MULTIPLE PRIORITY REGION OF INTEREST CODING WITH H.264

Dimitris Agrafiotis, David R. Bull, Nishan Canagarajah, Nawat Kamnoonwatana

Centre for Communications Research, University of Bristol, Woodland Road, Bristol BS8 1UB, UK
E-mail: d.agrafiotis@bristol.ac.uk

ABSTRACT

This paper describes a modified rate control algorithm for H.264 that can accommodate multiple priority levels given a region of interest (RoI). The modified method allows better control of the quality of the RoI and gradual variation of the quality in the rest of the video frame through a bit redistribution process that is based on a number of parameters, including characteristics of the RoI, user input and perceptual factors.

Index Terms— Video coding

1. INTRODUCTION

In standard hybrid video codecs the number of bits required for coding each macroblock in a video frame is controlled mainly by the quantisation step size, which in turn is determined by the quantisation parameter (QP). Rate control, employed when the coded video has to adhere to a specific target bit rate or delay constraints, aims at producing a coded stream that respects the rate and delay constraints mainly by manipulating the frame or macroblock QP. Normally a virtual buffer is employed for this purpose which is filled with the output bits and drained at the target rate.

With standard rate control methods all macroblocks / regions are treated as if they have the same importance (priority). Region of Interest (RoI) coding aims at providing preferential treatment for those regions of the video frame which are deemed to be of higher importance to the viewer. Such regions are usually identified by a pre-processing module or by the viewers themselves and can be for example a person's face in a videoconferencing scenario or an area of activity in a surveillance application.

The method described in [1] assigns a specific number of bits to an RoI and the rest of the video frame (considered background) out of the available bit budget based on parameters such as relative size, motion and desired priority. It then calculates appropriate QP values using the RM8 (H.261) rate control method. The methods described in [2][3] manipulate the rate distortion models used in the QP calculation process by employing distortion weights with a value greater than one for macroblocks belonging to the RoI, which effectively forces the TMN8 [4] rate control algorithm that is used in the above references to choose a lower QP for those MBs that belong to the RoI. Alternatively the rate prediction mechanism can also be biased, by weighting factors applied to the residual metric employed (for example the variance) [5] which can then fool the rate control algorithm (again TMN8 is used in [5]) with regard to the complexity of the RoI and background and thus influence the QP allocation process.

The rate control method adopted by the JM encoder [6][7] differs from previous approaches in that the QP values are chosen prior to the prediction taking place. More specifically, the existence in the H.264 standard of a number of coding modes for each MB - multiple inter and intra modes - and the use of rate distortion optimization in the JM encoder for selecting one makes the application of typical rate control strategies quite problematic. Most of the previously mentioned rate control methods rely on a rate model and a distortion model for choosing an optimal quantiser for each macroblock or frame, given a measure of the variance of the residual signal (the prediction difference signal) and a specific bit budget allocation. The rate model is used to predict the number of bits output after coding a macroblock or frame with a specific quantiser and the distortion model is used to predict the distortion associated with each quantiser. Lagrangian optimisation can then be employed to choose the best QP in a rate distortion sense. The problem with the JM H.264 encoder lies with the fact that the residual signal depends on the choice of coding mode and the choice of coding mode depends on the choice of QP which in turn depends on the residual signal (a chicken and egg type of problem).

The adopted solution [6][7] in the JM encoder is one where the choice of QP is made prior to the coding mode decision using a linear model for predicting the activity of the residual signal of the current basic unit (e.g. frame, slice, macroblock) based on the activity of the residual signal of past (co-located) basic units. Once the residual signal activity is predicted, the same rate model used in VM8 [8] is employed to find a QP which will lead to a bit stream that adheres to the specific bit budget allocation and the buffer restrictions. The size of the basic unit defines the number of rate control layers. Three layers are employed for a basic unit smaller than the frame size (in MBs), two otherwise. These are the GOP layer, the frame layer and the basic unit layer. We concentrate on a 2 layer rate control. Each of the 2 layers needs to be modified to accommodate RoI coding.

In this paper we suggest and describe such modifications for the case of one RoI and multiple priority levels. The structure of this paper is as follows. In section 2 we describe how the priority levels are created. In section 3 we present our RoI coding method. In section 4 we give indicative results taken with our method and finally in section 5 we conclude this work.

