Motion-based Adaptive GOP Algorithms for Efficient H.264/AVC Compression

Jun-Ren Ding, Ji-Kun Lin and Jar-Ferr Yang

Institute of Computer and Communication Engineering
Department of Electrical Engineering
National Cheng Kung University, Tainan, Taiwan 701
E-mail: jfyang@ee.ncku.edu.tw

Abstract

H.264/AVC provides several advanced features such as improved coding efficiency and error robustness for video storage and transmission. In this paper, to further improve the coding quality for encoding possible scene changes and fast variation sequences, we propose a new adaptive GOP scheme that uses the existing motion vectors and motion residuals. The proposed scheme effectively identifies the H.264/AVC coding core {I, P} frames using a statistical table that improves the coding efficiency for changed scene sequences and different video content variance. Simulation results show that the proposed adaptive GOP scheme can increase PSNR 0.65 dB on average and achieve a 92% correct rate for scene change detection from the original H.264 reference software JM10.1 operated in the fixed GOP size.

Keywords: AGD, SCD, H.264/AVC.

1. Introduction

The H.264 advanced video coder is the latest video standard. It outperforms the other standards, providing the best video compression [1][2]. Some highlighted features of H.264 include motion estimation in variable block sizes, multiple reference frame motion compensation, spatial prediction for intra coding, small block size (4×4) residual transform coding, adaptive and hierarchical block size transform, etc. For broadcasting video services, a group of pictures (GOP) with a proper intra frame insertion should be designed for practical considerations. However, the fixed GOP size (FGS) will use more bits to encode the intra frame and the inter frames, which consist of scene changes. To improve the coding performance, we should adapt the GOP size properly such that we can encode the intra frame in better locations. There are some researches that have suggested adaptive GOP detection (AGD) algorithms [3][8][9], but fail in needing a GOP-length delay with adjusted inter coding {B, P}, and not consider the better location decisions for intra frame {I} which hold a lot of bit-rate. The method in [3] does not consider scene change detection (SCD) in their adaptive GOP algorithm. The method in [10] uses only the mean square prediction error (MSPE) pixel by pixel to find the scene change and does not consider the motion analysis (MA) data. In this paper, we developed a new AGD method to further advance the video quality of H.264/AVC coders in the coding core {I, P}. Man-made gradual scene changes are not considered. In Section II, we propose a new adaptive GOP method based on the video content variance (VCV), which is computed by motion vectors. In Section III, we propose a SCD method that uses motion vectors and motion residuals, which are the sum of the absolute transformed differences (SATD). Simulation results and discussions are addressed in Section IV. Finally, we conclude the paper in Section V.

2. Adaptive GOP Based on Motion Information

Generally, if the video frames with smaller VCV are coded as intra frames, we will waste a lot of bits in video coding. Conversely, if two scene changed frames are coded using inter frames, it will also become inefficient. With FGS, the coding performance of H.264/AVC in the above two cases will not be effective. Usually, the VCV of two consecutive frames could be evaluated by computing the total difference between the two frames. Intuitively, the VCV could also be expressed using the motion vectors obtained after motion estimation. If the motion vector is zero, the video is actually a still image. Conceptually, small motion information means small VCV. In this paper, we utilize the motion information to design an AGD algorithm. The motion information will not introduce any computation during mode decision. The sum of the absolute motion vectors (*SAMV*) in the current frame in a 4×4 block size after normalization is expressed using

$$SAMV = \sum_{i=0}^{N}\sum_{k=0}^{M}[(|X_{sms_x}|+|Y_{sms_y}|)_k \times \rho_k], \qquad (1)$$

