

Improving Video Coding at Scene Cuts using Attention based Adaptive Bit Allocation

Zhibo Chen¹, Guoping Qiu², Yang Lu¹, Lihua Zhu¹, Quqing Chen¹, Xiaodong Gu¹, Charles Wang¹

¹ Corporate Research, Beijing, Thomson

² School of Computer Science, University of Nottingham, UK
chenzb@ieee.org

Abstract—

Existing video coding methods can cause visual quality and buffer occupancy to fluctuate significantly at scene cuts. To address this problem, we have developed a novel visual attention based adaptive bit allocation method. We first perform scene cut detection to extract frames in the vicinities of dramatic scene changes; we then perform visual saliency analysis on those frames to grade the macro-blocks according to their visual importance; and finally we devise a visual attention based adaptive bit allocation scheme which assigns more bits to visually salient blocks and fewer bits to visually less important blocks. We will present experimental results which demonstrate that at scene cut areas, coding quality in terms of PSNR of our method are both higher and much smoother than those of existing coding methods and the buffer occupancy of our method is also much more consistent and has less fluctuation. Our method is compatible with other rate control schemes and can be easily implemented to improve existing video coding standards such as MPEG-2, H.264/AVC, and others.

Keywords— scene cut, bit allocation, attention area

I. INTRODUCTION

The hybrid video coding framework has been widely adopted in existing video coding standard such as H.263, MPEG-2 and H.264/AVC. Extensive work has been carried out to improve visual quality within a certain bit rate constraint. Generally CBR (Constant Bit Rate-control) and VBR (Variable Bit Rate-control) are used to meet the tradeoff between quality and rate constraint according to the requirements of different applications. In CBR mode, the number of bits that can be transmitted to a video decoder in a given time interval is typically fixed. The decoder side will also use a buffer of pre-specified size referred to as the video buffer verifier (VBV) in MPEG2 and MPEG4-part2 or Hypothetical Reference Decoder (HRD) in h.263 and MPEG4-AVC/H.264. Related applications are broadcasting, cable transmission, and wireless transmission of compressed video, etc. In VBR mode, the total number of bit used to compress a long stretch of video is typically fixed, while no limits is imposed on instantaneous bit rate. Related applications are stored media application like DVD (Digital Versatile Discs) and PVR (Personal Video Recorder), etc.

Due to the high variability present in many video sources, a long video can be divided into video shots. A video shot may be defined as a sequence of frames captured by “a single camera in a single continuous action in time and space”.

Usually it is a group of frames that have consistent visual characteristics (including color, texture, motion, etc.). Consequently a large number of different types of scene changes can exist between shots. A **scene cut** is an abrupt transition between two adjacent frames, and many methods have been studied, e.g. using histograms to compare consecutive video frames [1] or measuring the Mutual Information (MI) can be used for detecting scene cut [2].

As shown in Figure. 1, usually the picture at scene cut is coded as Intra frame, and different bit allocation schemes will be adopted in CBR and VBR cases. In CBR cases, the encoder will try to keep the bit rate constant, which will often cause serious quality degradation in scene changing areas. In VBR case, more bits will be allocated to the frames at scene changing intervals and usually subsequent frames will be skipped according to buffer or transmission rate constraints, which will often cause jerk artifact in video display. If an algorithm doesn't handle scene change well, the encoder will usually consider this as a picture similar to previous one and allocate bits accordingly. If this picture is coded as P or B frames, the picture quality will be seriously distorted due to fewer allocated bits. Parameters from previous pictures are usually used to train parameters for future pictures for most rate control algorithms, which is not valid when a scene change occurs. This also results in a quality break and the more accurate the bit rate control is, the more severe the problem is.

[3] proposes a method for scene cut coding using non-reference frames, which compress the scene cut frame and its neighboring frames (before and after) as non-reference frame (B frame type) with increased quantization parameters to reduce bandwidth, as shown in Figure.2. In this case, the coding efficiency of the first P frame after scene cut is very low due to long prediction distance, a longer delay is needed, better performance in B frame coding and tradeoff between quality and rate constraint can't be assured.

Therefore existing scene change bit allocation schemes couldn't provide a good solution which will achieve a good tradeoff between smooth subjective quality and consistent buffer usage, especially in the intervals of large scene changes.

