Fast H.264/AVC-to-SVC Transcoding in a Mobile Television Environment

Glenn Van Wallendael[1], Sebastiaan Van Leuven[1], Rosario Garrido-Cantos[2], Jan De Cock[1], José Luis Martínez[2], Peter Lambert[1], Pedro Cuenca[2], and Rik Van de Walle[1]

[1] Ghent University - IBBT
Department of Electronics and Information Systems - Multimedia Lab
Gaston Crommenlaan 8 bus 201, B-9050 Ledeberg-Ghent, Belgium
{glenn.vanwallendael,sebastiaan.vanleuven,jan.decock,
peter.lambert,rik.vandewalle}@ugent.be
[2] Albacete Research Institute of Informatics
University of Castilla-La Mancha
02071 Albacete, Spain
{charo,joseluismm,pcuenca}@dsi.uclm.es

Abstract. Mobile TV environments demand flexible video compression like Scalable Video Coding (SVC) because of varying bandwidths and devices. Since existing infrastructures highly rely on H.264/AVC video compression, network providers could adapt the current H.264/AVC encoded video to SVC. This adaptation needs to be done efficiently to reduce processing power and operational cost. Since a cascaded decoder-encoder solution is too complex to be practical, we developed a mechanism to encode scalable video streams from existing H.264/AVC encoded video streams. This paper proposes a novel technique to accelerate the encoding of SVC streams by reusing information from the H.264/AVC stream and the base layer. We achieved a complexity reduction of 52%, while only an insignificant bit rate increase is reported (0.2%). According to these results, an H.264/AVC-to-SVC transcoder is usable with a low operational cost without compromising the coding efficiency.

Keywords: H.264/AVC-to-SVC transcoding, complexity reduction, fast mode decision, cascaded pixel-domain transcoding.

1 Introduction

A mobile television environment is characterized by fluctuating bandwidths and varying device capabilities. Because of these irregularities, it is necessary to adapt the video stream to the changing environment. Since this adaptability is not incorporated in H.264/AVC, Scalable Video Coding (SVC) [1] was introduced. The SVC video stream is divided into layers, each adding more spatial, temporal or quality resolution. By removing packets from these layers, spatial resolution, frame rate, or quality can be reduced. A distribution of the video stream with SVC therefore provides a low complexity solution for the adaptability problem.

J. Rodriguez, R. Tafazolli, C. Verikoukis (Eds.): MOBIMEDIA 2010, LNICST 77, pp. 607–618, 2012.

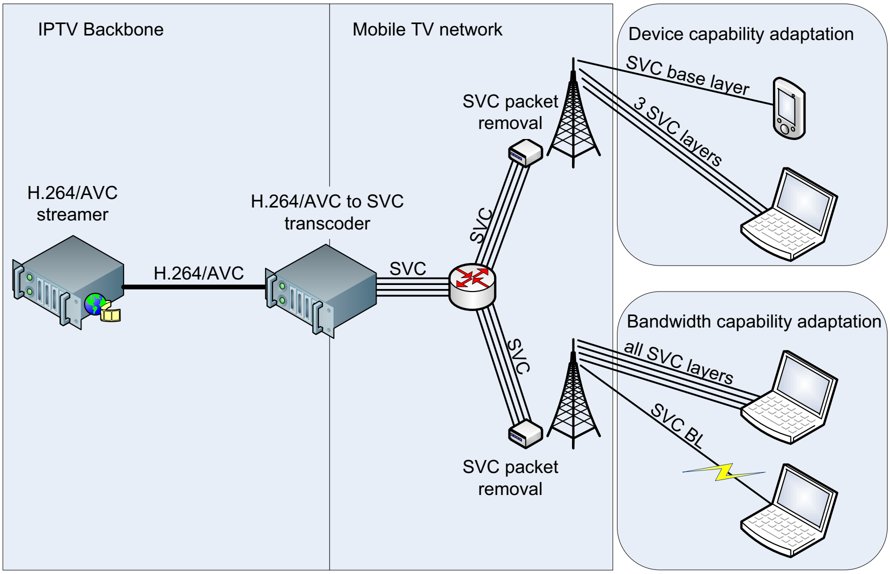
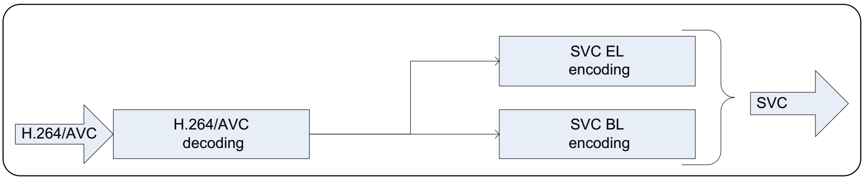