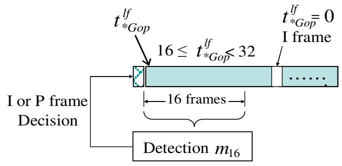
where N is the total number of macro-blocks in a frame, M is the total number of sub-macro-blocks in each macro-block, X and Y are the horizontal and vertical components of the motion vector after motion estimation, and the subscripts, sms_x and sms_y denote the width (x) and height (y) of the kth sub-macroblock size (*sms*), respectively. In (1), ρ_k, which is treated as a normalization factor for none 4×4 sub-macroblocks, is given by

$$\rho_k = (sms_x \times sms_y)_k / 16 \qquad (2)$$

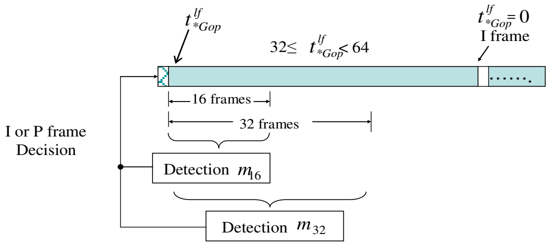
If a marcoblock, for example, is finalized to be split into two 8×16 sub-macro-blocks and their motion vectors are (6, 8) and (9, 8), so sms_x = 8, sms_y = 16 and ρ_k = 8. Therefore, to obtain the sum of total motion vectors in 4×4 sub-macro-block size, the *SAMV* of the macro-block is equal to (6 + 8) × 8 + (9 + 8) × 8 = 248. In (1), in order to simplify the computational complexity, we used motion vector addition instead of Euclidean distance. According to the variation in *SAMV*, we can adaptively change the GOP size for each group of frames in any location using the optimum statistics results in Table 1. We will make a decision to code {I, P} frame for the next frame, we compute the total VCV of the t^{th} frame (named as $SAMV_t$), where the logarithm of the mean of the $SAMV_t$ in the detection length d is denoted as

$$m_d = \log_2\left(\frac{1}{d}\sum_{t=t^{lf}_{*GOP}-d}^{t^{lf}_{*GOP}} SAMV_t\right) + \gamma \qquad (3)$$

where d represents one of five pre-determined detection lengths stated in Table 1, γ denotes an offset value, and t^{lf}_{*GOP} denotes the frame index of the last frame at the uncertain GOP (*GOP) size. For example, from starting detection length t^{lf}_{*GOP} = 16, as Fig. 1(a), If m_{16} is larger than or equal to a threshold T_1, said 6, the next frame will be decided to code as I frame, and t^{lf}_{*GOP} will be reset to be zero for the next AGD procedure. Then, we will encode 16 P frames, where we choose 16 as the minimum GOP size. Once $t^{lf}_{*GOP} \geq 16$, we then start to detect whether m_d satisfies the thresholds defined in Table 1 or not. If m_{16} is less than a threshold T_1, as Fig. 1(a), the detecting optimal GOP size will be increased one by $t^{lf}_{*GOP} = t^{lf}_{*GOP} + 1$ and proceed another detection cycle again. As Fig. 1(b), If t^{lf}_{*GOP} exceeds 32, we now should detect m_{32} and m_{16} in each adaptive procedure. After including one more frame, all m_d's for $d < t^{lf}_{*GOP}$ 256 of the maximum GOP size. Of course, the minimum and the maximum GOP sizes could be varied based on various considerations by the designers.



(a) Adaptive GOP Detection (16 < GOP Size < 32)



(b) Adaptive GOP Detection (32 < GOP Size < 64)

Figure 1. Example of Proposed AGD

Table 1. Thresholds used in the proposed AGD method

m_d d	$m_d < 0$	$m_d \geq 0$	$m_d \geq 3$	$m_d \geq 5$	$m_d \geq 6$
256	I				
128		I			
64			I		
32				I	
16					I

As to Table 1, we have analyzed 24 CIF-format videos and found all the thresholds of m_d for the proposed AGD method. With $\gamma = -14$, Table 1 shows the thresholds versus the detection lengths. Using 24 videos, we found that the motion velocity of the video frames could be classified into low motion (73%), medium motion (6%), high motion (0.5%), pans and zooms/rollins (20%) frames. The thresholds {0, 3, 5, 6} are used to detect the GOP sizes, {256, 128, 64, 32, 16}. Generally, for the higher VCV frames, the proposed AGD will achieve a smaller GOP size. Conversely, for the lower VCV, a larger GOP size will be detected generally.