2. PRIORITY LEVELS

Given the RoI size, a fixed number of priority levels are created around the RoI. The number of priority regions is determined by the number of desired levels and the maximum number of levels that can be accommodated if the RoI was positioned at the centre of the frame. Each region, including the RoI, is specified at the macroblock level.

If the RoI is specified as a circular region, the part of the radius from the centre of the frame to one of the frame corners that is greater than the RoI radius is equally split among the priority levels (i.e. if the RoI was centred on the video frame, the part of the radius lying outside the RoI). If the RoI is rectangular the split is done in a similar fashion but using the vertical and horizontal distances from the corresponding edge of the RoI instead, again done as if the RoI was centred on the frame. Alternatively foveation methods as those described in [9] can be used for defining priority regions based on eccentricity values[1], assuming that the fixation point lies on the centre of the RoI.

The calculation of both the extent and the number of the priority regions is done on a frame by frame basis with the maximum number of levels being fixed for the lifetime of a RoI (i.e. from the frame in which the RoI first appears until the frame it disappears). The priority levels are anchored at the centred of the RoI surrounding it on every frame as shown in the example of figure 1 for all three types of level definition.

3. ROI RATE CONTROL

3.1 GOP layer

In the first rate control layer a specific bit budget is allocated to the remaining pictures within a GOP, based on the coding rate and the occupancy of the virtual buffer employed for the rate regulation. The occupancy of the buffer is updated after coding each picture. The GOP layer also assigns a QP - $QP_i(1)$ - to the intra (IDR) and first predicted picture (P) of the i^{th} GOP based on the average QP assigned to the P pictures of the previous GOP as well as the respective $QP_{i-1}(1)$ allocation made at the beginning of the previous GOP (eq.1). In our modified rate control this QP assignment is done separately for each priority region, i.e. a $QP_{i,j}(1)$ is assigned to the j^{th} priority region based on the average QP assigned to this region in the P pictures of the previous GOP and the value of $QP_{i-1,j}(1)$ (eq. 2).

[1] The eccentricity value of a macroblock depends on its distance from the fixation point and the distance of the viewer from the screen.

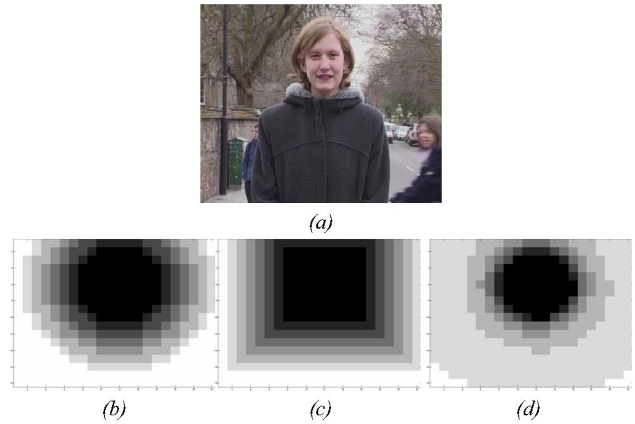

Figure 1 :Eight priority levels for an RoI centred on the face of the woman shown in (a) with a circular RoI definition (b), a rectangular (c) and based on foveation (d). In all cases a radius of 60 pixels was used (rectangle height = width = 120 pixels)

$$QP_i(1) = \max\left\{QP_{i-1}(1) - 2, \min\left\{QP_{i-1}(1) + 2, \frac{SumPQP_{i-1}}{NP_{i-1}} - \min\left\{2, \frac{N_{i-1}}{15}\right\}\right\}\right\} \quad (1)$$

$$QP_{i,j}(1) = \max\left\{QP_{i-1,j}(1) - 2, \min\left\{QP_{i-1,j}(1) + 2, \frac{SumPQP_{i-1,j}}{NP_{i-1}} - \min\left\{2, \frac{N_{i-1}}{15}\right\}\right\}\right\} \quad (2)$$

SumPQP is the sum of QP values assigned to the P pictures (eq.1) or a priority region of the P pictures (eq.2) in the previous GOP. *NP* is the number of P frames in the GOP and *N* is the total number of frames in the GOP. The restriction set in eq.2 for the $QP_{i,j}$ value to differ by a maximum of 2 from $QP_{i-1,j}$ is lifted for the second GOP in the sequence so that the QP values of the priority regions converge to the desired levels faster.

