DYNAMIC GOP STRUCTURE DETERMINATION FOR REAL-TIME MPEG-4 ADVANCED SIMPLE PROFILE VIDEO ENCODER

Yu-Lin Wang, Jing-Xin Wang, Yen-Wen Lai, Alvin W.Y. Su

Studio of the Computer REseArch on the Music & Media(SCREAM)
National Cheng Kung University

ABSTRACT

MPEG-4 Advanced Simple Profile video provides I, P, and B –type frames in each GOP (Group Of Pictures). To maximize the coding efficiency, it is important to determine the distribution of the three frame types, also called the GOP structure. In this paper, the GOP structure is determined dynamically in real time by using the information, such as Sad (Sum of Absolute Difference) and Mad (Mean of Absolute Difference), generated in the encoding process and little amount of computation is used. The algorithm first determines whether the current frame is an Intra-picture or an Inter-picture. If it is an Inter-picture, one has to decide whether it is a P_picture or a B_picture. No more than 4 consecutive B_pictures can be employed to reduce the video buffer size. The proposed method is tested over a wide range of video sequences at different data rate conditions and produces consistent improvement compared to fixed GOP methods.

1. INTRODUCTION

MPEG-4 ASP (Advanced Simple Profile) video provides I, P, and B –type frames in each GOP. Usually, a fixed GOP structure is used in most MPEG-4 ASP video encoder and the coding efficiency is better than the MPEG-4 SP (Simple Profile) encoder. Since the content of any image sequence varies from time to time, the performance of a fixed GOP ASP encoder can be worse than its SP version, sometimes. Therefore, to determine a good GOP structure for a video sequence dynamically is necessary. Quite a few related studies have had fine results in this respect.

In [1] and [2], a relatively simple method was proposed only for MPEG-1 and MPEG-2 video encoders, respectively. In [3], the purpose of this paper was mainly for quality scalability instead of optimizing the GOP structure and required more computation to explore the temporal relation of consecutive frames. In [4], it provided a simple method to determine the GOP structure, but failed to produce constant improvement at conditions of various data rate conditions and picture complexities. In [5], an optimal GOP structure determination algorithm was proposed for MPEG-1 or MPEG-2 video. However, this method needs to use a whole GOP to find the optimal distribution which requires a huge memory buffer.

The method proposed in this paper is briefed as follows. In order to minimize the additional computation, three parameters are used. They are *Sad*, *Mad* and *Var_sad* (variance of *Sad*) calculated during the encoding process. A current frame and the next image are used for the decision making. We first decide whether the current frame is to be encoded in Intra mode or in Inter mode. If it is in Inter mode, we need to decide whether it is to be encoded as a P-picture or a B-picture. Once a P-picture or an I-picture is met, the encoder starts to encode this picture and then those previously un-coded frames as B-pictures. Simple decision rules are used in this paper. However, since the situations are quite different for low QP (Quantization Parameter) and high QP cases, the rules are also different in their parameters. The method is tested over various video sequences at different data rate conditions. It is found that it provides constant improvement compared to various fixed GOP methods.

In section 2, the proposed algorithm is described. Computer simulations are reported in section 3. Conclusion and future works are given in section 4.

2. ADAPTIVE GOP STRUCTURE DETERMINATION

The overall GOP structure determination flow is show in Fig.1 and Fig.2. To determine the type of the Current Image, *Mad*, Sad_1, and Var_sad_1 after the motion estimation operation is performed with respect to the reference frame are first calculated as follows.

$$Mad = \sum_{i=1, j=1}^{16,16} |current - MB\_mean| * (!(Alpha_{current} == 0)) \tag{1}$$

$$Sad^m = \sum_{i=1, j=1}^{16,16} |current - previous| * (!(Alpha_{current} == 0)) \tag{2}$$

$$Var\_sad = N_{MB} \sum_{m=1}^{N_{MB}} (Sad^m)^2 - (\sum_{m=1}^{N_{MB}} Sad^m)^2 \Big/ N_{MB}(N_{MB} - 1) \tag{3}$$

0-7803-9332-5/05/$20.00 ©2005 IEEE

$$Sad = \sum_{m=1}^{N_{MB}} Sad^{m} / N_{MB} \quad (4)$$

Sad^m: The Sad value of the m^{th} macro block.
N_{MB} : Total number of macro blocks in a frame.