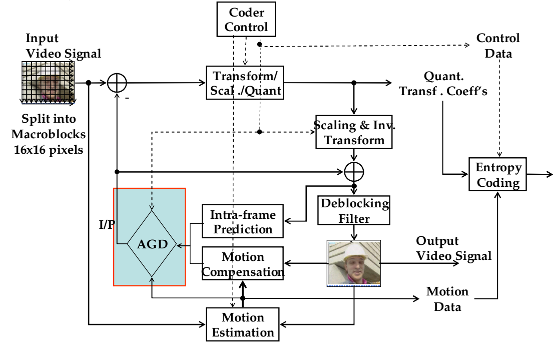


Figure 2. Insertion of the MV-based adaptive GOP detection function in H.264/AVC encoding loop

As shown in Figure 2, we can insert the AGD function before {I, P} frame detection in the H.264/AVC encoding loop. The proposed MV-based AGD method effectively uses the VCV characteristics without physically increasing the computation. We can improve the video quality using AGD if we can choose a better {I, P} coding decision for the next frame. If we simply spend 1-bit to represent the {I, P} coding in the header information for each frame, we can also save some computation for the encoder by performing only the intra frame coding. To precisely detect the true changed scene frame, we suggest that the proposed AGD method could be co-worked with a simple SCD algorithm, which will be addressed in the next section.

3. SCD Based on Motion and Residual Information

To improve the coding performance, the SCD is very important for many entertainment videos. Many advanced video applications also require the SCD algorithm to discriminate video content. The proposed AGD should further combine the SCD to avoid the inter frame coding of scene-changed frames in some movie sequences. In this section, we also use the motion information to achieve an efficient SCD method. With the existing motion information, again, the SCD method will not take any extra computation. Here, we suggest that the SCD needs to monitor the variation in $SAMV_t$ and $SAMV_{t-1}$ and detect the discrepancy of motion residuals at the t^{th} and $(t-1)^{th}$ frames. If the t^{th} frame has scene change, we observe that $SAMV_t$ will extremely small than $SAMV_{t-1}$ since the scene change frames are uncorrelated such that the most of the macro-blocks will be coded with intra coding modes. At the same time, the scene change will result in large sum of absolute transformed differences (*SATD*). Hence, we propose the SCD method could simply perform the motion VCV ratio (MVR) test depicted by

$$\frac{SAMV_t+\varepsilon}{SAMV_{t-1}+\varepsilon}\le sp_1 \qquad (4)$$

and the motion residual ratio (MRR) test given by

$$\frac{SATD_t}{SATD_{t-1}}\ge sp_2 \qquad (5)$$

where sp_1 and sp_2, which are statistical parameters are set to 0.15 and 1.3 respectively. The first VCV ratio test will detect most of the scene changes. In (4), we need to add a small constant ε to avoid singularity problem since the smaller VCV will lead to the $SAMV_t = 0$. However, the true still image sequence might have small $SAMV_t$. To avoid the static case, the motion residual ratio test can robust the $SAMV_t$ method in successfully detect the scene change.

There are advantages using the above SCD in the H.264/AVC: 1) we can further introduce some flexibility for video summarization, annotation, and indexing in video database management systems; 2) we can slightly improve the video compress efficiency and robust the AGD.

4. Simulation Results

For H.264/AVC, we compare the performance of the propose system with that of JM10.1 software. The peak signal to noise ratio (PSNR) is measured and the test conditions are defined as: 1) Search range: 16, 2) Total number of reference: 1, 3) Reference for P slices: 1, 4) Sequence type: IPPP, 5) Entropy coding method: CAVLC, 6) ME scheme: Full search, 7) Intra Period (Fixed GOP size): {16, 32, 64, 128, 256} and 8) RD Optimization: 0.

As shown in Figures. 3 for the smaller VCV sequence “News_Akiyo” and the larger VCV sequence “Football_Stefan”. They have very similar VCV but yet should be coded by different GOP size, respectively. The proposed algorithm will lead to better coding improvement. The proposed algorithm will lead to a lower coding improvement than the original H.264 which is used FGS. A key contribution is that we can determine the VCV before coding procedure. If we did not properly use the VCV, we will waste a lot of bits to encode I-frame or P-frame for scene change periods and different VCV.

