Bit Rate Control Schemes for ROI Video Coding

P Subramanian, *Member, IACSIT*, N R Alamelu, and M Aramudhan

Abstract—**Bit Rate control plays an important role in video coding. Region of Interest (ROI) based rate control has been attracting great attention due to the rapid demands in the region of interest in video coding. The main issue in video coding is the trade off between compression ratio and quality of the reconstructed signal. It is obvious that better quality can be achieved with smaller compression ratio and higher encoded stream bit rate. An optimal coder requires knowledge of the rate distortion (RD) model for the coding scheme. The R-D model is generally built in such a way that the quality of whole frames in a video sequence is taken into account. However in many applications like video monitoring and surveillance, telemedicine, videophone and videoconferencing, some areas in the consecutive frames of the video sequence are more important than others. It is desirable to encode those areas, called region of interest (ROI) with smaller distortion than the rest of the sequence (background). This paper presents a review of the available schemes for bit rate control in region of interest based video coding.**

Index Terms—**Background skipping, bit rate control, content based bit rate allocation, macroblock layer control, ROI video coding, SSIM QP.**

I. Introduction

The region of interest (ROI) video coding is an efficient scheme to enhance the quality of relatively important areas in a video frame, such applications include video telephony and video conferencing. In such video applications, there exist one or more regions in one video frame have higher importance than rest of the frame. In the literature, ROI coding has been discussed as a tool of improving the perceptual quality of more important areas in a video sequence and as an error resilience tool. Former is widely used in low bit rate conversational video communication applications since it provides a way of utilizing the available bit rate to maximize the effectiveness of the visual communication. This bit rate is often insufficient to retain more important visual clues in conversational applications such as facial expressions. However, low quality background is not unacceptable for the human visual system since it pays less attention to the background. Therefore, more bits can be allocated to the foreground without distracting the overall quality of the video. In this way, the foreground quality can be boosted. Rate control is essential in low bit rate video coding because it is the mechanism responsible to optimize the video quality with a given target bit rate. As a very useful technique, Region-of-Interest (ROI) video coding provides users more flexibility and interactivity in specifying their desires and enable encoders more efficiency in controlling the visual quality of coded video sequences. This way, the perceptually-important region, for example human faces, can be coded at higher quality to effectively improve the subjective quality of the coded video sequence. Based on region of interest, many schemes have been developed for bit rate control. The following sections discuss four of the different schemes available namely Content based bit rate allocation in section II, content adaptive background skipping in section III, Macroblock layer control in section IV and SSIM-QP based control in section V.

II. Content Based Bit Allocation Model

Traditional video coding schemes are more concerned about the compression ratio, computing complexity, SNR etc. But studies based on human visual perception indicate that humans tend to concentrate on small but important areas of an image or video. In the content based bit allocation model [13, 14, 21and 22] proposed by Wei Lai et al, a video encoding scheme is proposed based on visual attention model. Here the video is divided into several regions of interest with different attention levels. The segmentation of the ROIs is based on the saliency map provided by the attention model [2]. In an image sequence, there are many visual features including motion, motion, color, texture, shape, text region, etc. Also, some recognizable objects, such as face, will more likely attract human attention. Besides, camera operations are often used to induce reviewer's attention. Therefore visual attention models are proposed to model the visual effects due to motion, static, face, and camera attention. In static attention model, a saliency map is generated from each frame by the 3 channel saliency maps computation namely color contrasts, intensity contrasts, and orientation contrasts [2]. We detect the regions that are most attractive to humans by binarising the saliency map. Motion attention model as dynamic, is built based on motion vector field (MVF). It is assumed that a MVF has three inductors: Intensity inductor, Spatial Coherence inductor, and Temporal Coherence inductor. These normalized outputs of inductors are fused into a saliency map by linear combination [3]. Apart from the above features, face is one of the most salient characters of human beings. The appearance of dominant faces in video frames certainly attracts viewers' attention. Camera motion are always utilized to emphasize or neglect a certain objects or a segment of video, that is, to guide viewers' attentions. At last, an overall saliency map is built up by integrating all the

Manuscript received June 23, 2011; revised September 22, 2011.

Mr P Subramanian is a research scholar at JNTU, Kakinada, India and is with Department of Electronics and Communication Engineering, Aarupadai Veedu Institute of Technology, Vinayaka Missions University, Chennai, India (e-mail: psurya602@ gmail.com).

Dr N R Alamelu is now the Principal of Aarupadai Veedu Institute of Technology, Vinayaka Missions University, Chennai, India.