Fig. 1. Proposed mobile TV architecture

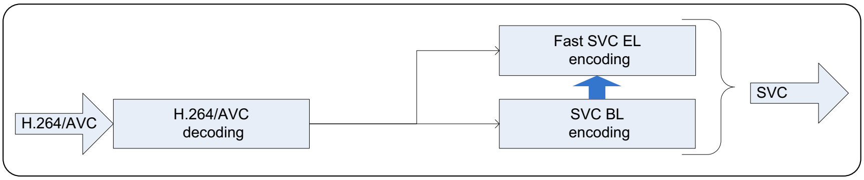
Nowadays, broadcasting for TV and mobile TV is largely based on H.264/AVC. As a result, to extend existing infrastructure with scalable capabilities, H.264 /AVC-to-SVC transcoding is needed. When applying this transcoding step at the broadcaster's premises, existing H.264/AVC infrastructure [2] can be maintained and low complexity adaptations can be made in the broadcast network, where needed. The resulting network architecture is illustrated in Fig. 1. In this figure, an H.264/AVC-to-SVC transcoder is added on the transition from the IPTV backbone to the mobile TV network. Also low complexity nodes for SVC packet removal are included in the mobile TV network to adapt the SVC video stream to the device and network characteristics.

Transcoding comes in two flavors, frequency domain transcoding and spatial domain transcoding [3]. With frequency domain transcoding, the H.264/AVC input video stream is not fully decoded [4]. The transform coefficients obtained after entropy decoding and scaling are utilized for the encoded SVC video stream. This results in a very fast transcoding paradigm, but the quality degradation of the process is noticeable. Because the focus in this work lies on distribution of television, a higher quality constraint is necessary. Spatial domain transcoding, or Cascaded Pixel-Domain Transcoding (CPDT), offers better quality and is therefore preferred in this paper. The downside of this transcoding paradigm is that it is slower compared to frequency domain transcoding.

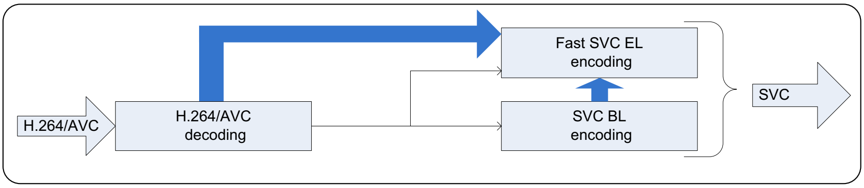
When no acceleration is applied in the CPDT encoding step, the process is called recoding instead of transcoding. Recoding the video stream by decoding



(a) Recoding configuration.



(b) Transcoding with SVC BL information.



(c) Transcoding with SVC BL and H.264/AVC information.

Fig. 2. Different transcoding architectures

the H.264/AVC video followed by encoding to SVC is the slowest possible process, but the resulting quality is minimally degraded. In Fig. 2(a), the recoding architecture is visualized schematically.

In a CPDT process, execution time is largely spent on the encoding steps. As such, acceleration of these steps would impact the overall performance the most. In the encoding process itself, execution time is divided in 1/3 for the base layer (BL) and 2/3 for the enhancement layer (EL) due to inter-layer prediction in the EL. Therefore, optimizing the enhancement layer results in the largest performance gain.

We propose an acceleration that can be achieved with two sources of information. First, information from the SVC base layer, or lower quality layer, will facilitate an acceleration in the enhancement layer encoding process [5] [6] (illustrated in Fig. 2(b)).

