

OPTIMIZING ALGORITHMS FOR REGION-OF-INTEREST VIDEO COMPRESSION, WITH APPLICATION TO MOBILE TELEHEALTH

Sira Rao and Nikil Jayant

Department of Electrical and Computer Engineering
Georgia Institute of Technology, Atlanta, GA, USA
{parshu, jayant}@ece.gatech.edu

ABSTRACT

Wireless communication of video poses constraints on information capacity. Region-of-Interest (ROI) video coding provides higher quality in the ROI, but poorer quality in the background (BKGRND), for a given total bitrate (TBR). Researchers, including the authors, have also proposed more graceful quality management methods, using what is referred to here as an Extended-Region-of-Interest (EROI). We consider three levels of losslessness – mathematical, diagnostic, and perceptual, with the goal of associating them with the above-mentioned regions. We describe work in progress aimed at optimizing an elastic expert system based on the above methodology, with telehealth video as its anchor. The optimizations are threefold – user, perceptual, and network oriented, and are incorporated in the rate control algorithm. We propose a rate control method where, unlike conventional methods, bit allocation is shifted from the frame level to individual regions within the frame. Thereafter, the above-mentioned criteria are used to determine regional bit allocation. Peak-Signal-to-Noise-Ratio (PSNR) results show, as expected, that the proposed scheme achieves higher ROI-EROI quality than the verification model VM8 of MPEG4. This is illustrated with four examples of pediatrics video. The value and design of the proposed methodology is being corroborated by subjective testing involving medical experts. We are independently researching another standing issue in the telehealth application, that of low complexity segmentation and tracking of the ROI-EROI boundaries.

1. INTRODUCTION

Recently, researchers have considered an elastic approach to wireless video communications. This has facilitated mobile telehealth, which aids in treatment of medical emergencies and will likely provide pervasive healthcare services in the future. ROI coding, where regions of greater diagnostic importance to the medical expert are coded with greater quality than the BKGRND, is a useful and popular elastic tool. In this paper, we also apply the notion of an EROI [1,2] – an intermediate region between the ROI and BKGRND, that enhances the elasticity of an ROI based video coding scheme by allowing for a more perceptually pleasing degradation in quality from the ROI. When the total bitrate (TBR)

permits, the ROI can be coded no loss in quality compared to the original, called mathematical losslessness (M-losslessness). In general, actual qualities of each region depend on the TBR and the rate control algorithm does the assignment of the governing quantization parameters. The following are some of the key works in the literature on spatial quality control and rate control methods.

In [1], the algorithm achieves a target ROI quality while allowing perceptually tolerable exponential degradation in quality for the background. In [3], a flexible bit allocation algorithm is proposed where the user can adjust bit consumption using a scale that ranges from “null-ROI” to “coding ROI only”. However, within each region, the quantization parameter for all the macroblocks (MB) is the same. In [4], a rate control technique for selecting quantization parameters in typical discrete cosine transform coders is presented that is based on distortion weights. This method has been adopted as a rate control tool for the TMN8 version of H.263+ and VM8 version of MPEG4.

In perceptually optimized rate control schemes, human visual system (HVS) factors are incorporated to determine some parameter in the algorithm, and the results are compared with those obtained without HVS factors. In [5], the rate control method in [4] is used and the distortion weights are determined by local visual perceptual cues. In [6], the edge masking property of the HVS is used to control the coding mode via the distortion weight. In [7], a visual distortion sensitivity index is obtained based on the HVS property that distortions in smoothly textured objects with regular motions are more noticeable than distortions in moving areas with random textures. In perceptually lossless (P-lossless) image coding, HVS factors are incorporated into a distortion measure to determine the optimum value of a parameter that satisfies a given threshold. When this threshold equals the Just-not-noticeable-difference (JNND) [8, 9], the result is P-lossless when compared with the original. To the authors’ knowledge, there are no definitive solutions for P-lossless video coding in the literature, although most commercial video coders incorporate different levels of empirical perceptual tuning.

Given that we are dealing with a medical application, the diagnostic value of the coded video is of prime importance. If an expert feels confident making an assessment with the coded video,

Invited paper: Special Session on Perceptual Visual Processing, IEEE ICME 2006.

the coding is said to diagnostically lossless (D-lossless). Thus, in this context, it is necessary to explore the relationships between perceptual and diagnostic criteria. With P-lossless image coding, if the correct values of the thresholds are not chosen, the system is either under-designed or over-designed with respect to D-losslessness. Furthermore, the computational complexity of a P-lossless video coding scheme is likely to be prohibitively expensive especially for wireless applications. Thus, the proposed system design incorporates perceptually optimized criteria with the goal of achieving D-losslessness.