If Intra mode is selected, we start to encode this frame as an I_picture and set it as the new reference frame. All previous B_pictures are encoded and output to the bitstream buffer, too. If Inter mode is selected, we perform motion estimation of a new image called the Next Image (shown in Fig.2) to obtain the corresponding parameters, Sad_2, and Var_sad_2. Then, we use the parameters to determine the type of the Current Image. If it is a P_picture, it is encoded and set to the new reference frame. All previous B_pictures are encoded and output to the bitstream buffer, too. Otherwise, the searching process continues until an I_picture or a P_picture is met. We next describe how the above decision making works.

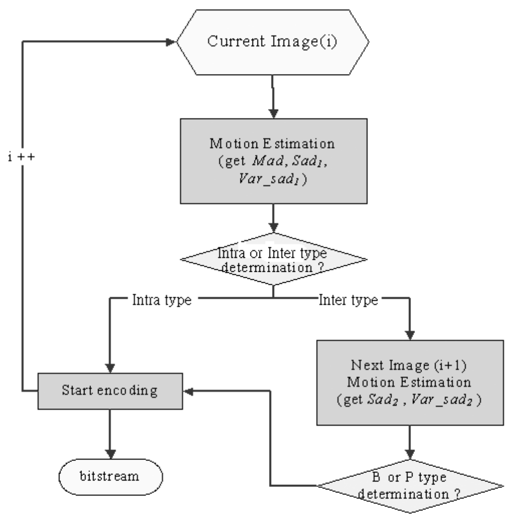

Fig. 1 Decision Flowchart

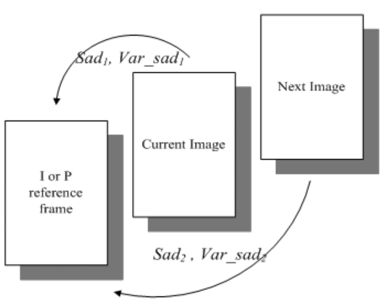

Fig.2 GOP Structure Decision Flow

2.1. Intra/Inter Determination

In general, an I_picture uses more bits than a P_picture. However, if the scene changes too rapidly, using Intra mode sometimes is better than using Inter mode. It is clear that detection of scene changes to determine which coding mode is to be used is vital. We focus on the relationship among *Mad*、 Sad_1 and Var_sad_1. Fig.3 shows a typical population of choosing between I_picture and P_picture based on two parameters. The X-axis uses the ratio of *Mad* and Sad_1, and the Y-axis uses Var_sad_1.

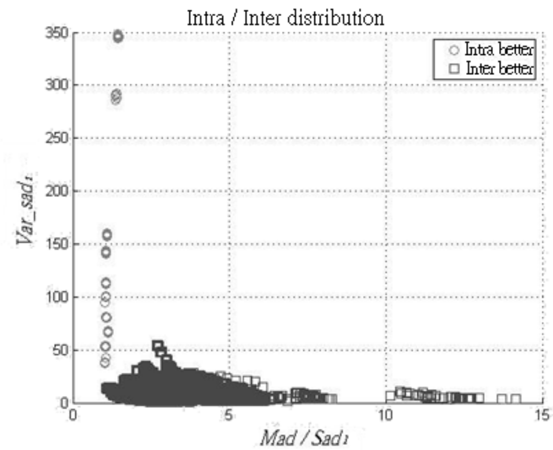

Fig. 3 Distribution of choosing Intra or Inter.

The density function for choosing the I-type is derived from Fig.3 and (5)-(8) as follows. The Gaussian Mixture Models (GMM) and Expectation Maximization (EM) method are used [6]. Statistics of both X-axis and Y-axis are considered separately by using χ as the argument. The results are shown in Fig.4.