The second source of information that can be used is the H.264/AVC video stream. This paper proposes to combine both SVC BL and H.264/AVC information to speed up the SVC EL encoding process with minor quality degradation (see Fig. 2(c)). This acceleration enables a capacity increase of the existing H.264/AVC-to-SVC transcoders resulting in more efficient SVC video stream generation. This, in its turn, could lead to a faster acceptance of SVC in the broadcast chain with better user experience as a consequence.

The remainder of this paper is organized as follows. First, quality scalability of SVC is explained in Section 2. Section 3 describes an analysis indicating the relation between the input H.264/AVC video stream and the generated SVC video stream. Then, in Section 4, the proposed transcoding algorithm is provided. Subsequently, in Section 5, the algorithm is verified with results and Section 6 ends this paper with a conclusion.

2 Quality Scalability in SVC

In SVC, two different types of quality scalability are incorporated, namely Medium Grain Scalability (MGS) and Coarse Grain Scalability (CGS) [1]. In both types of scalability, the quality of the decoded video stream increases with every additional layer. Efficient compression of these layers is obtained by three types of inter-layer prediction. These tools are named Inter-Layer Motion Prediction, Inter-Layer Residual Prediction, and Inter-Layer Intra Prediction. For every macroblock mode that is tested during the encoding of the enhancement layer, a performance comparison needs to be made both with and without this inter-layer prediction. This doubles the complexity of the EL coding compared to BL coding.

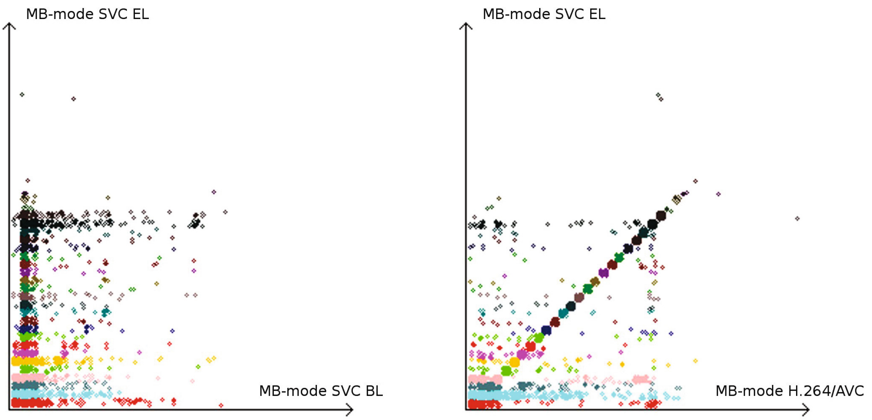
The difference between MGS and CGS is that MGS offers more flexibility when dropping quality layers. The disadvantage related to this flexibility is that an error drift is introduced when a lower quality layer is selected. Because quality is key in broadcasting mobile TV, CGS is preferred to MGS for this application.

Using quality scalability, a complex mode decision procedure needs to be performed for every quality layer. This mode decision process is an evaluation of different modes. For every mode, a cost is calculated with a Lagrangian cost function. To do this calculation, the best prediction for every mode must be looked for, which requires a lot of execution time. Our proposed algorithm accelerates this mode decision operation of the EL.

3 Analysis

To analyze the MB modes, a test set with varying and realistic properties is created. Sequences with varying characteristics are used (Harbour, Ice, Rushhour, Soccer, Station, Tractor), all at 4CIF resolution. The codec used for compression and decompression is JSVM (Joint Scalable Video Model) 9.19.7 [7]. This is the reference software for the SVC project of the Joint Video Team of MPEG and VCEG. A hierarchical GOP structure is used with a length of eight frames together with an intra frame period of 32 frames. In total, 64 frames are encoded for every sequence.

The transcoding process starts with H.264/AVC video streams as input, producing SVC video streams on the output. To get an overall result over a variety of qualities, the input H.264/AVC sequences are compressed with a Quantization



(a) Relation between MB mode of BL and MB mode of EL.

