTECTION SCHEME FOR OBJECT BASED VIDEO COMMUNICATIONS

S. Nasir, S. Worrall, M. Mrak and A.M. Kondoz

Centre for Communication Systems Research, University of Surrey,
Guildford, Surrey, GU2 7XH, U.K.
{s.nasir, s.worrall, m.mrak, a.kondoz}@surrey.ac.uk

ABSTRACT

This paper presents a novel adaptation technique for optimisation of the source and error protection rates in response to varying channel conditions in wireless networks. The optimisation methodology is based on accurate modelling of video frame distortion at the encoder. An optimal value for the quantisation parameter is calculated for video frames based on these estimates, and then video packets are mapped to one of a number of different priority radio bearers, according to their expected distortion. The algorithm performance is demonstrated for object based MPEG-4 video transmission over a UMTS/FDD system. The results demonstrate that a significant performance gain is achieved with the proposed scheme, compared with the equal error protection scheme for video transmission over a fixed bandwidth channel.

Index Terms— joint source channel rate allocation, unequal error protection, distortion modelling

1. INTRODUCTION

High quality video communication applications over wireless networks still pose a challenging problem due to the hostile and highly variable nature of the radio channels. Carefully designed error resilience techniques and channel protection can make the video data more resilient to channel degradations. However, as they are separately optimised, they are still insufficiently effective for application over mobile channels. Joint source-channel coding approaches have been proven to provide optimal performances for video application over practical systems [1]. In such an approach, either source coding is optimised for fixed channel coding or an optimal channel coding is derived for a fixed source code. However, adaptive source-channel code optimisation has been proven to provide better perceptual video quality [2].

The provision of object based video coding in video coding standards has generated a significant research interest in prioritising the transmission of segmented video objects and ultimately to achieve robust region of interest coding [3-5]. In object based video coding a video sequence is

separated into a number of objects, each of which is encoded in a separate compressed bit-stream.

Channel coding is usually performed at the physical link layer to overcome the effect of propagation errors. When the channel protection mechanisms are coupled with the video compression formatting at the application layer as in [6], this requires modification of all underlying network protocol layers in order to avoid redundancy for efficient resource utilisation. This problem can be solved by prioritisation of different parts of the video bit stream by sending them using multiple bearers with different characteristics such as different channel coding, modulation, etc. In order to ensure good performance of this method, it is necessary to separate the encoded bit-stream optimally into a number of sub-streams.

This paper addresses the prioritisation of different parts of the video bit-stream to provide maximum error robustness to the most vulnerable video packets, for transmission over wireless networks. The proposed scheme applies a packet level Unequal Error Protection (UEP) scheme to obtain optimal video quality over a wide range of channel conditions. The proposed scheme is compatible with the codec standard and is also transparent to the underlying network allowing video services over heterogeneous (fourth generation) wireless networks.

2. FORMULATION OF THE PROPOSED OPTIMISATION SCHEME

In this section, a novel UEP scheme is introduced with a goal to optimise the transmission of multiple video objects over a wireless communication system like UMTS that supports multiple radio bearers (sub-channels) with different channel coding schemes.

2.1 Transmission Channel Resource Optimisation

The underlying idea in this work is to optimise the use of available transmission channel resources for enhancing the output video quality. A transmission channel is characterised by channel bandwidth R_{ch} . The bandwidth is shared amongst available sub-channels. For sub-channels with indices $c = 0, \dots, C - 1$, where C is the number of sub-channels, the

corresponding bandwidth is $R(c)$, where $R_{ch} = \sum_{c=0}^{C-1} R(c)$, and for each sub-channel a probability of channel bit errors is known. Additionally, for each sub-channel a channel coding rate $\chi(c)$ is also defined. The combination of spreading gain, channel coding and modulation schemes provide different radio bearer configurations offering flexibility in the degree of protection, coding complexity and traffic capacity, which has been used in this work to prioritise objects of the video content on the basis of their importance.

2.2 Object Based Prioritisation

In an object-based video coding scenario, a video frame is separated into sub-streams at the bit-stream level. In our approach the number of these sub-streams is fixed over a video sequence. However, their rate and related contribution to the overall distortion varies. The sub-streams are associated with the available bearers. For given channel configuration and given content, the optimisation process determines the allocation of different sub-streams to video packets of different objects. The goal set here is to find the optimal sub-stream separation and mapping of video packets onto multiple bearers in such a way as to maximise the received video quality for video transmission over a bandwidth limited error prone channel.