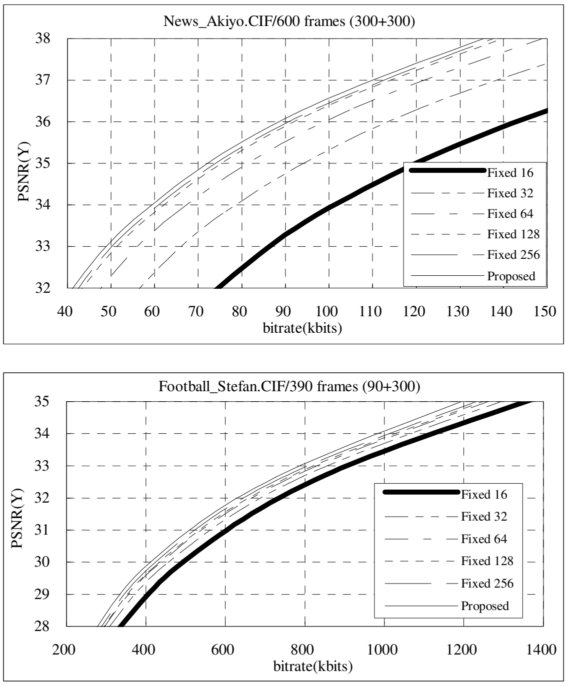


Figure 3. RD curves of the proposed and the original method for the News_Akiyo and Football_Stefan combined sequences.

As shown in Table 2, the proposed AGD or SCD method could properly detect all suitable intra frames for the combined or no combined video sequences to improve the coding performances. Figure 4 shows the variations in *SAMV* and *SATD* in detecting the combined sequence, which is composed of Football, Bus, Basketball, Table tennis, Singer, and News sequences. The dash lines are properly marked at the scene change positions, which are detected using the proposed method. Six frames were found using SCD because the “Table tennis” sub-sequence already contained a scene change in its content. Simulations show that the SCD achieved robust performance for AGD. With the AGD and the SCD methods, we could improve the video quality in 0.65dB PSNR on average in comparison to the original JM10.1 references software using various FGS. Note that the GOP mechanism will restrain the error propagation before I frames and improve the coding performance for the changed scene sequences.

We discovered something else in our experiences. The AGD has a smoother distortion curve. As shown in Fig. 5, the H.264 with a FGS has higher variation in video quality (PSNR) and wasted more bit-rate than the proposed method even if we used the same QP selection.

With the proposed adaptive GOP methods, H.264 becomes more suitable for modern video compression technology.

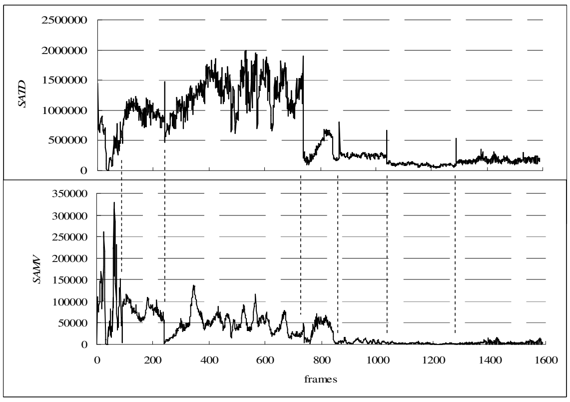


Figure 4. SCD in combined sequences "Football_Bus_Basketball_Table_Singer_News.CIF"

