Video abrupt cut detection in H.264 compressed domain

Máťuš Tomáš · Elektrotechnika, Informačné technológie

02.01.2012

In this paper we propose a method to detect abrupt cut changes in H.264 coded video that operates directly in the compressed domain. The proposed algorithm is fast and simple and it is suitable for the real-time implementation. This method is based on monitoring the number of I macroblocks in frames P and B. The ability to find cuts using this method for different GoP structures was analyzed in the experiment. Analysis was focused on sensitivity of this method to various evaluation thresholds that determine if the cut was occured. The evaluation was performed on the base of three metrics: precision, recall and F-measure.

1. Introduction

Video coding is a very important computational application, mainly due to the growth of video manipulation by many electronic devices as mobile phones, PDAs, digital television, etc. The H.264/AVC is currently the most efficient video coding standard [1]. The available H.264 software uses three frame types for the video coding: I (Intra), P (Predictive) and B frame (Bi-Predictive). Intra frames are coded using only the intra frame prediction and they are used as references for the P frame and B frame prediction. P frames use only past frames as a reference. B frames use the past and the future frames as a reference.

The efficiency of the H.264 coding depends directly on the GoP size and on its internal structure. Most available H.264 encoders use a static size for the GoP to encode video sequences. The GoP size can assume different values, however, after a given GoP size is selected, the whole coding process uses the same size. The frames I, P and B are distributed statically inside the GoP. Unfortunately, encoders known to support adaptive GoP structure, such as the Main Concept codec, often do not disclose the algorithms used to define the GOP size, or the I, P and B frames distribution [2].

The way how to find the exact video is a very important field of the research. The detection of shot boundaries provides a base for nearly all video abstraction and high-level video segmentation approaches. Therefore, solving the problem of shot-boundary detection is one of the major prerequisites for revealing higher level video content structure. Moreover, other research areas can profit considerably from successful automation of shot-boundary detection processes as well.

There is a number of different types of transitions or boundaries between shots. Two

shots can be combined by an abrupt cut or a gradual transition. The abrupt cut is an instantaneous transition of the content from one shot to another without any special effects. It is a direct change of shots. The gradual transition is a transition where special effects are used to artificially combine two shots. The change of a shot takes more frames [3].

Different approaches have been proposed to extract shots. The major techniques used for the shot boundary detection are pixel differences, statistical differences, histogram comparisons [4], edge differences, compression differences and motion vectors [5, 6, 7].
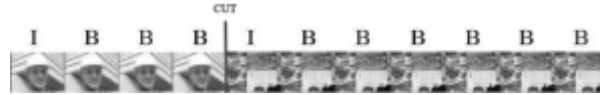
Figure 1: An example of a video cut.

2. Evaluation Techniques

To compare various GoP structures, we have chosen three evaluation techniques: precision, recall and F-measure.

2.1 Precision Measure

The Precision measure is defined as the ratio of correct video cut detections over the number of all video cut detections [3].

$$\text{Precision} = \frac{|Det \cap GT|}{|Det|} \tag{1}$$

GT: denote the correct cut detection. Det: denote the all detected (correct and false) cuts.

2.2 Recall Measure

The Recall measure is defined as the ratio of correct video cut detections over the number of correct video cut detections [3].

$$\text{Recall} = \frac{|Det \cap GT|}{|GT|} \tag{2}$$

2.3 F – Measure

The F – Measure combines precision and recall and is defined as the two times ratio of precision times recall over precision plus recall [8].

$$F = 2\frac{Recall.Precision}{Recall+Precision} \tag{3}$$

3. GoP Structures

We have chosen three GoP structures: IPPPP, IPBPB and IPBBP structure. In each structure we have tried to find cuts and detect, if the cut was detected correctly. For

each structure we have made a table with a count of all and right cuts. On the base of three metrics we have evaluated, which GoP structure is the best for H.264 video compressed domain. H.264 video has 1989 frames and 7 abrupt cuts.
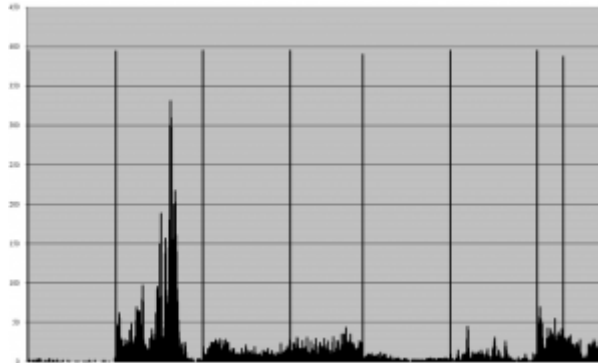
3.1 IPBPB structure

Figure 2: Graph of all detected video cuts in IPBPB structure.