Dr M Aramudhan is with Department of Information Technology, Perunthalaivar Kamarajar Institute of Engineering and Technology, Karaikal, India.

features above.

The ROI extraction process consists of the following steps. Firstly, the ROIs are detected from the saliency map of each frame. These ROIs are classified into several tracks. Each track represents the movement trajectory of a ROI along with time. Then these tracks are combined (if the end-point of one track is close to the begin-point of the other), removed (if one track is too short), and filtered in spatial and temporal. The smoothed tracks are used to generate new saliency map sequence to keep the ROIs in accurate position and size, and avoid the jitter effect which will bring down the visual quality. Rate control is a fundamental technique in the encoding process, which **is** based on the rate distortion theory [4]. For each ROI the attention level is calculated with the non-ROI region treated as minimum attention level ROI. Regions with different attention levels will have different sensitivity to coding errors. Hence a region-weighted distortion model is developed to allocate suitable bits to ROI and non ROI regions thereby resulting in the least distortion. The codec used here is based on the MPEG-4 framework. For base layer encoding, bit rate allocation is performed by using different QP in different spatial regions. The base unit of spatial region is macro block. To encode the MBs with Different QP, the overhead ROI shape information is inserted into the MPEG-4 bitstream. Assume that the maximum of the number of ROIs N<4, then add 2 bits for each MB, represents 4 kinds of value of QP. The decoder received these 2 bits, and chose the corresponding QP to decode the MBs [5]. In FGS layer, selective enhancement technique can he used to emphasis a part of a frame, a bit-plane shifting method is used to put the bit-planes of the MBs of interest earlier in the bit stream so that the are more likely to be included in a truncated bit stream [6]. So that when the bit stream is truncated, the parts been shifted up can get higher quality than the other parts. To avoid shape coding of bit-plane shift, a MAXSHIFT method [7] or Bitplane-by-Bitplane Shift (BbBShift) / Generalized BbBshift (GBbBShift) method [8] can be used. This model has achieved better subjective quality than normal method under constant bandwidth. Experiments conducted with five different test sequences have shown 60% the users as saying the attention model to be better than the normal coding.

III. Content Adaptive Background Skipping

The content adaptive background skipping scheme [9] proposed by Haohong Wang et al, dynamically decides whether to skip Non-ROI macro blocks of current frame and reallocate saved bits to ROI and coded non-ROI based on the real-time content information of the current and previous frames, such as foreground shape deformation, foreground and background motion, and background texture complexity. A generalized adaptive background skipping scheme that takes into account the frame content variation in skip mode decision and rate control is proposed. A previous algorithm [10] used a prototyped unit-based background skipping approach where every two consecutive frames are grouped into a unit in which the non-ROI of the second frame is skipped (not coded but replaced by the macro blocks of the first frame in the same locations) if the distortion caused by the skipping is smaller than a predefined threshold. In this schemer, a much more flexible background skipping scheme is proposed which considers the real-time frame content statistics, such as foreground (or ROI) shape deformation, foreground and background (or non-ROI) movement, and the accumulated skipped distortion due to skipped background, in a jointly framework to make runtime skip mode decisions. In addition, the algorithm dynamically reallocates the saved bits due to skipping to other regions, and adjusts the bit allocation in both frame and macroblock levels. In this scheme, it is assumed the ROI is known at the encoder, which is possible if the ROI is either automatically detected or specified by the end-user. A general problem is to find out the number and locations of the frames in a video sequence whose non-ROIs are to be skipped and the number of bits to be allocated to each frame that ensures the best visual quality of the video sequence. However, to obtain an ideal optimal solution is almost impossible, because the visual quality metric that balances the spatial and temporal quality for a video sequence is still an open issue and in real-time communication systems, the future frames normally are not available when the encoder processes the current frame, therefore the optimality is not achievable. Our goal is to find a low-complexity practical solution for real-time applications that achieve good perceptual video quality. Some perceptual rationales, as used in [11-12], have been considered in this algorithm design, for example, the human visual system (HVS) is more sensitive to the temporal changes than to the spatial details when the frame contains high motion activities. The proposed coding algorithm is briefly described as follows: For each encoded frame, an initial frame-level bit allocation is done by allocating available bits uniformly among the remaining frames in the rate control window. Then, based on a number of content cues, for example, ROI shape deformation and motion activities, and a set of predefined rules, the decision whether to skip the current non- ROI is made, and then the budget for current frame is adjusted to favor the bit reallocation on ROI macroblocks. Next, an optimized macroblock-level bit allocation is conducted, and the frame is coded with the assigned quantization parameters. Clearly, the proposed design favors both spatial and temporal quality. By background skipping and reallocating bits from non-ROIs to ROIs, the spatial visual quality of the frames are improved. On the other hand, the solution is content-adaptive in the sense that it follows some human perceptual rationales for improving temporal video quality. the mode of background skipping is dynamically determined based on the content context such as background and foreground activities. Both foreground and background activities in the framework have been considered. When large amount of motion occurs in background regions, the frequency of background skipping should be reduced. On the other hand, when the foreground contains large amount of activities, skipping background might be helpful to reallocate more bits to code the foreground. Experimental results indicate that the proposed scheme has significant gains of up to 2.5 dB over other approaches.