3.2 Picture layer

At the 2nd layer the frame bit budget allocation for P frames takes place. There are two stages in the frame bit allocation process. The first stage takes place after coding the first two pictures (IDR,P) of the current GOP and determines a target buffer level for each remaining P picture in the GOP based on the bit usage of these first two frames. The second stage of the frame bit allocation process takes place at each remaining P frame and determines the bits that will be allocated for the current P picture in the current GOP based on the target buffer level, the frame rate, the available channel bandwidth and the actual buffer occupancy. The actual allocated (target) bits are a weighted combination of the outputs of these two stages. Once a specific number of bits has been allocated to the current frame the algorithm normally proceeds to compute the frame QP. This involves the use of the linear model in equation 3 which predicts the MAD (σ) of the current frame based on the actual MAD of the previous P picture as follows:

$$\tilde{\sigma}_i(k) = a_1 \times \sigma_i(k-1) + a_2 \quad (3)$$

where a_1 and a_2 are the two coefficients of the model updated after coding each frame. The quantization step corresponding to the target bits is then computed using the following quadratic model:

$$T_i(k) = c_1 \times \frac{\tilde{\sigma}_i(k)}{Q_{step,i}(k)} + c_2 \times \frac{\tilde{\sigma}_i(k)}{Q^2_{step,i}(k)} + m_{h,i}(k) \qquad (4)$$

where $T_i(k)$ is the target bits for frame k of the i^{th} GOP, $m_{h,i}(k)$ is the total number of header bits and motion vector bits and c_1 and c_2 are two coefficients of the quadratic model again updated after coding each frame. The computed $QP_i(k)$ is restricted to lie within a range of [-2, +2] of the previous frame QP.

For variable priority coding we distribute the texture bits allocated to the current frame among the different regions in a way that reflects their relative importance. More specifically the amount of bits allocated to each region is calculated based on the size of the region, the predicted MAD value of the region and its priority. At a first stage the frame texture bits are distributed to each region based on the normalized size and normalized predicted MAD of each region as follows:

$$T_{i,j}(k) = \left(w_s S_{i,j}(k) + w_{MAD} MAD_{i,j}(k)\right) x \left(T_i(k) - m_{h,i}(k)\right) \qquad (5)$$

where $S_{i,j}(k)$ is the normalized size of the j^{th} region in frame k of the i^{th} GOP, $MAD_{i,j}(k)$ is the normalized predicted MAD of the region and w_s and w_{MAD} are the weights of the two factors indicating how much they influence the bit allocation process, with $w_s + w_{MAD} = 1$. The MAD prediction in our rate control is done independently for each region with the coefficients a_1 and a_2 of the linear model in eq.5 having separate values for each region and the MAD value of the j^{th} priority region in the previous frame being used instead of that of the picture as shown in equation (6).

$$\tilde{\sigma}_{i,j}(k) = a_{1,j} \times \sigma_{i,j}(k-1) + a_{2,j} \qquad (6)$$

Once the initial allocation has taken place, we employ a priority constant P_0 (ranging from 0 to 1) to specify the percentage of texture bits that will be redistributed from the lower priority regions to the highest priority region 0. The priority of the rest of the $n-1$ regions (i.e. all apart from the lowest priority region) is then calculated using the following exponential model:

$$P_j = e^{-\frac{j}{3}} \times P_0, \forall\ j = 1,..,n-1 \qquad (7)$$

Each j^{th} region is assigned a percentage P_j of the remaining texture bits of all lower priority regions. The number of texture bits remaining in each of the lower priority regions is updated before each redistribution step. The target bits for each region is thus given by the following equation:

$$T_{i,j}(k) = P_j \times \left(\sum_{m=(j+1):n} T_{i,j+m}(k) \right) \qquad (8)$$

The QP of each region is then calculated using the region's target bits and the quadratic model of eq.4 albeit with region specific coefficients $c1$ and $c2$ and predicted MAD values as shown below. The QP values for all regions are restricted to lie within a range of (-2, +2) of the QP of the same priority region in the previous frame.

$$T_{i,j}(k) = c_{1,j} \times \frac{\tilde{\sigma}_{i,j}(k)}{Q_{step,i,j}(k)} + c_{2,j} \times \frac{\tilde{\sigma}_{i,j}(k)}{Q^2_{step,i,j}(k)} \qquad (9)$$