Received visual quality is determined by the amount of distortion incurred during transmission. Let the input video sequence consists of M video frames. Each video frame is separated into L_m number of objects. Each object is coded separately using its own segmentation mask. Every object is divided into $N_{m,l}$ number of video packets. If expected distortion due to the corruption of the n^{th} packet of the m^{th} frame of the l^{th} video object is $\alpha(l, m, n)$, then the total expected distortion $E(D)$ of the sequence becomes

$$E(D) = \sum_{l=0}^{L-1} E(D_l) = \sum_{m=0}^{M-1} \sum_{l=0}^{L_m-1} \sum_{n=0}^{N_{m,l}-1} \alpha(l, m, n) \quad (1)$$

Thus the goal of the optimisation process is to minimise $E(D)$.

Now the optimisation problem can be formulated as optimal allocation of bits within each packet of K sub-streams, subject to the bandwidth limitations of the available communication channel. This is achieved by estimating the expected distortion for each packet of the current frame and then mapping them to the available sub-channels so that the highly protected sub-channel can carry most of the packets with the highest expected distortion.

A further level of optimisation is achieved by incrementing the value of the QP for each of the objects and recalculating the total cost in terms of expected distortion for the combination of packets of each of the sub-channels. The

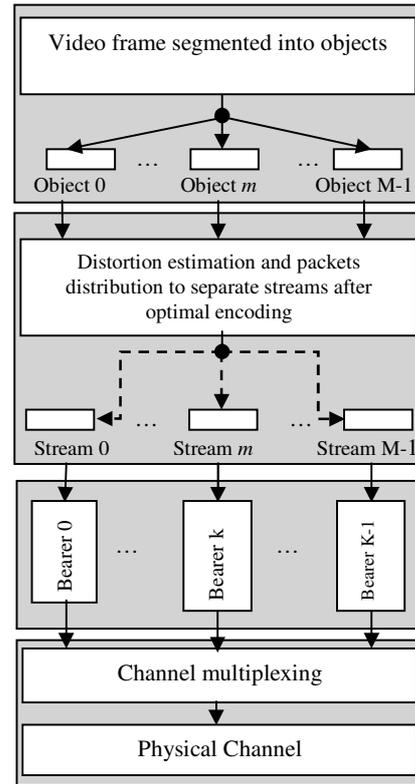


Figure 1: Object based channel resource allocation

process is repeated until the distortion cost is reduced to a minimum. The video frame is then coded with that optimal value of QP and packets are separated into sub-streams accordingly. Figure 1 outlines the proposed idea.

3. DISTORTION MODELLING

The sub-stream distribution algorithm described in the previous section relies on accurate distortion estimation at the video packet level. This section describes the estimation of the distortion due to the corruption of data in a video packet.

In wireless networks, the video quality is affected by the quantisation distortion $E(Q^D)$ and channel distortion [7]. Channel distortion is composed of concealment distortion and distortion caused from error propagation over frames coded using temporal prediction. Concealment distortion depends on the concealment techniques applied at the decoder. If errors in the video data are concealed with concealment data from the previous frames, then it is called temporal concealment and distortion caused by such a scheme is called as temporal concealment distortion $E(D^T)$. Errors can propagate in two ways, either in the temporal direction or in the spatial direction. Frame to frame error propagation through motion prediction and temporal concealment is called temporal domain error propagation f^{pp} . The distortion estimation model used in this work has

similarities with the methods proposed in [7, 8]. However, the distortion induced due to error propagation is calculated differently and adaptive intra refresh technique is used to minimise the error propagation factor, even though the same assumption, namely the uniform distribution of the video reconstruction error, is made. Previous research has suggested that spatial error concealment is less important than the temporal error concealment [9] so in this work only temporal concealment is considered.

Taking video packet as the base unit, the expected frame quality can be written as:

$$E(Q_{l,m}) = 10 \cdot \log \left(g / \sum_{n=0}^{N-1} E(D_{l,m,n}) \right) \quad (2)$$

where $E(Q_{l,m})$ is the expected quality of m^{th} frame of the l^{th} object, $E(D_{l,m,n})$ is the expected distortion of n^{th} video packet and N is the total number of video packets. g is a constant defined by the colour depth and by the size of given composite video frame, $g = cd^2 \cdot w \cdot h$, where cd is the colour depth factor, w and h are the width and height of the video frame, respectively. $E(D_{l,m,n})$ can be written as:

$$E(D_{l,m,n}) = E(D_{l,m,n}^O) + \rho^e(l, m, n) \cdot E(D_{l,m,n}^T) + f^{pp}(l, m, n) \quad (3)$$

Calculation of each term shown in (3) depends on the implemented concealment techniques, and the applied video coding scheme and its parameters.