Visual temporal masking occurs when there is a sudden change of stimulus, a situation usually happens during scene changes. Based on the effect of temporal masking [4] and theory of “change blindness” which illuminated that change blindness is evident when mistakes occur far from the viewer's focus of attention [5], a novel solution named **ASBA**

(Attention based Scene cut Bit Allocation) with optimal perceptual video coding based on attention area extraction is proposed in this paper, which can smooth the output of the video encoder without serious reduction of the video subject quality.

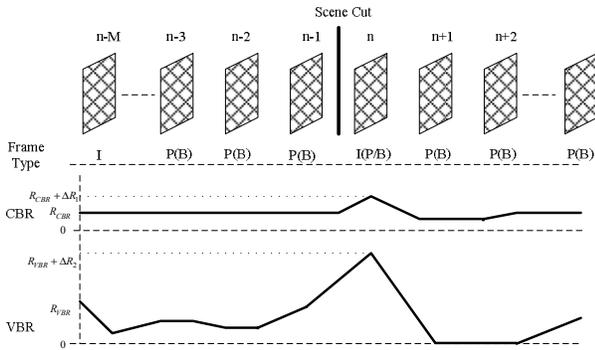


Figure 1. Typical process of CBR and VBR in scene cut

Section II describes the basic framework of the proposed scheme, experiments results on analysis of perceptual quality and buffer variance performance are introduced in Section III. Conclusion is given in the last section.

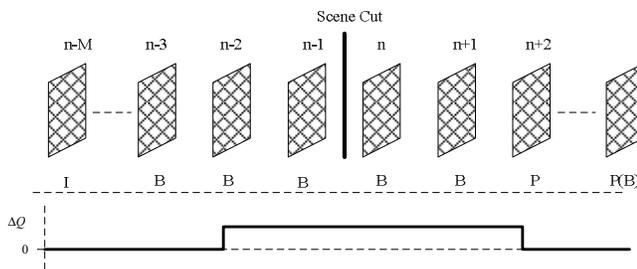


Figure 2. Coding of scene cuts in video seq. using non-reference frames

II. FRAMEWORK AND FLOWCHART OF ASBA SCHEME

A. ASBA Framework

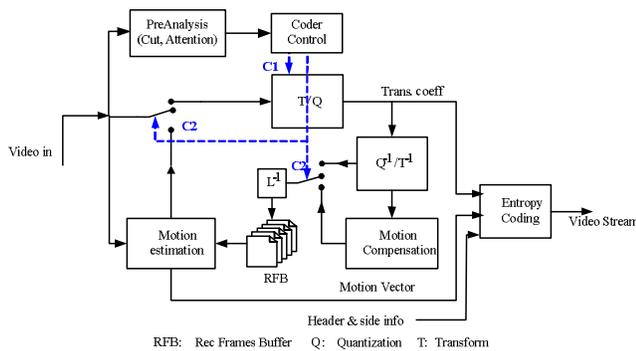


Figure 3. Framework of ASBA scheme

Figure.3 illustrates the framework of the proposed ASBA scheme. It adds two modules to the traditional hybrid coding framework, one is PreAnalysis module and the other is Coder Control module. In this paper, the PreAnalysis module includes operations of scene cut detection and attention area extraction. Based on the information from PreAnalysis module, the Coder Control part send control information **C1** to Transform/Quantization module (T/Q) to control quantization of each block, send control information **C2** to mode selection module (frame or MacroBlock mode) to control the suitable mode for each frame and MB (MacroBlock).

For attention area, human always pay more attention to some parts of the picture others, we call the areas that attract human visual attention “attention area”. In classical attention model [6], first a set of feature maps in gray level are extracted from visual input of a given image. The features include intensity, color, orientation, and so on. Then in each feature map, the most salient areas are picked out. Finally all feature maps are integrated, in a purely bottom-up manner, into a master “saliency map”, which is regarded as the attention information of a picture [6]. Then an attention mask can be obtained for each picture which describes the different attention importance of different area of a picture.

