International Journal of Information, Control and Computer Sciences
ISSN: 2517-9942
Vol:7, No:7, 2013

# Adaptive Group of Pictures Structure Based On the Positions of Video Cuts

Lenka Krulikovská, Jaroslav Polec, and Michal Martinovič

***Abstract*****—**In this paper we propose a method which improves the efficiency of video coding. Our method combines an adaptive GOP (group of pictures) structure and the shot cut detection. We have analyzed different approaches for shot cut detection with aim to choose the most appropriate one. The next step is to situate $N$ frames to the positions of detected cuts during the process of video encoding. Finally the efficiency of the proposed method is confirmed by simulations and the obtained results are compared with fixed GOP structures of sizes 4, 8, 12, 16, 32, 64, 128 and GOP structure with length of entire video. Proposed method achieved the gain in bit rate from 0.37% to 50.59%, while providing PSNR (Peak Signal-to-Noise Ratio) gain from 1.33% to 0.26% in comparison to simulated fixed GOP structures.

***Keywords*****—**Adaptive GOP structure, video coding, video content, shot cut detection.

## I. Introduction

IN the multimedia compression technology and computer performance is recorded a progress. It has led us to the widespread availability of digital video. The efficiency of video encoding became very important task, due to the need of flexibly delivering multimedia data to users with different available interests, access networks and resources.

The newest and the most efficient video coding standard is the H.264/AVC [1], [2]. New features of H.264 include motion estimation in variable block sizes, multiple reference frame motion compensation, spatial prediction for intra coding, small block size residual transform coding, adaptive and hierarchical block size transform, etc. [2]. The encoders mostly used fixed group of pictures (GOP) size to encode video sequences. The GOP size can achieve different values, but once target size for GOP is selected, it is applied to whole coded sequence.

While fixed GOP structures are easy to implement, they prevent encoders from adapting to temporal variations in video sequences and thus prevent encoders from improving coding efficiency by selecting the frame type of each frame adaptively. The transitions between shots are the region, where static GOP structures achieved poor performance. Generally, if the video frames with smaller video content variance are coded as intra frames, we will waste a lot of bits in video coding. Conversely, if two shots changed and frames are coded using inter frames, it will also become inefficient. This can be solved by using adaptive GOP structure with positioning $N$ frames to the places of shot changes.

Adaptive GOP structure (AGS) is a new technique that can be used for enhancing the coding performance of the scalable extension of H.264/AVC [3], [4]. The AGS scheme adaptively changes the sizes of GOP structure according to the temporal characteristics of a video sequence to improve the coding efficiency.

In this paper we propose a novel method of GOP structure adaptable to the positions of shot transitions. This approach is based on shot cut detection and subsequently the size of GOP structure is adapt to the video content by placing $N$ frames to detected abrupt cut. The proposed method was evaluated through experiments and obtained results were compared with selected sizes of fixed GOP structure.

The paper is structured as follows: in the second section is described a proposed method of adaptive GOP structure. Results obtained by the simulations for adaptive and fixed GOP structures are displayed in the third section. In conclusion we summarize and discuss all results.

## II. Theoretical Basis and Model

Our proposal of GOP structure is based on two principles – the detection of shot transitions and applying $N$ frames to the positions of detected cuts.

### *A. Detection of Video Cuts*

The majority of shot transitions used in video are abrupt cuts in general, because more than 99% of all transitions found in video [5]. It is the reason why we focused only on the detection of an abrupt cut.

Various approaches have been proposed to extract shots. The same basic scheme is used for the most of published methods– two consecutive frames are compared and evaluated by selected metrics for chosen technique. As the next step, the fixed or adaptive threshold is applied to determine the occurrence of shot cut [6].

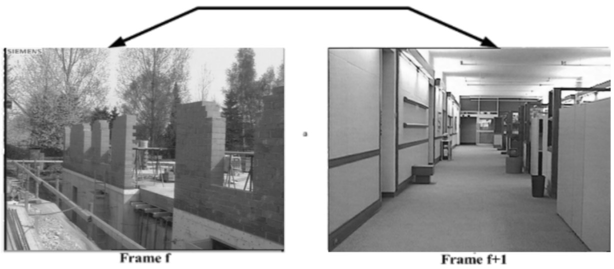
For the detection of abrupt cuts we have used a method described in [7]. The novelty of this method is in comparing frame with its motion estimated prediction instead of pair of successive frames. This logic of comparison in combination

L. Krulikovská works for AT&T Slovakia and in 2004–2012 was with the Institute of Telecommunication, Slovak University of Technology, Ilkovičova 3, 812 12, Bratislava, Slovak Republic.