Quantisation distortion is computed from the squared difference of original and reconstructed luminance values for every macroblock in the particular video packet.

Probability $\rho^e(l, m, n)$ of receiving a video packet with errors depends on the channel's Bit Error Rate (BER) and size of the video packet. If ρ^b is the BER of the transmission channel then $\rho^e(l, m, n)$ is given as:

$$\rho^e(l, m, n) = 1 - (1 - \rho^b)^{p_b(l, m, n)} \quad (4)$$

where $p_b(l, m, n)$ represents the size of the packet in bits.

The extent of temporal concealment distortion depends on the algorithm used for the temporal concealment. Here a simple concealment approach is followed that macroblocks in the corrupted video packet are replaced by data of the corresponding macroblocks of the previous frame. In estimation it is assumed that the neighbouring video packets and reference frames are received correctly. Therefore, the concealment distortion is given as:

$$E(D_{l,m,n}^T) = \|Y_{l,m,n} - Y'_{l,m-1,n}\| \quad (5)$$

which is the squared difference of the luminance components of macroblocks of reconstructed part of current frame $Y_{m,n}$ and the same spatial area in the previous frame $Y'_{m-1,n}$.

The temporal error propagation due to MB mismatch between adjacent video frames is quantified by the term $f^{pp}(l, m, n)$ in (3), which is computed as:

$$f^{pp}(l, m, n) = (1 - \rho^e(l, m, n)) \cdot p_{l,m-1}^{pp} \cdot \sum_{k=0}^{K-1} [\rho^e(l, m-1, n) \cdot E(D_{k,l,m-1,n}^T)] + (1 - \rho^e(m-1, n, l)) \cdot p_{l,m-2}^{pp} \quad (6)$$

where K is the total number of macroblocks in the frame. The summation in (6) represents the error propagation through MBs. $p_m^{pp} = 1 - (1 - \rho_b)^{F_m}$ quantifies the fraction of distortion of the m^{th} reference video frame, which should be considered in the propagation loss calculation and F_m is the size of the frame area contained in the observed packet.

Packets are then distributed on the basis of the estimated distortion to respective bit-streams. Packets with higher expected distortion are put on the sub-channel with higher protection.

3. EXPERIMENTAL RESULTS

The proposed scheme has been tested for performance over a simulated link level UMTS channel. A detailed explanation of the developed link level simulator can be found in [10].

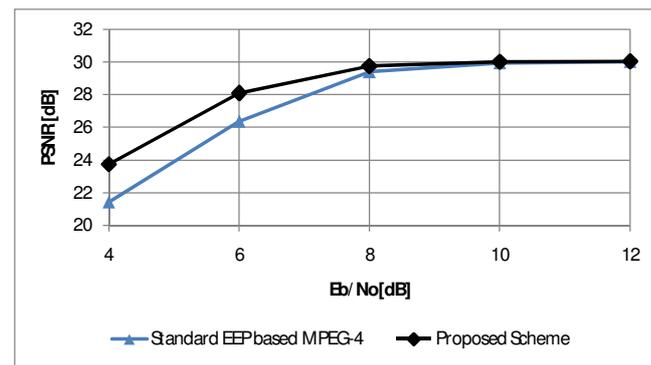
Experiments have been performed with two video sequences, 'Singer' and 'Dancer'. Each sequence is segmented into two objects, where packets of each object are distributed to either of the sub bit-streams following the proposed optimisation scheme. The 'Singer' is a QCIF resolution (176×144) sequence that has moving clouds in its background while the 'Dancer' object is used with the 'Highway' sequence in background at CIF resolution (352×288). Two channel bearers are used to transmit one sub bit-stream over each of them. A spreading factor of 16 is used in the physical channel configuration with 1/2 rate and 1/3 rate convolutional coding over the two sub-channels. These channels permit a maximum of 208 kbps and 138 kbps respectively. Therefore, 90 frames of the QCIF resolution 'Singer' sequence are encoded at 30 frames per second (fps) while 180 frames of the CIF resolution 'Dancer' sequence are encoded at 15 fps. Experiments have been conducted for a range of channel conditions for the proposed optimisation scheme and a comparison is made with standard MPEG-4 object based coding.

The video performances are shown in terms of average frame PSNR vs Eb/No in Figure 2. Each point of the graph has been taken as the average of 40 runs of the decoder for each of the 90 frames long test sequences. The figure illustrates the performance of the proposed method compared to standard MPEG-4 encoding with traditional equal error protection (EEP). The main strength of the proposed scheme is efficient packet prioritisation, which is most effective in bad channel conditions. For test points at lower Eb/No the gain is significant and reaches up to 2 dB.