(b) Relation between MB mode of H.264/AVC and MB mode of EL.

Fig. 3. Relations between MB modes

Parameter (QP) varied between 22 and 37 in steps of five. For the outputted SVC sequences, the QP of the base layer is fixed to 42 and the QP of the enhancement layer is varied between 22 and 37 in steps of five, conforming to the H.264/AVC streams.

In Fig. 3, the relation between MB types is visualized. Fig. 3(a) shows the relation between MB types of the SVC base layer compared to the SVC enhancement layer. From the graph, it can be noticed that no clear relationship is present between the modes. However, by analyzing the data thoroughly, a more complex relationship can be found resulting in an acceleration of the mode decision process [8] [5], but this is outside the scope of this paper.

The relation between H.264/AVC modes of the input stream and the SVC enhancement layer is depicted in Fig. 3(b). In contrast to the previous observation, a more pronounced relation can be observed. This clear linear relation suggests that an improvement of the existing fast mode decision algorithms is possible. In the proposed algorithm both information sources, namely input video stream and base layer, are used to quicken the mode decision process in the enhancement layer.

4 Proposed Transcoding Algorithm

When recoding H.264/AVC to SVC, the most complex part of the process is encoding the SVC EL. As stated before, it takes 2/3 of the total encoding process. Our proposed algorithm uses MB modes from both the SVC BL and the H.264/AVC input video stream to construct a fast mode decision algorithm (see Fig. 2(c)). The algorithm comes in two varieties.

In both proposed algorithms, all intra prediction modes are evaluated, since intra prediction does not take up a large amount of processing power. Consequently, only evaluating a selection of the intra modes would result in minor complexity reduction. In addition, skip modes like P-skip, B-skip, and BL-skip are also evaluated for every macroblock during encoding, because of their effective RD performance.

The distinction between both proposed algorithms is that the first algorithm will only accelerate on MB mode level whereas the second one will also take submacroblock modes into account. The first algorithm reduces complexity by evaluating the subset of MB modes that occur at the co-located MB in the H.264/AVC input video stream or the base layer. For example, if a certain macroblock was encoded using 16x8 partitions in the H.264/AVC video stream and it was encoded using a 16x16 partition in the base layer, then only 16x8 and 16x16 partitions are evaluated for the enhancement layer.

For the second acceleration scheme, this idea is expanded to submacroblock modes. Previously when an 8x8 mode was evaluated, all subpartitions (8x8, 8x4, 4x8, and 4x4) were evaluated as well. The second algorithm only evaluates the subpartitioning of the corresponding MB. As a result, if an MB is partitioned into 8x4 submacroblocks in H.264/AVC or in the BL then only this subpartition is evaluated when encoding the EL.

Both algorithms are summarized in a flow chart in Fig. 4.