IV. Macroblock Layer Rate Control

In this scheme proposed by Lin Tong and K. R. Rao [13], an advanced segmentation scheme for real time video at the pre-processing stage has been proposed. Bits are then assigned to different macro blocks based on the segmentation information. A face detection system that combines a feature invariant (skin color) method and a knowledge-based (mosaic rule-based) method is proposed. In the first stage, chrominance value of each pixel is processed by a skin color filter which is defined in Cb-Cr domain. The pixel-based skin color classification information is integrated into macro block level. Then, the classification results are written into a binary mask image, where '1' indicates a macro block is face candidate, '0' means a macro block is background. A 3x3 median filter is applied to smooth out the binary mask image and to remove the noise. The filtered binary mask image is then projected horizontally and vertically. Based on the zero-runs and nonzero-runs in these two detections, the rectangular face candidate region can be obtained. The output candidate face regions are the input to the second stage, mosaic rule-based detection. Adopting it in second stage of the face detection scheme for H.263 video can dramatically reduce the computation complexity.

For inter frames, computation burden and latency are critical. To meet the requirements of low complexity and accuracy, a face tracking method with only motion vector (MV) is proposed for inter frame face segmentation for the ROI based codec. Earlier face tracking with MV was proposed for video retrieval [14]. They use the accumulated motion vector over a number of frames to decide the tracking offset. The algorithm is stable for tracking slow motion; however it is not sensitive to abrupt motion. To balance sensitivity and stability of the tracking system, MV of each individual frame is also checked in this algorithm. Either the accumulated motion vector greater than 16 pixels (a macro block length) or the motion vector of individual frame greater than 8 pixels (half macro block length) is detected; the ROI location is modified along the motion direction. The ROI location size is designed to be elastic so that more face region can be located within the ROI location, i.e., when the ROI location moves half macro block length, the ROI location will extend one macro block in the moving direction, and the ROI location size will shrink back to original when more motion along the same direction is detected. With this modified face tracking method, good tracking efficiency can be obtained. The rate control for real time CBR video is very challenging due to the strict requirements on low latency and small buffer size [15]. The variation of bits-count for each frame should be very small, hence the TMN8 frame layer rate control is employed to provide near constant target number of bits for each frame. In the macro block layer rate control scheme, the major task is to determine QP for each macro block so that the rate constraint T_L x T < R < T_H x T can be satisfied, where T is the target bit budget for a frame, R is the estimated bits-count for the frame, T_L and T_H are the lower and higher percentage bound that restrict the current frame bits-count. The proposed rate control for VBR video includes two parts: a frame layer rate control to balance the temporal quality with spatial quality, and a macro block layer rate control to vary the spatial quality among different regions based on the visual priority.

V. SSIM-QP Based Rate Control

In the SSIM –QP based control scheme proposed by Ling Yang et al [16], in order to analyze the relationship between subjective quality and encoding parameters, a Structural Similarity Index Map – quantization parameter (SSIM-QP) model is established. Through this relation, the possible visual quality range of ROI is defined according to the range of ROI QP, which is predicted by rate control algorithm. Then, with interest levels being identified within the visual quality range, resources allocated to ROI is determined. Finally, considering both the quality of ROI and the entire frame, resource allocation is slightly adjusted. The chosen SSIM metric [17] follows the philosophy that ‘human visual system is highly adaptive in extracting structural information’. Therefore, it employs a modified measure of spatial correlation between the pixels of the original and distorted images to quantify the extent to which the image's structure has been distorted. Therefore, the frames of reconstructed sequences can be compared with the original ones for visual quality evaluation. To predict subjective quality at various encoding configuration, a uniform relation between visual quality and quantization of DCT coefficient is exploited. The visual quality is measured using SSIM metric. Conversely, the quantization parameter corresponds to a target visual quality SSIM can also be determined. Such SSIM-QP relation is used to properly allocate bits to different areas of the frame. The visual quality of ROI can be adjusted according to users’ interest level. The R-Q model is customized with the fact that the texture content and visual quality of ROI and non-ROI are different to achieve more accurate rate control. According to the customized R-Q model, the upper and lower bound of ROI QP, *QPROIMAX* and *QPROIMIN*, are obtained. Then, utilizing the SSIM-QP model established in the former section, the possible visual quality range is estimated and equal difference visual quality ranks of ROI are given. Therefore, the target subjective quality for ROI, also measured in SSIM metric, can be set according to the user’s interest. With the target SSIM, quantization parameter for encoding ROI is first predicted and then adjusted to further save bits for non-ROI region to enhance the entire frame quality. After ROI encoding is done, the rest bits are allocated for non-ROI region by FMO technique [18], and its encoding parameters can also be derived using customized R-Q model.