In our study, the bottom-up attention model in [6] is used for attention area extraction and our empirical results show that, when the size of extracted attention area is about 20%~40% of the picture size, it yields a satisfied performance for the attention based application [7]. The computation complexity for analyzing each feature map is high, but it is achievable to optimize this algorithm for real-time applications. The outcome of this algorithm is a macroblock based attention mask, which may be used conveniently for the macroblock based hybrid video coding framework.

Figure.4 shows a sample of attention mask, which divides macro-blocks in the picture into 4 level sets of attention importance, M_i ($i=1\sim 4$) represents the ratio of visual importance of each level set, a larger M_i means that larger distortion can be tolerated corresponding to visually less important area. We assume that attention mask can be obtained in the PreAnalysis module.

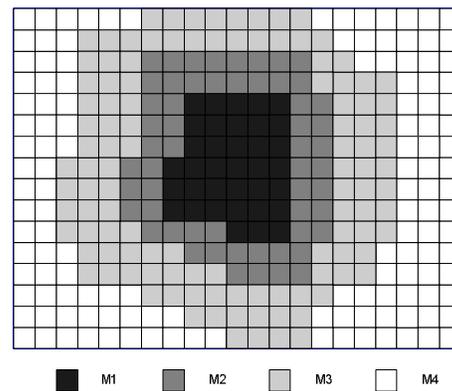


Figure 4. Sample of Attention Mask

B. ASBA flowchart

Figure.5 shows the concept of “scene cut period” which represents period of frames close to (before an after) scene cut. The window size for scene cut period is determined by the change blindness theory [4]. Usually we can set $L1 = 1 \sim 4$ and $L2 = 2 \sim 4$. Therefore scene cut period detection can be done with the same method of scene cut detection, only with a delay of $L1$ frames.

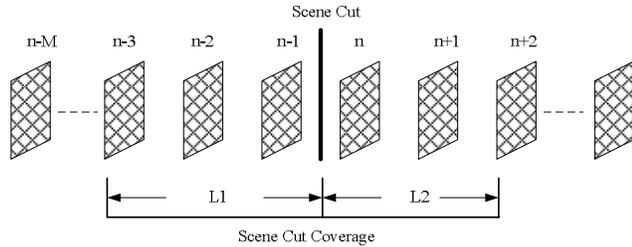


Figure 5. Illustration of scene cut period

Figure.6 shows the flowchart of ASBA scheme, where scene cut period detection is done in the PreAnalysis module after each video input.

If the video frame does not belong to a scene cut period, the frame type and MB type decision can be made according to default schemes (any existed rate control scheme) and coding can also be performed according to the default schemes.

If the frame belongs to scene cut period, the attention area of the frame is extracted based on methods proposed in [6][7].

In this paper, the frame level bit allocation adjustment is only applied to the frames belong to scene cut periods, and an attention based MB level bit allocation scheme is implemented in these frames.

As shown in Figure. 7,, where area with diagonal line indicates attention area and area with cross line indicates non-attention area. If scene cut happens, attention area has the highest priority to be allocated with more bits to keep the subjective quality in attention area, while non-attention area has the lower priority will be allocated fewer bits.

For the frames before the scene cut, bits can be removed from the non-attention area to free the buffer occupancy, which introduces neglectable degradation on subjective quality due to the effect of temporal “backward masking” effect [4].

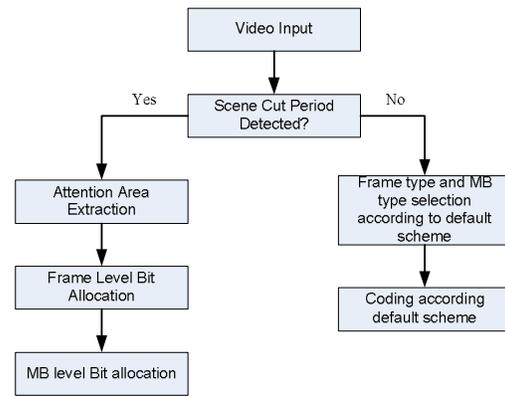


Figure 6. Flowchart of the ASBA scheme

As shown in Figure.7, the number of bits ΔR_i that is removed from non-attention area in the i -th frame inside scene cut period ($-L_1 \leq i \leq L_2$, $i = 0$ stands for the scene cut frame and $i < 0$ stands for the frames before the scene cut, $L1=2$ and $L2=4$ in this case), as defined in equation (1), where ρ is the ratio of attention area inside one frame, R_F is the total bits for one frame, and K_i are control factors ($0 < K_i \leq 1$) defined in real application to give a better tradeoff between attention area distortion and non-attention area distortion, and it can also be modulated adaptively in real application.