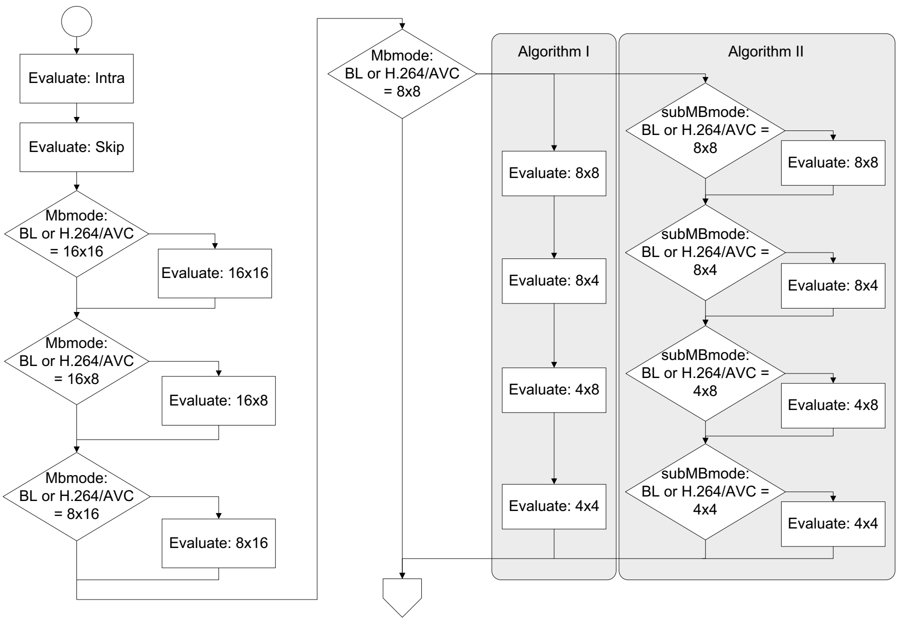


Fig. 4. Flow chart of the proposed fast mode decision algorithm with the differences between algorithm I and II indicated

5 Experimental Results

To evaluate the proposed algorithms, the test set of Section 3 is used. A transcoding algorithm is evaluated by rate-distortion performance and speed gain. Therefore, first we will evaluate the acceleration, followed by the resulting quality reduction. Comparisons are made between the proposed algorithms and three other transcoding configurations. All time and quality measurements will be referenced relative to the recoding time and quality, respectively (see Fig. 2(a)). Furthermore, because our algorithms combine information from the SVC BL and the H.264/AVC input video stream, also a comparison is made when only one source of information is used. When only the SVC BL is used, this is labeled as SVC fast mode decision (see Fig. 2(b)) and when only the H.264/AVC video stream is used, the result is labeled as H.264/AVC fast mode decision. The comparison shows that both techniques separately always have the disadvantage of resulting in significant quality reductions.

The results of time gain measurements are summarized in Table 1. Independent of the sequence, a constant gain is noticed, so further conclusions will be made according to the averages. For the first proposed algorithm, a speed increase of the EL with 78% can be noticed. Because of the acceleration of submacroblock modes in the second proposed algorithm, an extra 3% increase is realized. Notice that only using the H.264/AVC input stream or the SVC base layer will accelerate the execution time with 87% and 82% respectively. Although, H.264/AVC fast mode decision is 7% faster compared to the second proposed algorithm, this gain will not justify the resulting bandwidth increase as discussed below.

In Table 2, the acceleration of the total H.264/AVC-to-SVC transcoder is summarized. With the proposed techniques, total transcoding times are reduced with 51% on average.

The impact on bandwidth is calculated according to the Bjøntegaard delta metric [9] and shown in Table 3. These bandwidth penalties resulting from this metric are relative with respect to the recoding scenario. For both proposed algorithms, a 0.2% bandwidth increase is measured. Note that this increase is

Table 1. Acceleration of enhancement layer execution time relative to execution time in the recoding configuration

Sequence	Prop. Alg. I	Prop. Alg. II	H.264/AVC fast mode dec.	SVC fast mode dec.
Harbour	74%	78%	84%	82%
Ice	80%	81%	91%	83%
Rushhour	78%	80%	87%	83%
Soccer	78%	80%	88%	82%
Station	80%	81%	89%	83%
Tractor	76%	79%	85%	82%
Average	78%	80%	87%	82%

Table 2. Time saving of full H.264/AVC-to-SVC transcoder relative to the recoding configuration

Sequence	Prop. Alg. I	Prop. Alg. II	H.264/AVC fast mode dec.	SVC fast mode dec.
Harbour	49%	51%	55%	54%
Ice	52%	53%	59%	54%
Rushhour	52%	53%	58%	55%
Soccer	51%	52%	58%	54%
Station	53%	53%	59%	54%
Tractor	50%	51%	55%	54%
Average	51%	52%	57%	54%

Table 3. BD bandwidth increase relative to the recoding configuration

Sequence	Prop. Alg. I	Prop. Alg. II	H.264/AVC fast mode dec.	SVC fast mode dec.
Harbour	0.5%	0.5%	4.7%	35.9%
Ice	0.0%	0.0%	2.1%	29.2%
Rushhour	0.0%	0.0%	3.0%	29.9%
Soccer	0.3%	0.2%	11.8%	25.9%
Station	0.0%	0.0%	13.8%	21.7%
Tractor	0.5%	0.4%	4.7%	25.7%
Average	0.2%	0.2%	6.7%	28.0%

negligible and will not impact the bandwidth requirements of the application. For the H.264/AVC fast mode decision algorithm, an increase of 6.7% in bandwidth needs to be taken into account. The SVC fast mode decision algorithm performs the worst by increasing the bandwidth with 28%.