The current H.264/AVC standard utilizes a quadratic rate-distortion model [19], [20] to calculate the corresponding quantization parameter of a given bit budget. For video coding with ROI, an improved R-VQ performance can be expected if mean absolute difference (MAD), is customized applied with different scene content and regions of different VQ in a sequence. Therefore, the MAD of ROI and non-ROI is predicted separately. And the R-D model is also customized. Since the purpose of rate control is to maintain consistency buffer status, when proper bits are allocated to a frame by bit allocation scheme, the encoder parameter is adjusted to satisfy the bit budget. In the proposed scheme, an extreme condition is that the entire bit budget for one frame is used solely on ROI. At this stage, it reaches to the minimum possible encode parameter for ROI region the proposed scheme performs flexible macroblock order to encode ROI region prior to non-ROI region. Bit budget on ROI is set according the interest level, and also adaptively save bits for non-ROI region when surplus bits are not necessary for ROI region. Simulation results show that our proposed encoder can adjust subjective visual quality of ROI in relatively linear steps. The visual quality of ROI is improved and the visual quality of the entire frame remains acceptable.

References

[1] Wei lai., et al, "A content based bit allocation model for video streaming", IEEE International Conference on Multimedia and Expo, 2004

[2] Itti L, Koch C. "A Comparison of Feature Combination Strategies for Saliency-Based Visual Attention Systems", *SPIE Human Vision and Electronic Imaging IV (HVEl'99).* San Jose, CA, January 1999 Vol. 3644, pp. 373-382.

[3] Yu-Fei Ma, Lie Lu, Hong-Jiang Zhang and Minding Li, "A User Attention Model for Video Summarization", *ACM Mulrimedia'O2,* December, *2002.*

[4] Li Zhao. Qi Wang, *er al.* "A Novel Content-Based Video Streaming Algorithm for Fine Granular Coding" *,International Conference on Managemenr of Multimedia Networks and Services. MMNS200l.* Oct29-Novl,2001.

[5] Ouerhani. N., el *al.* "Adaptive Calor Image Compression based on Visual Attention", IEEE *Image Annlysis and Processing, 2001,* pp.416-421.

[6] Weiping Li, "Overview of Fine Granularity Scalability in MPEG4 Video Standard", *IEEE Trans on Circuits and Sysrems for Video Technology,* Vol. 11, No. 3, March 2001

[7] Christopoulos, C., el *nl,* "Efficient Region of Interest Coding Techniques in the Upcoming JPEG2000 Still Image Coding Standard", *IEEE Image Processing, 2000.* Vo1.2, pp. 41-44.

[8] Zhou Wang., *er al.* "Generalized Bitplane-by-Bitplane Shift Method for PEG2000 KO1 Coding", IEEE *Image Processing. 2002.* Vo1.3, pp. 81-84.

[9] Haohong Wang., et al, "Real Time Region of Interest Video coding using content adaptive background skipping with dynamic bit reallocation", ICASSP,2006.

[10] H. Wang and K. El-Maleh, "Joint adaptive background skipping and weighted bit allocation for wireless video telephony", in *Proc.International Conference on Wireless Networks, Communications and Mobile Computing*, Maui, Hawaii, USA, June 2005.

[11] F. C. M. Martins, W. Ding, and E. Feig, "Joint control of spatial quantization and temporal sampling for very low bit rate video", in *Proc. ICASSP*, May 1996, pp. 2072-2075.