$$\Delta R_i = \left(\frac{K_i \rho}{1 - \rho + K_i \rho} - \rho \right) R_{F_i} \quad (1)$$

In case of $i < 0$, the removed bits from the non-attention area are saved to free the buffer occupancy. While in case of $i \geq 0$, the removed bits are reallocated to the attention area to improve the perceptual video quality.

The adjustment of the bit allocation for the frames inside scene cut period can efficiently improve the subjective quality under the same frame level bit allocation, as well as efficiently reduce the buffer occupancy and reduce the probability of buffer overflow after the scene cut.

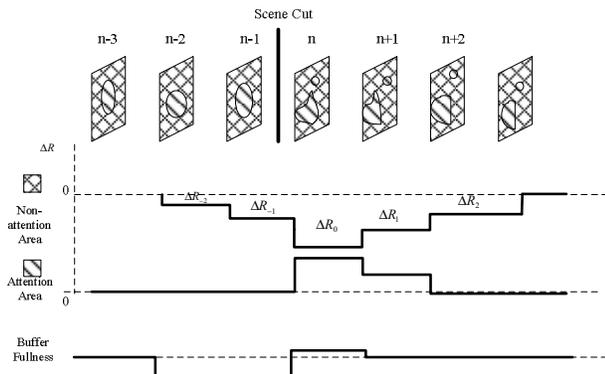


Figure 7. Illumination of ASBA in CBR case and the buffer variance analysis

III. EXPERIMENTAL RESULTS

In this section, simulation results are given to illustrate the effectiveness of the proposed ASBA scheme. The ASBA frame level bit allocation scheme is compared with CBR and VBR frame level rate control algorithms. As constant QP usually gives the smooth quality fluctuation, the VBR rate control uses constant QP for all frames in the sequence. The rho-domain based bit allocation [8] is employed for macroblock level rate control.

The movie trailer of “delovely” with the format of 720×480 is used as the test sequence, which includes many scene cuts and Jm1.0 is used as the test bed. The target bit rate for CBR, VBR and ASBA is set as 600kbps. Using the scene cut detection algorithm in [2], the scene cut will appear at the 36-th, 64-th, 110-th and 133-th frame. For each i th frame, $K_i = 1$ means change blindness effect is not considered during bit allocation. The smaller K_i is, the larger quality difference between attention area and non-attention area is, whose value can be adjusted adaptively according to the content property and bit rate constraint. In our experiment, it is fixed as 0.3 to show the primary performance improvement.

The frame level PSNR variation is illustrated in Figure.8. And since theory of “change blindness” illuminates that change blindness is evident when mistakes occur far from the viewer’s focus of attention, the PSNR variation of the attention area is specially described in Figure.9 to reflect the real subjective quality experience in case of scene cut. From Figure.8 and Figure.9, it can be seen that compared with CBR, the ASBA significantly reduces the quality variation to keep a consistent quality at the scene cut period, and subjective quality verification also proves this results.

The buffer fullness status comparison among CBR, VBR and the proposed ASBA scheme is illustrated in Figure.10. It is obviously that during the scene cut period, the buffer occupancy fluctuation of ASBA will be much smaller than VBR, and very close to that of CBR.

Therefore, by exploiting the temporal masking effects and theory of change blindness, the ASBA scheme achieves a better tradeoff between consistent quality and smooth buffer fluctuation.