$$f_I^{Mad\ /\ Sad_1}(\chi) = \sum_k \alpha_k G(\chi_{I_k}, \mu_{I_k}, \sigma)$$
$$f_P^{Mad\ /\ Sad_1}(\chi) = \sum_k \beta_k G(\chi_{P_k}, \mu_{P_k}, \sigma) \quad (5)$$

$$f_I^{Var\_sad_1}(\chi) = \sum_k \alpha_k G(\chi_{I_k}, \mu_{I_k}, \sigma)$$
$$f_P^{Var\_sad_1}(\chi) = \sum_k \beta_k G(\chi_{P_k}, \mu_{P_k}, \sigma) \quad (6)$$

$$f_{Mad\ /\ Sad\ 1}(\chi) = f_I^{Mad\ /\ Sad_1}(\chi) \big/ (f_I^{Mad\ /\ Sad_1}(\chi) + f_P^{Mad\ /\ Sad_1}(\chi)) \quad (7)$$
$$f_{Var\_sad\ 1}(\chi) = f_I^{Var\_sad_1}(\chi) \big/ (f_I^{Var\_sad_1}(\chi) + f_P^{Var\_sad_1}(\chi)) \quad (8)$$

$f_I^*(\chi)$: Estimated density given that Intra-type is better.
$f_P^*(\chi)$: Estimated density given that Inter-type is better.
$f_{Mad\ /Sad_1}(\chi)$: Estimated density of choosing the Intra-type corresponding to parameter, (Mad/Sad_1).
$f_{Var\_sad_1}(\chi)$: Estimated density of choosing the Intra-type corresponding to parameter, Var_sad_1.

2.1.1 Intra/Inter determination Procedure

If ($f_{Var\_sad_1}(\chi) > 80$){ Choose Intra_picture.}
Else if ($f_{Var\_sad_1}(\chi) < 30$){Choose Inter_picture.}

Else if ($f_{Mad/Sad_1}(\chi)$ >1.4){Choose Inter_picture.}
Else {Choose Intra_picture.}

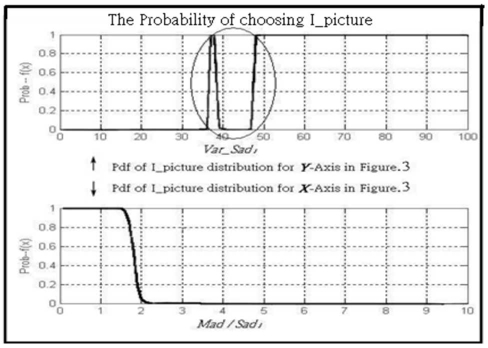

Fig.4 Two estimated densities of choosing Intra-type.

2.2. P/B Determination

More B_pictures should be inserted in a slowly changing scene generally and less number of B_pictures should be inserted in a fast changing one. However, inserting too many B_pictures also degrades the quality of pictures. For real-time encoding, the maximum number of B_pictures in a row is four. Four parameters are used. They are Sad_1 and Var_Sad_1 of the Current Image and Sad_2 and Var_sad_2 of the Next Image.

To obtain the statistics, we encoded many sequences and compare the data rates generated using different coding schemes. The following procedure is employed.

1. If the previous frame is an I_picture, we compared ***IPP*** and ***IBP*** coding sequence. If ***IPP*** uses less number of bits, P_type is used for the Current Image. Otherwise B_type is used.
2. If the previous frame is a P_picture, for example, we compared two difference schemes, $\boldsymbol{IB_{11}P_{12}P_{13}}$ and $\boldsymbol{IB_{11}P_{12}B_{21}P_{22}}$. If $\boldsymbol{P_{13}}$ uses less number of bits than the average of $\boldsymbol{B_{21}}$ and $\boldsymbol{P_{22}}$, P_type is used for the Current Image ($\boldsymbol{P_{13}}$). Otherwise B_type is used.
3. If there are N previous B_pictures, the scheme shown in Fig.5 is employed. Let B_k, B'_k, P and P' be the bits used for the respective frames. If $\frac{\sum_{k=1}^{N} B_k + P}{N+1} < \frac{\sum_{k=1}^{N+1} B'_k + P'}{N+2}$, P_type is used for the Current Image. Otherwise, B_type is used.