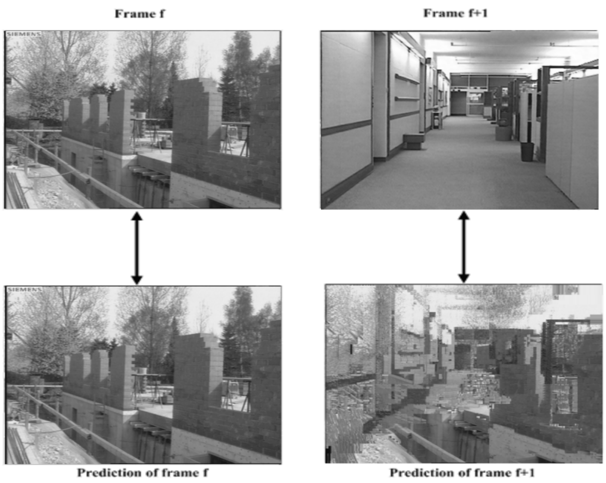
J. Polec is with the Institute of Telecommunication, Slovak University of Technology, Ilkovičova 3, 812 12, Bratislava, Slovak Republic (phone: +421268279409, e-mail: polec@ktl.elf.stuba.sk).

M. Martinovič is with the Institute of Telecommunication, Slovak University of Technology, Ilkovičova 3, 812 12, Bratislava, Slovak Republic (e-mail: martinovic@ut.fei.stuba.sk).

with adaptive threshold lead to significant improvement of the efficiency and accuracy in shot cut detection. Fig. 1 illustrates us the principle of comparison for the majority of shot cut detection methods and for the method in [7]. The arrows indicate which frames are compared.




(a)




(b)

Fig. 1 The principle of comparison in shot cut detection process for the (a) majority of published methods and (b) selected method [7]

*B. Adaptive GOP Structure*

We place $N$ frames to the positions of abrupt cuts as we mentioned before. Thus we adapt the size of GOP structure to the content of video sequences.

According to positioning frame types between two consecutive $N$ frames, $M$ and $B$ frames are most commonly used. We have made decision to use only $M$ frames for simulation, because $B$ frames are computationally high demanding and can caused unwanted delays.

An example of proposed and fixed GOP structure is shown in Fig. 2. There are two abrupt cuts. Namely: between third and fourth frame and between sixth and seventh frame.

Fixed GOP structure selects $M$ frames for encoding of shot transitions. This decision causes the degradation of coding efficiency and increase the number of bits used for encoding, because $M$ frame will contains mainly $N$ macro blocks. Proposed GOP structure adapts to video content and place $N$ frames to the positions of cuts.

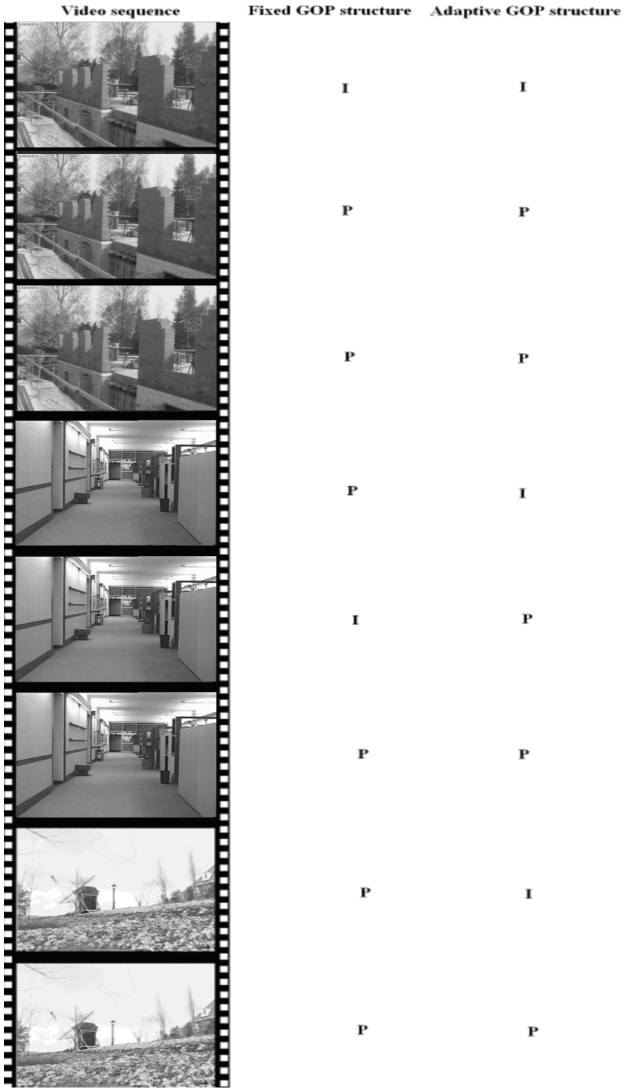



Fig. 2 Example of fixed GOP structure (size 4) and proposed adaptive GOP structure