When looking at the individual results of the different sequences for the H.264/AVC fast mode algorithm, the sequences Soccer and Station perform considerably worse. This can also be noticed in the rate distortion graphs described next.

Rate distortion graphs of the Station sequence are visualized in Fig. 5. The Soccer sequence shows similar characteristics and RD graphs for this sequence are added at the end of this paper in Fig. 7. The recoding and both proposed algorithms all lay indistinguishably on the top curve because of the insignificant losses introduced by the proposed algorithms. Only bandwidth increases of the H.264/AVC and SVC fast mode decision algorithms are noticeable. The graphs also show that the bandwidth increase of the H.264/AVC fast mode decision algorithm is almost equal in size over the entire bandwidth range. This is because both the H.264/AVC video stream as well as the SVC enhancement layer are coded with equal quality. This finding is opposed to the SVC fast mode decision algorithm, where the bandwidth increase is more significant for higher qualities. An explanation can be found at higher bitrates where the quality difference

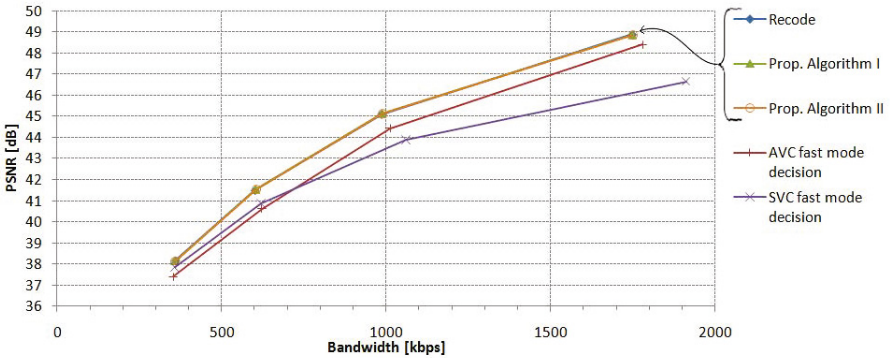


Fig. 5. RD graphs of Station sequence

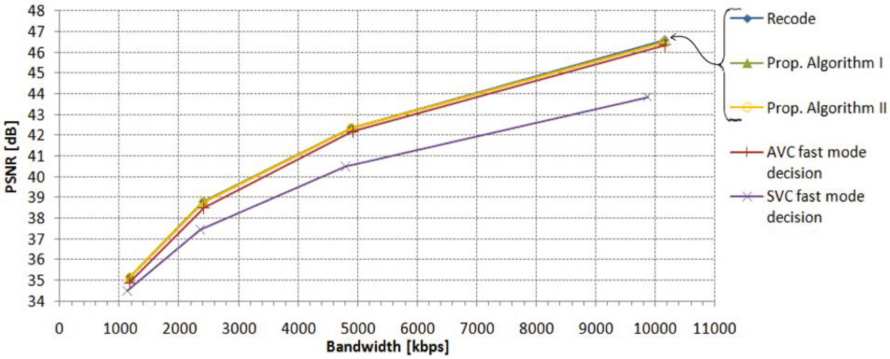


Fig. 6. RD graphs of Harbour sequence

between BL and EL becomes larger. The larger this difference, the smaller the correlation between the BL and EL modes. For example, a lower quality will result in larger partition sizes. Therefore, only these larger partitions are used in the higher quality layer, which will result in less efficient compression. When the quality difference increases, the base layer partition will have lower probability of being the optimal partition in the EL, resulting in an RD penalty.

When studying the graphs of the other sequences (Fig. 6, and Fig. 8-10), a great resemblance with the results of the Harbour sequence (Fig. 6) is noticed. In these graphs, only the SVC fast mode decision quality loss is noticeable. The small loss of the H.264/AVC fast mode decision curve can only be perceived by close inspection. As in all other graphs, no distinction can be made between the recoding quality and the quality of the proposed algorithms.