Table 1: Number of all (Det) and correct (GT) cuts and calculated value for Recall (R), Precision (P) and F-measure (F) for the various count of Macroblocks (M)

M	GT	Det	R	P	F
0	7	1989	1	0,004	0,007
50	7	63	1	0,111	0,200
100	7	27	1	0,259	0,412
150	7	21	1	0,333	0,500
200	7	14	1	0,500	0,667
250	7	11	1	0,636	0,778
300	7	10	1	0,700	0,824
350	7	7	1	1,000	1,000
400	0	0	0	0,000	0,000
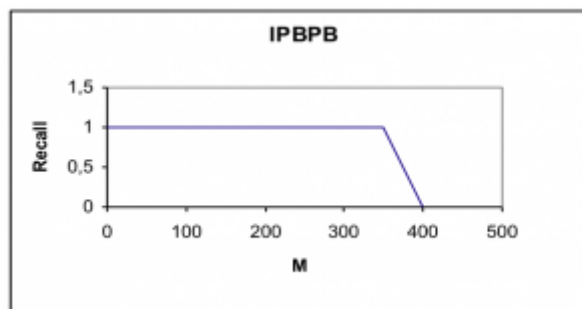
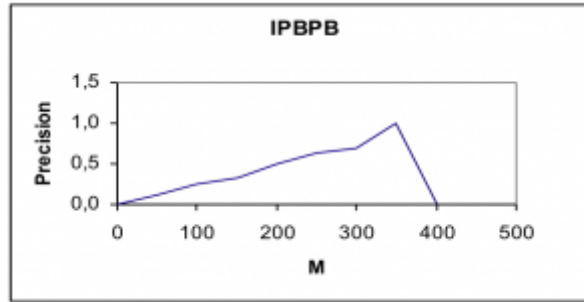*Figure 3: Graph of the Recall measure depending on the count of Macroblocks (M).*

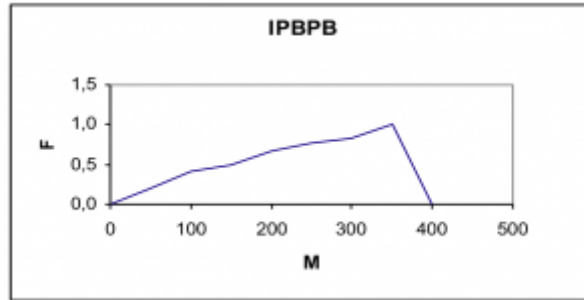*Figure 4: Graph of the Precision measure depending on the count of Macroblocks (M).*

Figure 5: Graph of the F-measure depending on the count of Macroblocks (M).
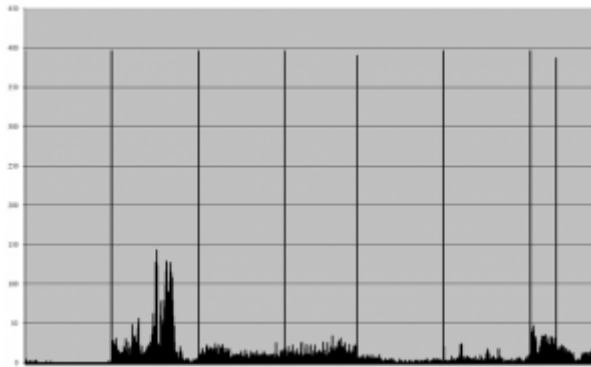
3.2 IPPPP structure

Figure 6: Graph of all detected video cuts in IPPPP structure.

Table 2: Values for the IPPPP structure.

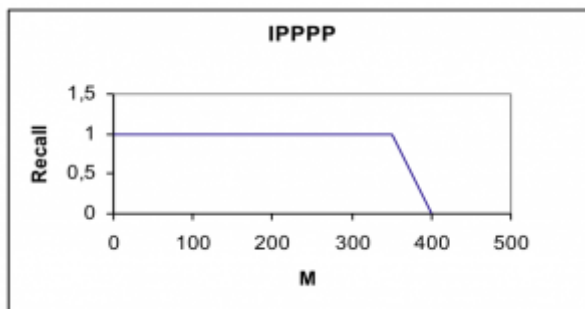| M	GT	Det	R	P	F
0	7	1989	1	0,004	0,007
50	7	56	1	0,125	0,222
100	7	23	1	0,304	0,467
150	7	7	1	1,000	1,000
200	7	7	1	1,000	1,000
250	7	7	1	1,000	1,000
300	7	7	1	1,000	1,000
350	7	7	1	1,000	1,000
400	0	0	0	0,000	0,000