## III. Results of Experiments

We confirmed the effectiveness of proposed method through a test experiment. For test purposes we created a video sequence with 1989 frames at CIF resolution (352 x 288 pixels) with 7 abrupt cuts sampled at rate of 30 frames per second. The test video sequence consists of eight standard test sequences: akyio, foreman, hall, flower, mobile, mother-daughter, stephan and bus.

Obtained results for proposed GOP structure were compared with the situation, when whole sequence is coded only with one $N$ frame at the beginning and the rest are $M$ frames (fixed GOP structure IPPP with the length of entire video sequence). With effort to provide similar comparison as in [8], adaptive GOP structure was additionally compared with fixed GOP structures of size 4, 8, 12, 16, 32, 64 and 128 (it means $N$ frame is followed by size-1 $M$ frames). Following five conditions of H.264/AVC encoder were performed for each simulation:

- QP (quantization parameter): 28
- Reference for $M$ slices: 1
- Total number of reference: 1
- Search range: 16
- Entropy coding method: CABAC

The results obtained during simulations are shown in Table I.

The highest PSNR and the smallest bit rate were achieved by proposed method of adaptive GOP structure. The improvement is more significant for bit rate.

TABLE I
COMPARISON OF PSNR AND BIT RATE ACHIEVED BY FIXED AND ADAPTIVE GOP STRUCTURE

| GOP | PSNR [dB] | Bit rate [kb/s] |
|---|---|---|
| 4 | 37,9 | 977,51 |
| 8 | 37,3 | 729,89 |
| 12 | 37,7 | 648,5 |
| 16 | 37,6 | 606,29 |
| 32 | 37,6 | 546,35 |
| 64 | 37,5 | 515,06 |
| 128 | 37,5 | 501,15 |
| IPPP | 37,5 | 484,83 |
| Adaptive | 38 | 483,03 |

Table II shows the gain of adaptive GOP structure in comparison to fixed GOP structures. Proposed method achieved a bit rate reduction from 0.37% to 50.59%, while providing PSNR gain from 1.33% to 0.26%.

We can see that the highest bit rate reduction was achieved in comparison with fixed GOP structure of size 4. This case illustrates that if we select small size of GOP structure, we will force the encoder to use a lot of $N$ frames and also we increase the bit rate.

In the case of selected GOP with the size same as the length of entire video sequence (IPPP); we have forced encoder to use only one $N$ frame for whole sequence. Despite this fact proposed method achieved 0.37% bit rate reduction. The bit rate of fixed GOP structure was increased due to large amount of $N$ macro blocks used in $M$ frames in positions of abrupt cuts. 0.37% reduction is not so high; however it gives an assumption that the reduction becomes more significant for video sequences with more shot transitions.

TABLE II
PSNR AND BIT RATE GAIN ACHIEVED BY ADAPTIVE GOP STRUCTURE IN COMPARISON TO FIXED GOP STRUCTURES

| GOP | PSNR gain [%] | Bit rate gain [%] |
|---|---|---|
| 4 | 0,26 | 50,59 |
| 8 | 0,8 | 33,82 |
| 12 | 0,8 | 25,52 |
| 16 | 1,06 | 20,33 |
| 32 | 1,06 | 11,59 |
| 64 | 1,33 | 6,22 |
| 128 | 1,33 | 3,62 |
| IPPP | 1,33 | 0,37 |

Table III shows the difference in bit usage for $N$ and $M$ frames at the places of cuts for the same PSNR. With effort to provide as close PSNR as possible for adaptive and fixed GOP structure, we have used different quantization parameter for each GOP structure. The number of bits needed for encoding was higher if we used $M$ frames in detected cuts for each cut. This was caused by a large number of used Intra macro blocks as it can be seen from third column for fixed GOP structure.