In this paper, a rate control method is presented that improves upon conventional bit allocation methods by shifting bit allocation from the frame level to individual regions within the frame. On variable bitrate (VBR) wireless channels, the distortion weights [3,4,5] provide the same functionality at all bitrates. That is, all the regions are affected equally when the TBR changes. With the proposed scheme, it is possible to achieve different quality variation for each region, which is essential for mobile medical video. Our algorithm uses user (medical expert) defined criteria, perceptual criteria, and network criteria to perform regional bit allocation. Users are allowed to semantically select ROI and EROI, and HVS and network criteria may be incorporated to maximize diagnostic quality of the ROI.

The rest of the paper is organized as follows. Section 2 describes the proposed rate control method. Section 3 describes the performance results of the proposed methods. Section 4 concludes the paper and describes future work.

2. PROPOSED RATE CONTROL METHOD

Figure 1 shows the flowchart of the proposed method. By distributing the target bits for the frame between each region – ROI, EROI, and BKGRND, bit allocation is shifted from the frame level to individual regions. Then, Lagrangian optimization leads to the following expression for the ROI quantization parameter (Q):

$$Q_{i,ROI}^f = \sqrt{\frac{AK\sigma_{i,ROI}^f \sum_{i=1}^N \sigma_{i,ROI}^f \omega_{i,ROI}^f}{(B_{T,ROI}^f - AN_{ROI}C)\omega_{i,ROI}^f}}$$

The subscript i represents the macroblock (MB) number, the superscript f represents the frame number, $B_{T,ROI}^f$ is the target number of bits for the ROI, $\sigma_{T,ROI}^f$ is the variance of the i^{th} MB of the ROI, $\omega_{T,ROI}^f$ is the distortion weight for the ROI MB, A is the number of pixels in a MB, N_{ROI} is the number of MBs in the ROI, C is the overhead rate, and K is a constant. Note that this is only dependent on the bits allocated to the ROI ($B_{T,ROI}^f$), and the distortion weights for the ROI. The same is true for the EROI and BKGRND. $B_{T,ROI}^f$ provides an additional degree of freedom over distortion weights to control ROI quality.

Subject to the constraint,

$$B_{T,ROI}^f + B_{T,EROI}^f + B_{T,BKGRND}^f = B_T^f$$

selecting these parameters for the three regions is a major step towards achieving the envisioned elastic system. As detailed in [8], perceptual factors may be used to determine the distortion weights

in the Q equation. These perceptual weights are incorporated here, together with size and motion features, to achieve regional bit allocation as:

$$\begin{aligned} B_{T,ROI}^f &= (\alpha_S S_{ROI} + \alpha_M M_{ROI} + \alpha_W W_{ROI}) B_T^f \\ B_{T,EROI}^f &= (\alpha_S S_{EROI} + \alpha_M M_{EROI} + \alpha_W W_{EROI}) B_T^f \\ B_{T,BKGRND}^f &= (\alpha_S S_{BKGRND} + \alpha_M M_{BKGRND} + \alpha_W W_{BKGRND}) B_T^f \end{aligned}$$

where,

$$\begin{aligned} W_x &= \frac{\sum_i (\omega_{i,x} / S_x)}{\sum_y \sum_j (\omega_{j,y} / S_y)} \\ S_{ROI} &= \frac{N_{ROI}}{N}, S_{EROI} = \frac{N_{EROI}}{N}, S_{BKGRND} = \frac{N_{BKGRND}}{N} \\ M_{ROI} &= \frac{\sum_{ROI} |MV|}{\sum_{FRAME} |MV|}, M_{EROI} = \frac{\sum_{EROI} |MV|}{\sum_{FRAME} |MV|}, M_{BKGRND} = \frac{\sum_{BKGRND} |MV|}{\sum_{FRAME} |MV|} \end{aligned}$$

where the numerator in the equation for perceptual parameters W_x is the normalized perceptual weight of region x . The S parameters are the normalized size parameters, and the M parameters are the normalized activity descriptors - obtained from the motion vectors (MV) of each region. The α parameters are the weights for the size, motion, and perceptual parameters, with $\alpha_S + \alpha_M + \alpha_W = 1$. Thus, for a region, the larger its perceptual weight, size, or greater its motion, the higher the number of bits allocated.