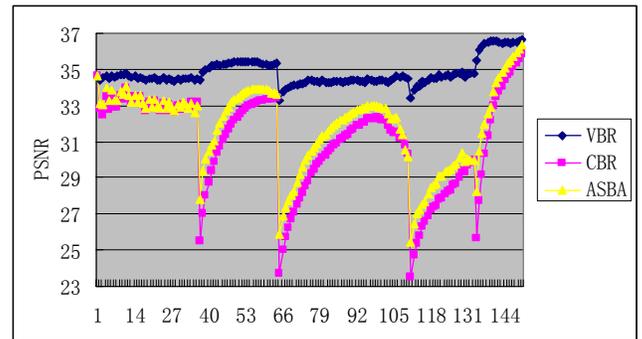


Figure 8. The average PSNR of each frame

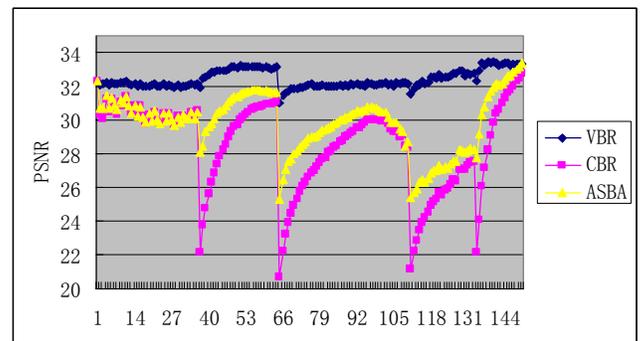


Figure 9. The average PSNR of the attention area for each frame

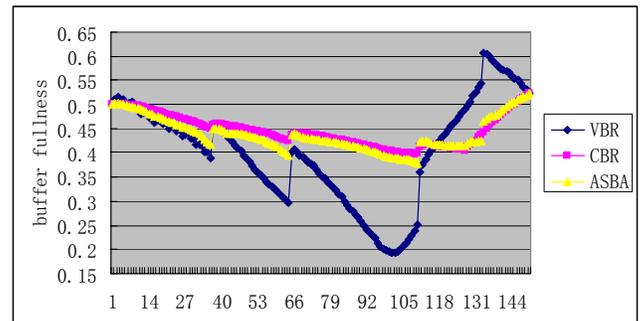


Figure 10. Buffer fullness status comparison

IV. CONCLUSION

Using the method in this paper for video coding, skipped frames can be reduced by putting limited resource to attention area to reduce necessary bits for scene cut frames, therefore to improve decoder synchronization, and remove jerk artifact. Due to the change blindness property of human vision, a better tradeoff between spatial/temporal quality assurance and buffer rate constraint can be achieved. The proposed scheme is

compatible with any other rate control schemes and suitable for any existed video coding standard like MPEG-2, H.264/AVC, H.263, etc. And advantageously, the ASBA scheme can also be applied in case of special scene changes like fade in/out or dissolve.

REFERENCES

- [1] H.J.Zhang, A.Kankanhalli and S.W.Smoliar, "Automatic partitioning of full-motion video", in *Multimedia Systems*, volume 1, pages 10-28, 1993
- [2] Zuzana Cernekova, Ioannis Pitas and Christophoros Nikou, "Information Theory-Based Shot Cut/Fade Detection and Video Summarization", *IEEE CSVT*, pages 82-91,2006
- [3] "Coding of scene cuts in video sequences using non-reference frames", US Patent 2005/0286629 A1, Dec.29, 2005
- [4] Breitmeyer, B.G., Ogmen, "Recent Models and findings in visual backward masking: A comparison, review and update", *perception & psychophysics* 72(8): 1572-1595
- [5] Daniel J Simons, Christopher F Chabris, "Gorillas in our midst: sustained inattentional blindness for dynamic events", *Perception*, 1999, volume 28, pages 1059-1074
- [6] L. Itti, C. Koch, E. Niebur, "A Model of Saliency-based Visual Attention for Rapid Scene Analysis," *IEEE Transactions On Pattern Analysis and Machine Intelligence*, vol. 20, no. 11, Nov 1998, pp.1254-1259
- [7] Quqing Chen, Zhibo Chen, Xiaodong Gu, Charles Wang, "Attention based Adaptive Intra Refresh for Error Resilient Video Coding", 25th Picture Coding Symposium, Beijing, 2006
- [8] Z. He and S. K. Mitra, "A linear source model and a unified rate control algorithm for DCT video coding", *IEEE Trans. Circuits and System for Video Technology*, Vol. 12, No.11, Nov. 2002. pp. 970-982