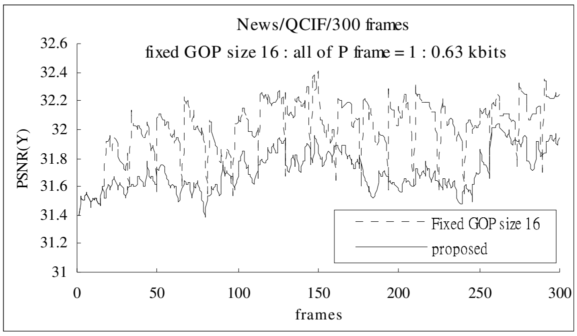


Figure 5. Distortion curve with AGS

5. conclusions

H.264/AVC is known as the most advanced video compression method. However, it is not so efficient for coding high VCV and changed scene sequences. To further improve the coding performance of H.264/AVC, we proposed the AGD method and SCD algorithm based on motion information, which includes the motion vectors and the residual after motion compensation. Simulations exhibit that the proposed methods can improve the coding performance greater for the lower VCV than for the higher VCV. In the future, we will apply the VCV to smooth the distortion for vision subjectivity quality. We believe that the proposed methods combined with the rate distortion optimum (RDO) [4][5] could further improve the coding performance. Using the RDO with shape information to dispose more bit-rate for motion objects in a static camera [6][7] is also possible.

References

[1] T. Wiegand, G. J. Sullivan, G. Bjontegard, and A. Luthra, "Overview of the H.264/AVC video coding standard,"IEEE Trans. on CSVT, vol. 13, no. 7, pp. 560-576, July 2003.

[2] "Draft ITU-T recommendation and final draft international standard of joint video specification (ITU-T Rec. H.264/ISO/IEC 14 496-10 AVC ", in Joint Video Team (JVT) of ISO/IEC MPEG and ITU-T VCEG, JVTG050, 2003.

[3] Y. Yokoyama, "Adaptive GOP Structure Selection for Real-Time MPEG-2 Video Encoding," IEEE International Conference on Image Processing, vol. 2, pp. 832-835, Sep. 2000.

[4] F. Pan, Z. Li, K. Lim, and G. Feng, "A study of MPEG-4 rate control scheme and its improvements,"IEEE Trans. on Circuit Syst. Video Technol., vol. 13, 440-446, 2003.

[5] Z. Li, F. Pan, K. P. Lim, G. Feng, X. Lin and S. Rahardja,"Adaptive basic unit layer rate control for JVT," JVT-G012-r1, 7th Meeting, Pattaya II, Thailand, Mar 2003.

[6] H. Luo,"Image-dependent shape coding and representation", IEEE Transactions on Circuits and System for Video Technology, vol. 15, no. 3, march 2005.

[7] D. Zhang and G. Lu, "Review of shape representation and description techniques," Pattern Recognition., vol. 37, no. 1, pp. 1–19, 2004.

[8] S.-C. Hsia, S.-C. Cheng and C.-L. Chen, "A real-time chip implementation for adaptive video coding control", IEEE Trans. Circuits and Systems on Video Technology, vol. 14, no. 8, pp. 1098-1104, Aug.2004.

[9] A. Lan, A. Nguyen, J.-N. Hwang, ``Scene Context Dependent Reference Frame Placement for MPEG Video Coding," IEEE Trans. Circuits and Systems for Video Technology, vol. 9, no. 3, 478-489, April 1990.

Table 2. Decision of intra coding of combined sequences of News_Akiyo and Football_Stefan by the proposed methods

Combined video(CIF)	Decision condition	Detected Intra Frame Location by Adaptive GOP Detection (AGD) or Scene Change Detection (SCD) Method					
News_Akiyo (300 + 300) frames	Intra Frame location	141 (AGD)	295(AGD)	300(SCD)	557(AGD)		
	Computed Detection Threshold	$m_d \geq 0$	$m_d \geq 0$	$\frac{SAMV_t+\varepsilon}{SAMV_{t-1}+\varepsilon}=0.1 \leq sp_1, \frac{SATD_t}{SATD_{t-1}}=8.76 \geq sp_2$	$m_d < 0$		
Football_Stefan (90 + 300) frames	Intra Frame location	34(AGD)	67(AGD)	90(SCD)	219(AGD)	284(AGD)	349(AGD)
	Computed Detection Threshold	$m_d \geq 5$	$m_d \geq 5$	$\frac{SAMV_t+\varepsilon}{SAMV_{t-1}+\varepsilon}=0.08 \leq sp_1, \frac{SATD_t}{SATD_{t-1}}=4.89 \geq sp_2$	$m_d \geq 0$	$m_d \geq 3$	$m_d \geq 3$