TABLE III
THE DIFFERENCE IN BIT USAGE FOR I AND P FRAMES AT THE PLACES OF CUTS

| Cut | *Adaptive GOP, QP 26* $M$ *frame placed to position of cut* | | | *Adaptive GOP, QP 28* $N$ *frame placed to position of cut* | |
|---|---|---|---|---|---|
| | PSNR [dB] | Used bits | Number of Intra MB | PSNR [dB] | Used bits |
| 1 | 39 | 52328 | 396 | 39 | 49920 |
| 2 | 39,8 | 52000 | 395 | 39,9 | 48184 |
| 3 | 36,4 | 154496 | 396 | 36,6 | 145960 |
| 4 | 35,3 | 184496 | 390 | 35,3 | 183048 |
| 5 | 40,7 | 34256 | 396 | 40,5 | 33224 |
| 6 | 37 | 112464 | 396 | 37,1 | 112032 |
| 7 | 36,1 | 109928 | 385 | 36,2 | 109032 |

## IV. CONCLUSION

In this paper we present an adaptive GOP method for improving coding efficiency of H.264/AVC. The novelty of proposed method is in the using of method for shot cut detection with new logic of frames comparison and in adapting GOP structure to the video characteristics. Our method places $N$ frames to the positions of detected abrupt cuts, the rest of sequence is encoded as $M$ frames.

We make experimental tests and we prove correctness and efficiency of our approach in a comparison to fixed GOP structures of sizes 4, 8, 12, 16, 32, 64, 128 and length of the whole video sequence. Bit rate reduction is obtained by method of adaptive GOP differs from 0.37% (fixed GOP structure with length of video sequence) to 50.59% (fixed GOP structure of size 4), while providing PSNR gain from 1.33% to 0.26% at the same time.

0.37% reduction of bit rate in the case of fixed GOP structure, where only one $N$ frame is used for whole coded video sequence, shows the $M$ frames placed to the shot transition will increase the bit rate.

When we compare our method with solution in [8] we achieve higher bit rate reduction for the same GOP sizes 4, 8, 16 and 32. Furthermore our proposed method is not degrading PSNR.

The another important advantage of using proposed adaptive GOP structure, in addition to improvement of video coding efficiency while providing the same quality, is the simplification of later video segmentation. There is no need to run shot cut detection process again, because cuts can be identified in positions of $N$ frames.

In further research, we would like to examine the influence of various quantization parameters and of the composition of

GOP structure on the video coding efficiency and video quality.

## ACKNOWLEDGMENT

Research described in the paper was financially supported by the Slovak Research Grant Agency VEGA under grant No. 1/0602/11.

## 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] JVT Editors, Draft ITUT Recommendation and final draft international standard of joint video specification (ITU-T Rec.H.264 |ISO/IEC 14496-10 AVC), JVT-G050r1, Geneva, May 2003.

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

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

[5] S. Paschalakis and D. Simmons, (2008, April 24), "Detection of gradual transitions in video sequences" [Online]. Available: http://www.wipo.int/pctdb/en/wo.jsp?WO=2008046748&IA=EP2007060594&DISPLAY=STATUS

[6] A. Hanjalic, "Shot-boundary detection: unraveled and resolved?," IEEE Transactions on Circuits and Systems for Video Technology, vol. 12, no. 2, pp. 90–105, 2002.

[7] L. Krulikovska and J. Polec, "A novel method of comparison frames for video cuts detection" Proceeding of Redžúr 2011 5th International Workshop on Speech and Signal Processing, in press.

[8] B. Zatt, M. Porto, J. Scharcanski and S. Bampi, "Gop structure adaptive to the video content for efficient H.264/AVC encoding," Proceedings of 2010 IEEE 17th International Conference on Image Processing, pp. 3053-3056, September 2010.

**Lenka Krulikovská** received her Bc. (bachelor degree), Ing. (MSc.) and PhD. degrees in Telecommunication from the Slovak University of Technology (SUT) in Bratislava in 2007, 2009, 2012. Her research interests were concentrated on Unequal error control coding for telemetric information and image transmission. From 2012 until the presence, she has worked for AT&T Slovakia.

**Jaroslav Polec** received the Engineer and PhD. degrees in Telecommunication engineering from the Faculty of Electrical Engineering and Information Technology, Slovak University of Technology in 1987 and 1994, respectively. Since 1997 he has been associate professor and since 2007 professor at the Institute of Telecommunications of the Faculty of Electrical Engineering and Information Technology, Slovak University of Technology and since 1998 at the Department of Applied Informatics, Faculty of Mathematics, Physics and Informatics of the Comenius University. His research interests include Automatic-Repeat-Request (ARQ), channel modeling, image coding, interpolation and filtering.

**Michal Martinovič** received his Bc. (bachelor degree) and Ing. (MSc.) in Telecommunication from the Slovak University of Technology (SUT) in Bratislava, in 2010 and 2012, respectively. From 2012 until the presence, he has been an internal PhD. student at the Institute of Telecommunication, (SUT). His research interests are concentrated on Automatic-Repeat-Request (ARQ), Error control coding for wireless communications and also signal processing. From 2011 until the presence, he has been with the Intech Control Spol. S R. O.