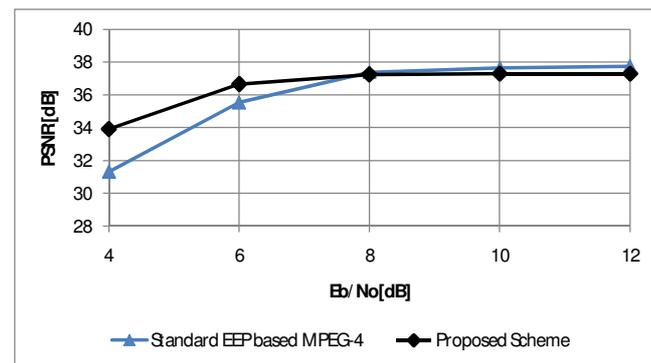
The computational overhead for the distortion calculation in the proposed scheme does not add any

significant complexity during video coding. However, it is assumed that the encoder will have some pre-computed information about the distortion cost for different QP values to minimise the computational complexity.

The scheme has been tested for a set of three QP values. 12, 16 and 20 for the background object and 12, 14, and 16 for the foreground object in all the tests. The optimal value of QP for each object is used to encode that particular frame. The QP values used for standard MPEG-4 that is used in comparison are 10 for I frames and 12 for P frames.



a)



b)

Figure 2: Performance Comparison for the test sequences a) 'Dancer' b) 'Singer'

4. CONCLUSIONS

An optimum channel resource allocation scheme is proposed for video communications over wireless networks. The rate allocation is based on accurate modelling of the expected distortion at the video packet level. A novel UEP scheme is incorporated into the rate allocation scheme to achieve optimum video quality. The video data is prioritised according to the relative perceived importance, which is estimated using the developed distortion model at the encoder. Prioritised streams are transmitted over the air interface by using multiple radio bearers with different error protection capabilities. The experiments that are carried out

over the simulated UMTS systems show that significant performance improvements can be achieved with the proposed scheme compared to traditional EEP schemes.

5. ACKNOWLEDGMENTS

The work presented was developed within VISNET II, a European Network of Excellence, funded under the European Commission IST FP6 programme.

6. REFERENCES

- [1] Bystrom M., Stockhammer T., "Dependent source and channel rate allocation for video transmission", *IEEE Trans. Wireless Comm.*, Vol.3, Iss. 1, pp. 258 – 268, Jan. 2004.
- [2] Dyck. R.E., Miller. D.J., "Transport of wireless video using separate, concatenated, and joint source-channel coding", *Proceedings of the IEEE*, Vol. 87 Iss. 10, pp. 1734 -1750, Oct. 1999.
- [3] Chen, O.T.-C., Chih-Chang Chen, "Automatically-Determined Region of Interest in JPEG 2000," *IEEE Trans. Multimedia.*, Vol. 9, Iss. 7, pp. 1333 - 1345, Nov. 2007.
- [4] Sanchez V., Basu A., Mandal M. K., "Prioritized region of interest coding in JPEG2000," *IEEE Trans. Circuits and Sys. for Video Tech.*, Vol. 14, Iss. 9, pp. 1149 - 1155, Sept. 2004.
- [5] Lee S.-H., Moon J., Lee M., "A Region of Interest Based Image Segmentation Method using a Biologically Motivated Selective Attention Model", *Int. Joint Conf. Neural Networks*, pp. 1413-1420, July 2006
- [6] Gharavi H., Alamouti S.M., "Video transmission for third generation wireless communication systems", *Journal of research of the national institute of standards and technology*, Vol. 106, pp. 455 - 469, March-April, 2001.
- [7] Katsaggelos A.K., Eisenberg Y., Zhai F., Berry R., Pappas T. N., "Advances in Efficient Resource Allocation for Packet-Based Real-Time Video Transmission", *Proceedings of the IEEE*, Vol. 93, Iss. 1, pp. 135-147, Jan 2005.
- [8] Kim I.-M., Kim H.-M., "An optimum power management scheme for wireless video service in CDMA systems", *IEEE Trans. Wireless Comm.*, Vol. 2, Iss. 1, pp. 81 – 91, Jan. 2003.
- [9] Wang Y., Wenger S., Wen J., Katsaggelos A.K., "Error resilient video coding techniques," *IEEE Signal Processing Magazine*, Vol. 17, No. 4, pp. 61 - 82, Jul 2000.
- [10] Kodikara, C., Worrall, S., Fabri, S., Kondo, A., "Performance evaluation of MPEG-4 video telephony over UMTS", 3G2003, pp. 73-77, June 2003.