2.1 Design value

At very low TBR, ROI quality will not be of a D-lossless level, and at very high TBR, the entire frame quality can be of M-lossless level. In between these TBR, the design is expected to have high value because of elastic associations between different levels of losslessness and different regions in the scene. For example, in narrowband environments, the design will achieve D-lossless ROI quality, whereas in broadband environments, the design can achieve M-lossless ROI, D-lossless EROI, and P-lossless BKGRND qualities.

3. PERFORMANCE RESULTS

The proposed rate control method was implemented within the MoMuSys version of the MPEG4 standard, and the original version was used for comparison purposes. Four pediatrics videos, named A, B, C, and D, were encoded with the above two methods. They have CIF (352×288) resolution and the frame rate is 30fps. To give high priority to perceptual parameters, $\alpha_W = 0.6$, and then $\alpha_S = 0.2$ was used. Perceptual criteria are currently being incorporated into the system design, therefore in this work results are demonstrated with representative values of the distortion weights – 4, 2, and 1, for the ROI, EROI, and BKGRND respectively. Note that these weights are used in the proposed method in the regional bit allocation and quantization parameter determination steps. In the original version, they are used only for quantization parameter determination. Also, PSNR is used as a

quality measure in this paper, and subjective testing with MCG medical experts is currently corroborating them with D-lossless notions.

Figure 2(a) visually illustrates how the EROI enhances the elasticity of the ROI. It shows one frame when video D is encoded as proposed at 208kbps. The quality transition is smooth – the PSNRs of the ROI, EROI, and BKGRND are 34.55dB, 31.11 dB, and 28.52dB respectively. Figure 2(b) shows the same frame when video D is encoded with a null EROI. The quality transition is very abrupt – from 34.55dB to 28.91dB.

Figure 3(a)-(i) illustrates the efficacy of the proposed method in two ways – visually, and by plotting the ROI, EROI, and BKGRND PSNRs versus TBR for both methods. This is done for videos A, B, and C. In each case, the ROI and EROI qualities are superior with the proposed method. There is a small degradation in the BKGRND quality that is not perceptible. The quantitative results of the experiment at 208kbps whose frames are displayed are tabulated in Table 1. The average gain in ROI PSNR is 2.50dB, in EROI PSNR is 1.19dB, and in BKGRND PSNR is -1.24dB. For the four videos, over a TBR ranging from 200kbps to 1600kbps, the ROI PSNR with the proposed method is, on an average, 3.47dB greater than that achieved with conventional rate control. The EROI PSNR is, on an average, 1.94dB greater. This is all achieved at a small cost – on an average, a 1.67dB reduction in the BKGRND PSNR.

4. CONCLUSIONS AND FUTURE WORK

This paper presents optimizing algorithms for ROI-EROI coding, with applications to mobile telehealth. These are user, perceptual, and network optimizations. A single process guided by the distortion weights achieves unequal bit allocation between regions of interest in many conventional rate control algorithms. This paper focuses on elastic methods of spatial quality distribution – through the notion of an EROI, and bit allocation – by shifting bit allocation from the frame level to individual regions within the frame, and introducing the above mentioned elastic criteria for regional bit allocation. This design is expected to result in elastic and desirable associations between different lossless notions and different regions in the scene. The results indicate that the proposed method performs better than VM8 of MPEG4 with the pediatrics database.

Future work will address two standing issues. It will aim at corroborating PSNR results with D-losslessness, thereby throwing light on the choices of optimum design parameters. Furthermore, as part of an overall system solution for mobile telehealth, this work will draw from existing techniques for ROI-EROI segmentation and tracking [4], but will again optimize them in automatic and semi-automatic forms for the mobile telehealth application.