These results show that both algorithms reduce the H.264/AVC-to-SVC transcoding process significantly (52%) without a noteworthy quality reduction. It can also be concluded that, when the quality difference between BL and EL is not small, a fast mode decision algorithm based on the input video stream outperforms an SVC fast mode decision algorithm in rate distortion sense.

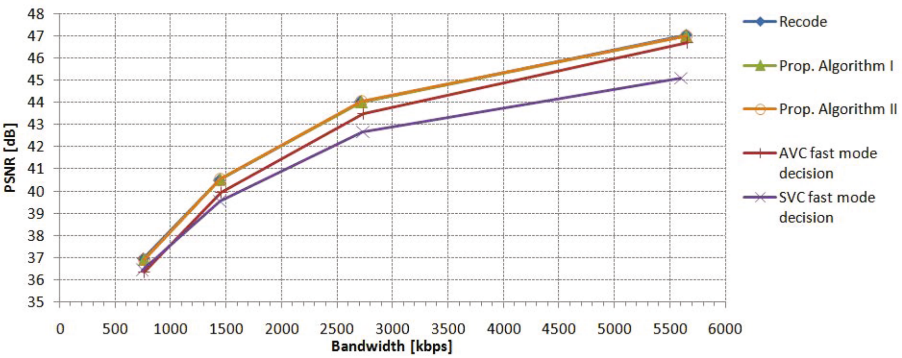


Fig. 7. RD graphs of Soccer sequence

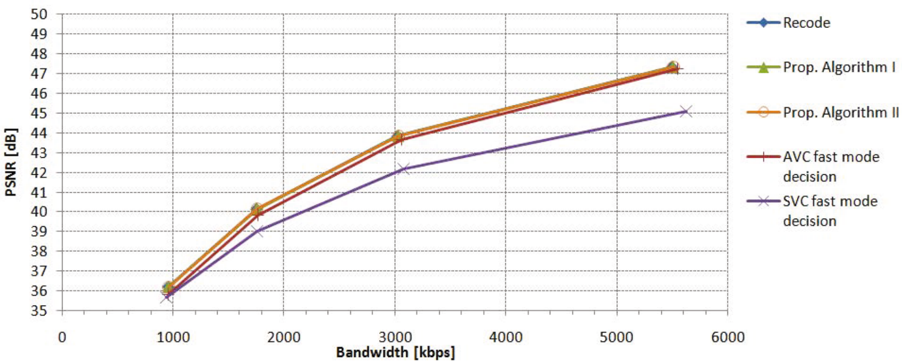


Fig. 8. RD graphs of Tractor sequence

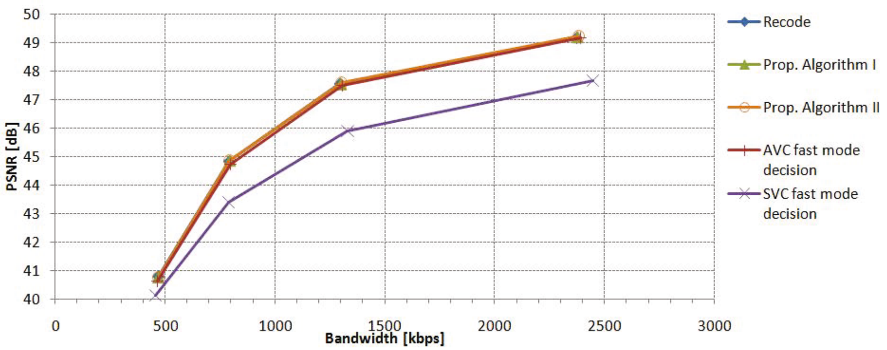


Fig. 9. RD graphs of Ice sequence

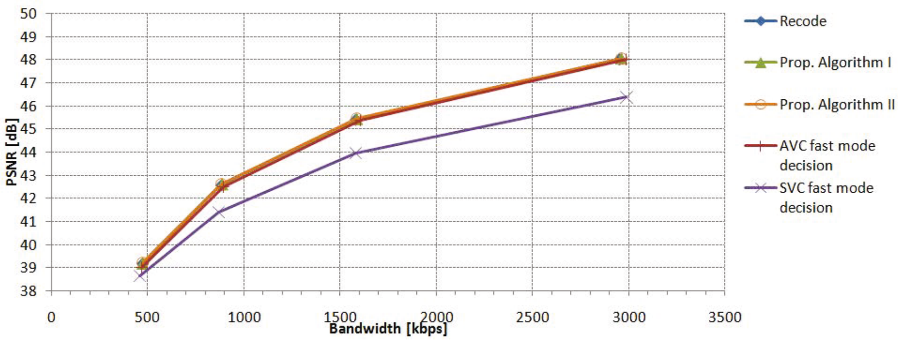


Fig. 10. RD graphs of Rushhour sequence

6 Conclusion and Future Work

In this paper, we proposed two fast mode decision algorithms using information from both the H.264/AVC input video stream and the SVC base layer. Using these algorithms, execution time of the H.264/AVC-to-SVC transcoding process is reduced with 52%. This time reduction only costs 0.2% bandwidth increase and thereby outperforms systems relying on only the input or the SVC base layer. Furthermore, it can be concluded that acceleration on the submacroblock level should be added as well because an extra time saving is realized without impacting bandwidth.

As an extension of this work, an acceleration of the SVC base layer will be considered. With the macroblock modes of the H.264/AVC video stream as input, the encoding process of the base layer can be accelerated as well.

Acknowledgements. The research activities that have been described in this paper were funded by Ghent University, University of Castilla-La Mancha, the Interdisciplinary Institute for Broadband Technology (IBBT), the Institute for the Promotion of Innovation by Science and Technology in Flanders (IWT-Flanders), the Fund for Scientific Research- Flanders (FWO-Flanders), and the European Union.