[12] F. Pan, Z. P. Lin, X. Lin, S. Rahardja, W. Juwono, and F. Slamet, "Content adaptive frame skipping for low bit rate video coding", in *Proc. 2003 Joint Conference of the Fourth International Conference on Information, Communications and Signal Processing, and the Fourth Pacific Rim Conference on Multimedia*, Vol. 1, Dec. 2003, Singapore, pp.230 – 234.

[13] Lin Tong and K R Rao, " Region of Interest based H.263 compatible codec and its rate control for low bit rate video conferencing", Proceedings of 2005 International Symposium on Intelligent Signal Processing and Communication Systems, December 2005, Hongkong

[14] V.Mezaris et al, "Real-time compressed-domain spatiotemporal segmentation and ontologies for video indexing and retrieval", IEEE Trans. CSVT, vol.14, pp. 606-621, May 2004.

[15] J.Ribas-Corbera and S.Lei, "Rate control in DCT video coding for low-delay communications", IEEE Trans. CSVT, vol.9, pp. 172-185, Feb. 1999.

[16] Ling yang et.., al, "A ROI quality adjustable rate control scheme for low bit rate video coding".

[17] Zhou Wang, Liguang Lu, Alan C. Bovik, "Video quality assessment based on structural distortion measurement," *Signal Processing: Image Communication*, 19(2004) 121-132

[18] Stephan Wenger, Michael Horowitz, FMO: Flexible Macroblock Ordering, Joint Video Team (JVT) of ISO/IEC MPEG & ITU-T VCEG (ISO/IEC JTC1/SC29/WG11 and ITU-T SG16 Doc. JVT-C089, Fairfax, Virginia, USA, 6-10 May, 2002

[19] Siwei Ma, Zhengguo Li, Feng Wu, Proposed draft of Adaptive rate control, Joint Video Team (JVT) of ISO/IEC MPEG & ITU-T VCEG(ISO/IEC JTC1/SC29/WG11 and ITU-T SG16 ,8th Meeting: Geneva, May 20-26,2003

[20] Yang Liu, Zhengguo Li, Yeng Chai Soh, "A novel rate control scheme for low delay video communication of H.264/AVC standard", IEEE Trans. On Circuits and Systems for Video Technology, Vol.17, no. 1, January 2007

[21] D.Vijendra Babu, Dr.N.R.Alamelu,"Wavelet Based Medical Image Compression Using ROI EZW",International Journal of Recent Trends in Engineering(1797-9617),Vol. 1, No. 3, May 2009.

[22] D.Vijendra Babu, Dr.N.R.Alamelu, P.Subramanian, "Energy Efficient Wavelet Based Medical Image Compression Using Modified ROI EZW",Proceedings of the Int. Conf. on Information Science and Applications, February 2010, ISBN 978-81-907677-9-8

P Subramanian completed his B.E. in Electrical and Electronics Engineering from Coimbatore Institute of Technology, Coimbatore, India in the year 1997 and his M.E. in Applied Electronics from Madurai Kamaraj University, India in the year 2000. He is currently pursuing his PhD at Jawaharlal Nehru Technological University, Kakinada, India.

He is currently working as Associate Professor in the Department of Electronics and Communication Engineering at Aarupadai Veedu Institute of Technology, Vinayaka Missions University, Chennai, India.

Mr Subramanian is a member of IEEE since 2006. He is also a member of IEEE Communications Society and IEEE Circuits and Systems Society.

Dr N R Alamelu completed her B.E. in Electronics and Communication Engineering with honors from PSG College of Technology, Coimbatore, India in the year 1981 and her M.E. in Applied Electronics from Government College of Technology, India in the year 1983. She obtained her PhD in the area of Broadband Networks from Bharathiyar University, India in 2004.

She is currently the Principal of Aarupadai Veedu Institute of Technology, Vinayaka Missions University, Chennai, India. She has more than 25 years of teaching experience and has published in a number of International Journals and Conferences.

Dr Alamelu is a member of IEEE since 2006. She is also a member of IEEE Communications Society. She is the Secretary of IEEE Madras Section and is the Chairperson of the IEEE Communications Society, Madras Section. She is a Life Member of ISTE and Fellow IETE

Dr M Aramudhan completed his B.E. in Computer Science and Engineering from Regional Engineering College, Trichirapalli, India in the year 1997. He completed his M.E. in Computer Science and Engineering from Regional Engineering College, Trichirapalli, India in the year 2002. He obtained his Ph D from Anna University, Chennai, India in the year 2008.

He is currently working as Assistant Professor in the Department of Information Technology at Perunthalaivar Kamarajar Institute of Engineering and Technology, Karaikal, India. He has more than 10 years of teaching experience and has published in a number of National and International Journals.