Fig.6 shows one of the statistics generated from the above process. Note that the figure is different for different QPs and different numbers of previous B_pictures. Similar to the previous the Intra-Inter determination method, the density function for choosing a B_picture is determined by using (9)-(12) The GMM and EM method are also used again [6].

$$\vec{x} = (x, y)^T \quad , \quad \vec{\mu} = (\mu_x, \mu_y)^T \tag{9}$$

$$f_B(\vec{x}) = \sum_{x \in I_B} \alpha_k G(\vec{x}, \vec{\mu}_{B_k}, \sigma) \tag{10}$$

$$f_P(\vec{x}) = \sum_{x \in I_P} \beta_k G(\vec{x}, \vec{\mu}_{P_k}, \sigma) \tag{11}$$

$$f(\vec{x}) = f_B(\vec{x}) \big/ (f_B(\vec{x}) + f_P(\vec{x})) \tag{12}$$

where I_B denotes the set where B-type is better for the data point in Fig.6. Similarly, I_P is the set for P-type.

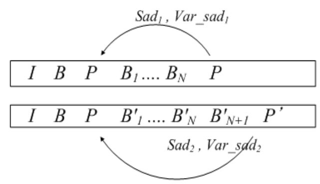

Fig.5 The scheme to collect data for analysis when the previous frame(s) is(are) B frame(s).

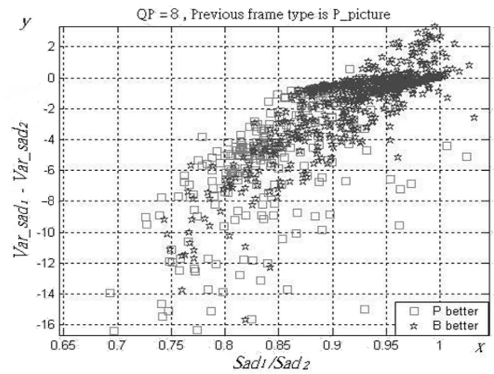

Fig.6 Distribution of choosing P or B when the previous frame is a P_picture and QP=8.

2.2.1. P/B determination procedure for QP=8

If ($f(\vec{x}) \geq 0.5$){Choose B_picture}
Else {Choose P_picture. }

3. COMPUTER SIMULATIONS

To evaluate our algorithm, two kinds of video sequences are used. The first kind contains two QCIF sequences, *[akiyo]* and *[foreman]*, both containing no scene change. The second kind contains two CIF sequences, *[table tennis]* and *[Lethal Weapon]*. Both contain possible scene changes. The *[Lethal Weapon]* sequence is extracted from a DVD using the Cyber-Link Power DVD® tool. The proposed method is compared to various fixed GOP methods at various bit rate conditions. For the fixed GOP methods, one to four B_pictures are inserted. As shown, our method gives consistent improvement over other methods.

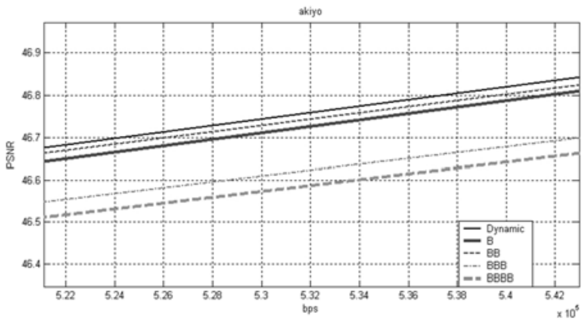

Figure.7 *[akiyo]*, 120 frames.