References

1. Schwarz, H., Marpe, D., Wiegand, T.: Overview of the Scalable Video Coding extension of the H.264/AVC standard. IEEE Trans. on Circuits and Systems for Video Technology 17(9), 1103–1120 (2007)
2. Palau, C., Martinez-Nohales, J., Mares, J., Molina, B., Esteve, M.: On Mobile Video Streaming IPTV. In: 10th International Conference on Telecommunications (ConTEL 2009), Zagreb, Croatia, June 08-10, pp. 457–462 (2009)
3. Vetro, A., Christopoulos, C., Sun, H.: Video transcoding architectures and techniques: an overview. IEEE Signal Processing Magazine 20(2), 18–29 (2003)

4. De Cock, J., Notebaert, S., Van de Walle, R.: Transcoding from H.264/AVC to SVC with CGS layers. In: IEEE International Conference on Image Processing (ICIP), San Antonio, TX, September 16-19, pp. 1769–1772 (2007)
5. Li, H., Li, Z., Wen, C., Xie, S.: Fast mode decision for coarse granular scalability via switched candidate mode set. In: 2007 IEEE International Conference on Multimedia and Expo., pp. 1323–1326 (July 2007)
6. Kim, S.-T., Konda, K., Park, C.-S., Cho, C.-S., Ko, S.-J.: Fast mode decision algorithm for inter-layer coding in scalable video coding. IEEE Transactions on Consumer Electronics 55(3), 1572–1580 (2009)
7. Joint Video Team (JVT) of ISO/IEC MPEG & ITU-T VCEG, Doc. JVT-AB203: Draft reference software for SVC, MPEG/ITU-T, Tech. Rep. (July 2008)
8. Van Leuven, S., De Wolf, K., Lambert, P., Van de Walle, R.: Probability analysis for macroblock types in spatial enhancement layers for SVC. In: Signal and Image Processing 2009: Proceedings of the IASTED International Conference, pp. 221–227 (August 2009)
9. Bjøntegaard, G.: Doc. VCEG-M33: Calculation of average PSNR differences between RD-curves, ITU-T, Video Coding Experts Group (VCEG) (April 2001)