4. RESULTS

We show representative results collected for an outdoor sequence showing a woman signing in front of the camera. Eight priority levels were defined using eccentricity values, assuming that the point of fixation lies on the face of the signer. This assumption has been shown to be valid for deaf viewers in [10]. The multiple priority levels can accommodate the viewing model followed by the human visual system whereby regions away from the fixation point are processed with lower acuity. Figure 2 shows average PSNR results for each region with standard rate control and our method (VPRC – variable priority rate control - in the figure) at different priorities for all tested bit rates (96 to 256 kbits/sec). It can be seen that the perceptually important regions (regions close the fixation point - 0 to 4) are favoured by VPRC at the expense of regions 5 to 7 which are less significant for sign language comprehension [10] as they are located further away from the face. On a frame by frame basis Figure 3 shows the quality improvements offered for the highest priority region by our method (VPRC at 25%), for the duration of two GOPs at a rate of 128 kbits/sec. The effectiveness of our method is particularly visible at the high activity frames shown in Figure 3 where the quality of the perceptually important regions is preserved in contrast to the standard approach that experiences dips in performance and can thus hamper comprehension. For the specific priority constant (25%) the minimum region 0 PSNR recorded with VPRC at 128 kbits/sec was 26.04 dB whereas the minimum region 0 PSNR recorded with standard RC was 23.2 dB. Similarly, the respective maximum region 0 PSNR values were 35.63 dB for VPRC and 33.87 dB for the standard RC. Similar benefits can be had with other types of sequences, e.g. surveillance sequences, when bit rates are limited.

5. CONCLUSIONS

We have presented a modified rate control method that allows the definition of multiple priority regions, the quality of which varies, based on the RoI characteristics, input by the user and the target bit rate. The results presented demonstrate the coding flexibility offered by our method which can lead to significant quality improvements for the RoI and a perceptually pleasing variation in quality for the rest of the frame, without violating the target bit rate.

6. REFERENCES

[1] D.Chai, K. N. Ngan, “Content based bit allocation and rate control for classical MC-DCT video coding systems”, ISCAS 1998

[2] S. Daly, K. Matthews, J. Ribas-Corbera, “Face-Based Visually-Optimized Image Sequence Coding”, ICIP 1998

[3] Chia-Wen Lin, Yao-Yen Chang and Yung-Chang Chen, “Low-Complexity Face-Assisted Video Coding”, ICIP 2000

[4] J. Ribas-Corbera and S. Lei, “Rate Control in DCT Video Coding for Low-Delay Communications”, IEEE Trans. on Circ. and Syst. for Video Tech., Vol.9, No.1, Feb. 1999

[5] Mei-Juan Chen, Ming-Chieh Chi, Ching-Ting Hsu and Jeng-Wei Chen, “ROI Video Coding Based on H.263+ with Robust Skin-Color Detection Technique”, IEEE Transactions on Consumer Electronics, Vol. 49, No. 3, August 2003

[6] G. Sullivan, T. Wiegand, Keng-Pang Lim, "Joint Model Reference Encoding Methods and Decoding Concealment Methods”, JVT-I049, September 2003.

[7] Zhengguo Li, Feng Pan, Keng Pang Lim, Xiao Lin, Susanto Rahardja, “Adaptive Rate Control for H.264”, ICIP 2004.

[8] T. Chiang and Y.-Q. Zhang, “A new rate control scheme using quadratic rate-distortion modelling”, IEEE Trans. on Circ. and Syst. for Video Tech., Feb. 1997.

[9] S. Lee and A. C. Bovik, “Fast algorithms for foveated video processing”, IEEE Trans. on Circ. and Syst. for Video Tech., Vol. 13 , No. 2, Feb. 2003.

[10] D. Agrafiotis, N. Canagarajah, D.R. Bull, J. Kyle, H. Seers, M. Dye, “A video coding system for sign language communication at low bit rates”, ICIP 2004.