5. REFERENCES

- [1] S. Sengupta et al, "Perceptually motivated bit-allocation for H.264 encoded video sequences," *Proc. 2003 International Conference on Image Processing*, Vol. 3, pp. 797-800, Sept. 2003.
- [2] S. Rao and N. Jayant, "A flexible approach to spatial quality distribution in wireless medical video", *Third IASTED International Conference on Circuits, Signals, and Systems*, pp. 194-199, Oct. 2005.
- [3] D. Chai et al, "Foreground/background bit allocation for region-of-interest coding," *Proc. 2000 International Conference on Image Processing*, Vol. 2, pp. 923-926, Sept. 2000.
- [4] J. Ribas-Corbera and S. Lei, "Rate control in DCT video coding for low-delay communications," *IEEE Transactions on Circuits and Systems for Video Technology*, Vol. 9, No. 1, pp. 172-185, Feb. 1999.
- [5] X. Yang et al, "Rate control for videophone using local perceptual cues," *IEEE Transactions on Circuits and Systems for Video Technology*, Vol. 15, No. 4, pp. 496-507, Apr. 2005.
- [6] H. Yu et al, "Applications and improvement of H.264 in medical video compression," *IEEE Transactions on Circuits and Systems – I: Regular Papers*, Vol. 52, No. 12, pp. 2707-2716, Dec. 2005.
- [7] C.W. Tang et al, "Visual sensitivity guided bit allocation for video coding," *IEEE Transactions on Multimedia*, Vol. 8, No. 1, pp. 11-18, Feb. 2006.
- [8] D. Wu et al, "A new visual pruning algorithm for perceptually lossless medical image coding," *Second International Conference on Video Processing and Quality Metrics for Consumer Electronics*, Jan. 2006.
- [9] D. Wu et al, "Perceptually lossless medical image coding," *IEEE Transactions on Medical Imaging*, Vol. 25, No. 3, pp. 1-10, Mar. 2006.

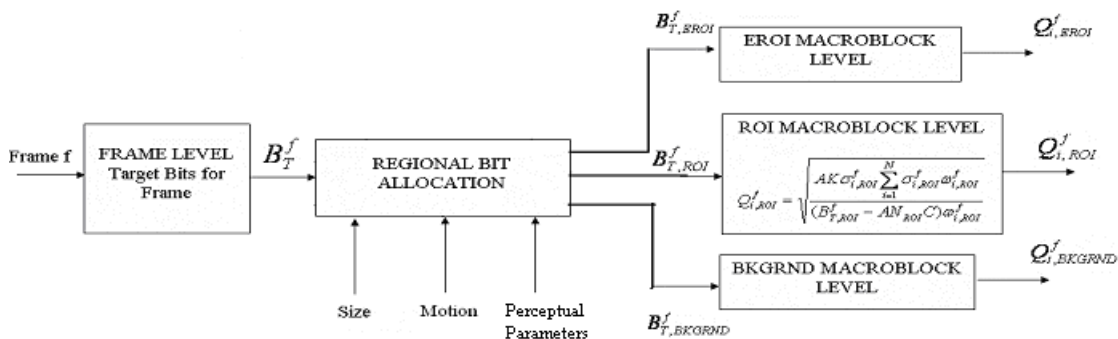
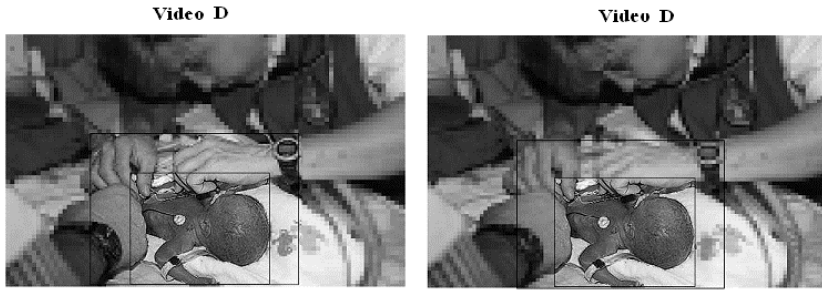


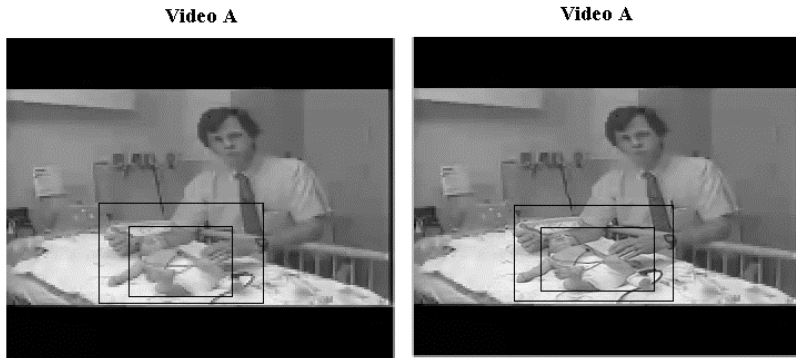
Figure 1. Proposed rate control method for ROI-EROI coding.