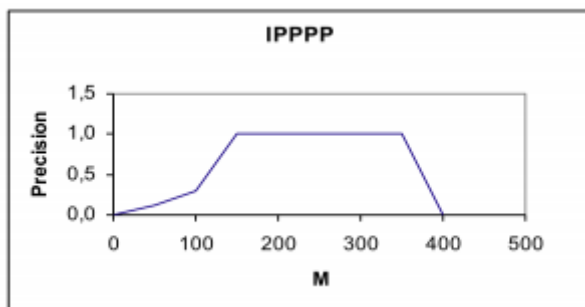*Figure 7: Graph of the Recall measure for the IPPPP structure.*

Figure 8: Graph of the Precision measure for the IPPPP structure.
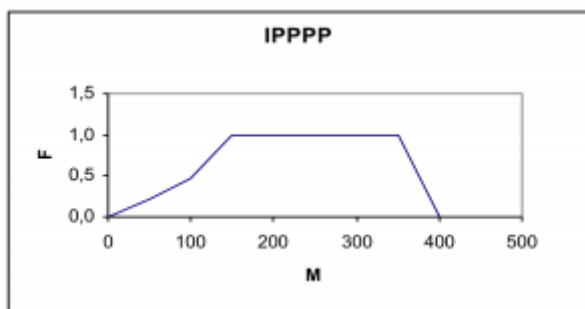

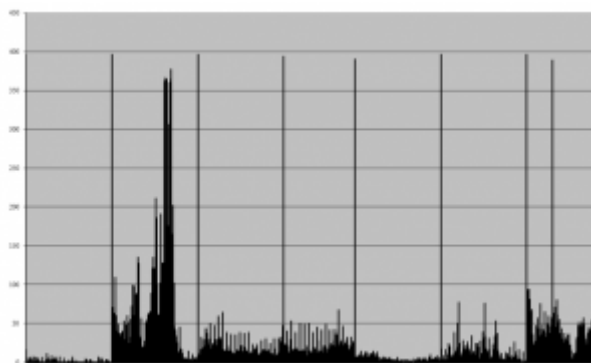*Figure 9: Graph of the F-measure for the IPPPP structure.*

3.3 IPBBP structure

Figure 10: Graph of all detected video cuts in IPBBP structure.

Table 3: Values for the IPBBP structure

M	GT	Det	R	P	F
0	7	1989	1	0,004	0,007
50	7	103	1	0,068	0,127
100	7	34	1	0,206	0,341

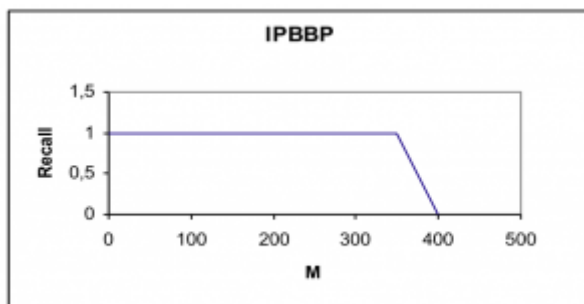| 150	7	21	1	0,333	0,500
200	7	17	1	0,412	0,583
250	7	15	1	0,467	0,636
300	7	15	1	0,467	0,636
350	7	14	1	0,500	0,667
400	0	0	0	0,000	0,000

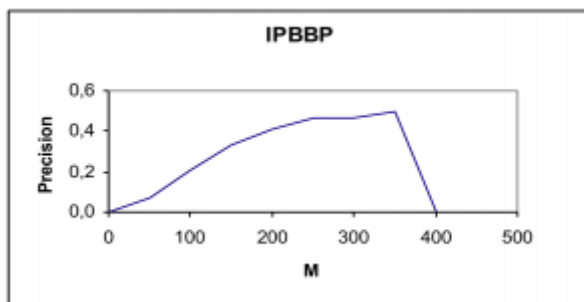*Figure 11: Graph of the Recall measure for the IPBBP structure.*