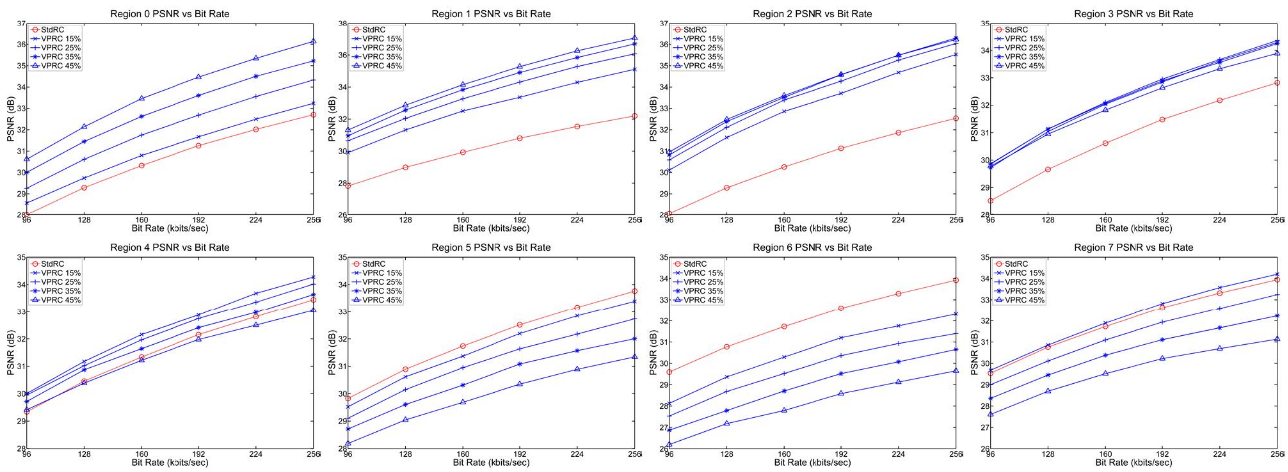

Figure 2 : “Outdoor2” average PSNR graphs for each priority region with standard RC and VPRC.

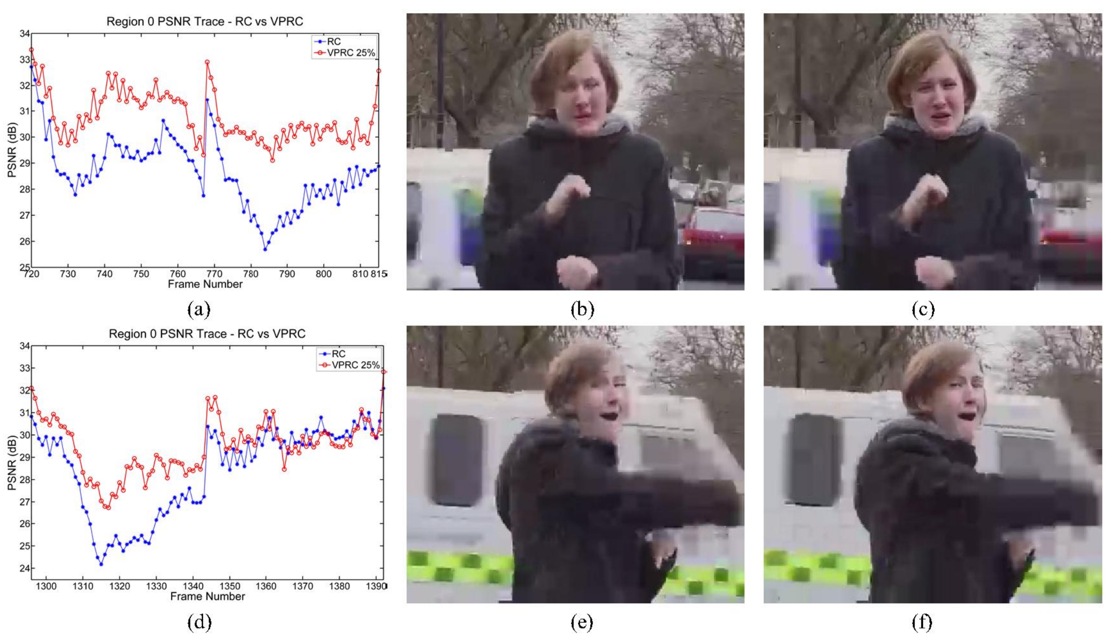

Figure 3 : Frame results for the “outdoor 2” sequence at 128 kbits/sec; a) PSNR trace for frames 720 to 815 (2 GOPs), b) frame 784 coded with standard RC, c) same frame with VPRC at 25% priority, d) PSNR trace for frames 1296 to 1392 (2 GOPs), e) frame 1313 coded with standard RC, f) same frame with VPRC at 25%