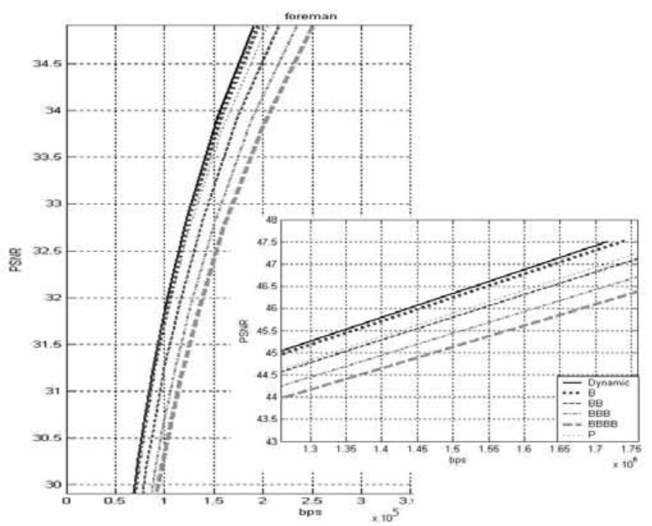

Fig.8 *[foreman],* 240 frames

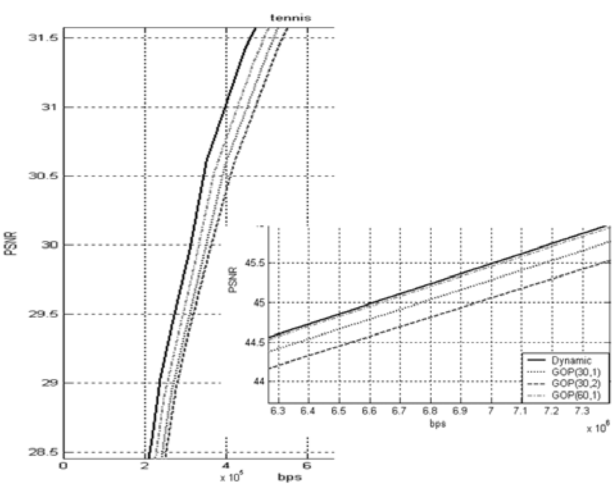

Fig.9 *[table tennis] ,* 110 frames

4. CONCLUSIONS AND FUTURE WORK

The method proposed in this paper has better performance than various static GOP encoding schemes at different bit rate conditions. No complicate computation is required. Thus, it is suitable for real time encoding and streaming applications.

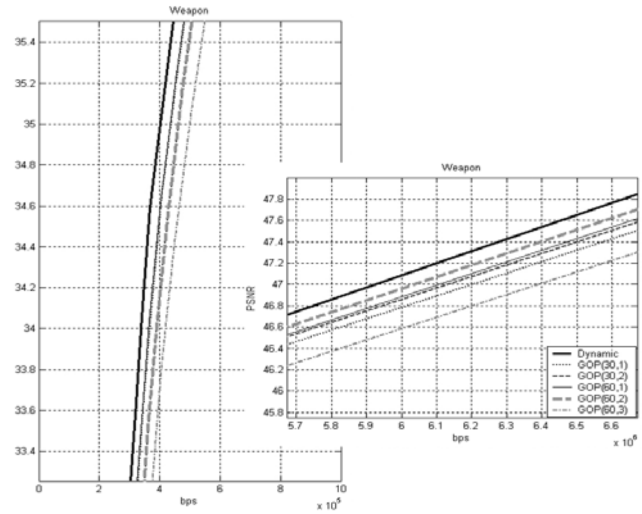

Fig.10 *[Lethal-Weapon]*, 270 frames

In the future, we will involve the motion vector information to make the determination more precisely. On the other hand, it is possible to research a backward process to re-determine a frame which has been determined as a B-picture if we find that the performance decreases because too many B_pictures are inserted. However, computation complexity will be increased inevitably. We will also compare the proposed method with the result produced with a truly optimal GOP structure.

5. REFERENCES

[1] Yutaka Yokoyama, "Adaptive GOP Structure Selection for Real-Time MPEG-2 Video Encoding," *Proceeding of IEEE International Conference on Image Processing*, vol. 2, pp 832-835, Sept. 2000.

[2] Akio Yoneyama, *et.al.*, "One-pass VBR MPEG encoder using scene adaptive dynamic GOP structure", *ICCE2001*, pp 174-175, June 2001.

[3] Xiaodong Gu, and Hongjian Zhang, "Implementing dynamic GOP in video encoding," *Proceedings of IEEE International Conference on Multimedia and Expo*, vol. 1, pp.349-352, July 2003.

[4] Akio Yoneyama, *et.al.*, "MPEG Encoding Algorithm with Scene Adaptive Dynamic GOP Structure," *IEEE 3rd Workshop on Multimedia Signal Processing*, pp 297-302, Sept. 1999.

[5] Jungwoo Lee, *et.al*, "Rate-Distortion Optimization Frame Type Selection for MPEG Encoding," *IEEE Trans. on Circuits and System for Video Technology*, vol. 7, No 3, pp.501-510, June 1997.

[6] David W. Scott, *Multivariate Density Estimation,* A Wiley-Interscience Publication, Feb. 1997.