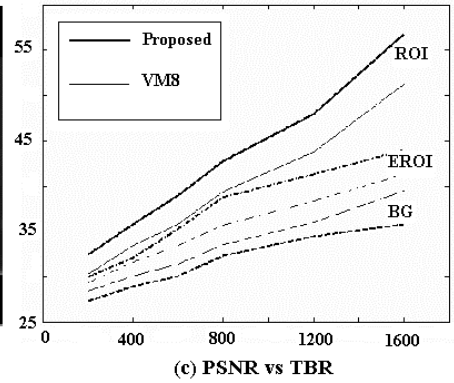
(a) With EROI (b) Without EROI
Figure 2. Video D coded at 208kbps (a) with EROI (b) with null-EROI.

Table 1. Comparison of Proposed method with VMS for Videos A, B, C, and D

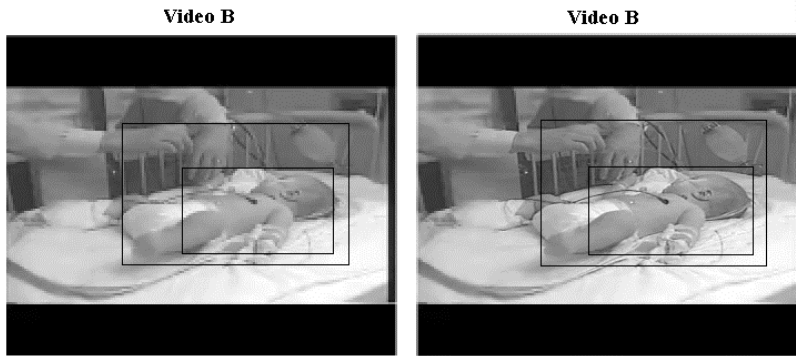
Seq.	TBR kbps	PSNR(dB) Proposed			PSNR(dB) VMS		
		ROI	EROI	BG	ROI	EROI	BG
A	208	32.43	29.97	27.34	30.23	29.37	28.38
B	208	33.38	31.20	27.94	30.72	29.96	29.14
C	208	32.54	30.11	26.80	29.81	28.24	28.14
D	208	34.55	31.11	28.52	32.13	29.73	29.96



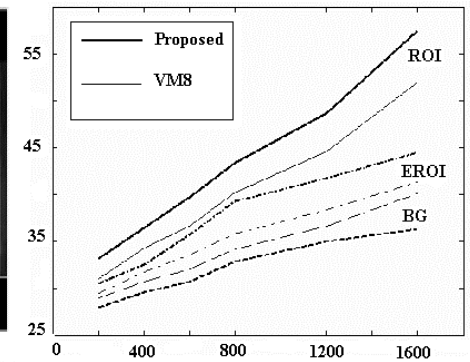
(a) VMS (b) Proposed
 (d) VMS (e) Proposed



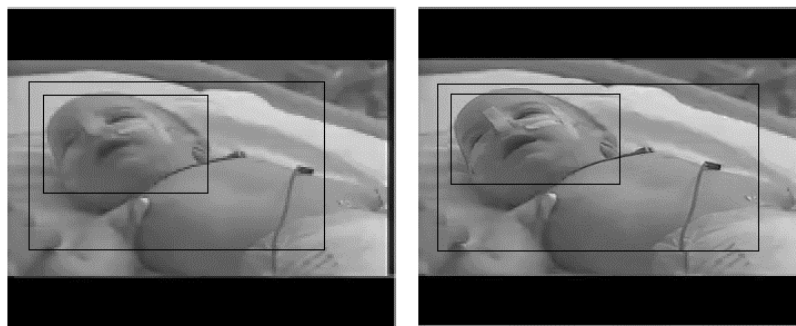
(c) PSNR vs TBR



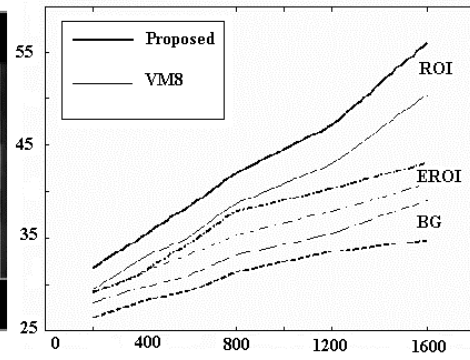
(d) VMS (e) Proposed
 (g) VMS (h) Proposed



(f) PSNR vs TBR



(g) VMS (h) Proposed



(i) PSNR vs TBR

Figure 3. Comparison of proposed rate control method with VMS, (a)-(c) Video A, (d)-(f) Video B, (g)-(i) Video C.