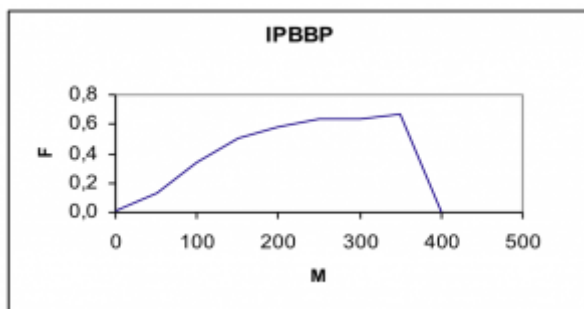*Figure 12: Graph of the Precision measure for the IPBBP structure.*

Figure 13: Graph of the F-measure for the IPBBP structure.
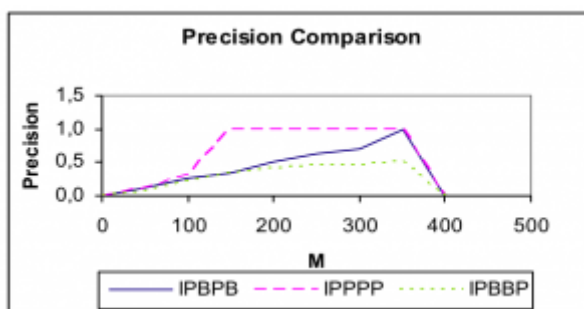
3.4 Structures Comparison

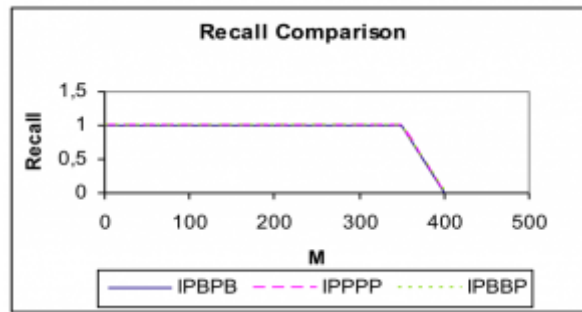*Figure 14: Graph of the Precision Comparison.*
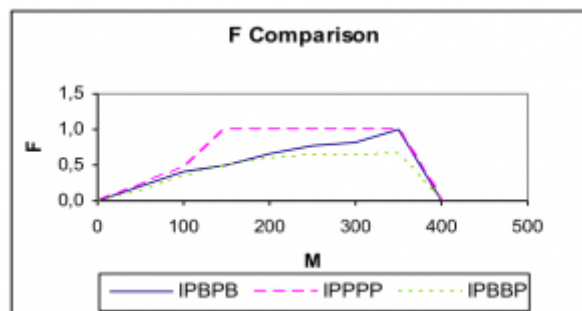
Figure 15: Graph of the Recall Comparison.

Figure 16: Graph of the F-measure Comparison.

4. Conclusion

We compared three various GoP structures for three different measure techniques. The best GoP structure in H.264 video compressed domain is IPPPP structure, where over the 150 macroblocks were only correct cuts and precision, recall and F-measure equated to 1. The worst GoP structure is IPBBP structure, where over the 350 macroblocks were either occurred false cut detections.

Precision value was smaller or equal to 0,5 and F-measure value was smaller or equal to 0,68. In IPBPB structure were false cuts occured up to 350 macroblocks. We can see detailed structure comparison in the attached graphs and tables.

Acknowledgements

References

1. JVT Editors (T. Wiegand, G. Sullivan, A. Luthra), 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.
2. B. Zatt, M. Porto, J. Scharcanski, S. Bampi, GoP Structure Adaptive to the Video Content for Efficient H.264/AVC Encoding, Hong Kong September 2010.
3. Z. Černeková, Temporal video segmentation and video summarization, (PhD Thesis), Bratislava 2009.
4. A. Amiri and M. Fathy, "Video shot boundary detection using QR-decomposition and Gaussian transition detection", EURASIP Journal on Advances in Signal Processing, Volume 2009, Article ID 509438.
5. 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.

6. J. S. Boreczky and L. A. Rowe, "Comparison of video shot boundary detection techniques," in Storage and Retrieval for Still Image and Video Databases IV, Proc. SPIE 2664, pp. 170-179, Jan. 1996.

7. R. Lienhart, "Comparison of automatic shot boundary detection algorithms," in Storage and Retrieval for Image and Video Databases VII, vol. 3656 of Proceedings of SPIE, pp. 290–301, San Jose, Ca, USA, January 1999.

8. Steven M. Beitzel, On understanding and classifying web queries (PhD Thesis), 2006.

Coauthor of this paper is prof. Ing. Jaroslav Polec, PhD., Dept. of Telecommunications, Faculty of Electrical Engineering and Information Technology, Slovak University of Technology