A LOW-COMPLEXITY CLOSED-LOOP H.264/AVC TO QUALITY-SCALABLE SVC TRANSCODER

Sebastiaan Van Leuven, Jan De Cock,
Glenn Van Wallendael, Rik Van de Walle[*]

Ghent University - IBBT
ELIS Department - Multimedia Lab
Gaston Crommenlaan 8 b 201,
B-9050 Ledeberg-Ghent

Rosario Garrido-Cantos,
José Luis Martínez, Pedro Cuenca

Albacete Research Institute of Informatics
University of Castilla-La Mancha
Albacete, Spain

ABSTRACT

Efficient broadcasting of video content to end-users often requires one or more adaptations of the bitstream, due to varying network conditions and different end-user device characteristics. To ensure a high quality of experience for all end-users, the highest possible quality of the bitstream and, contradictory, connectivity for low bandwidth devices should be guaranteed. H.264/AVC allows only a single (high quality) bitstream. Therefore, the lower bit rates need adaptations of the input bitstream, requiring processing power, delay and energy. By transcoding the existing H.264/AVC bitstream to SVC, bit rate adaptations can be efficiently performed in the network. Consequently, only the cost of one transcoding step is required.

To ensure optimal transcoding, we present a low-complexity solution for transcoding H.264/AVC bitstreams to SVC. The proposed system can be applied in a broadcasting environment, since less than 10% of the normal transcoding complexity is needed, while coding efficiency is maintained.

Index Terms— Closed-loop transcoding, Scalable video coding, quality scalability, heterogeneous networks

1. INTRODUCTION

The broadcasting of video intended solely for television sets belongs to the past. Pervasive environments and mobile devices gain momentum while internet television and IPTV services are becoming widespread. To ensure interoperability, broadcasters should be aware that their content is subject to different network conditions and device capabilities. To guarantee an optimal quality of experience (QoE) for the end-user, the broadcasted video stream, encoded using H.264/AVC [1], should be adapted to the varying network conditions and end-user device capabilities. For example, on the convergence point of a broadband and an access network (e.g. mobile network), the resolution of the video can be reduced, since high definition resolution content is most likely not required for mobile devices. After routing the scaled bitstream through the mobile network, the stream can be adapted again on the last transmission point, typically a mobile network base station. Performing these transcoding operations each time for an H.264/AVC bitstream requires extra delay, processing power and energy consumption. On the other hand, not performing these steps will require more energy consumption down the line because of the increased bandwidth for the mobile link or processing power for the mobile devices.

With the advent of scalable video coding (SVC), the scalable extension of H.264/AVC, the problem of introducing multiple transcoding steps can be resolved. Encoding the video stream with SVC instead of H.264/AVC allows the network to scale the bitstream accordingly on the fly, without requiring significantly more processing power and energy consumption than parsing the bitstream. However, since SVC is not frequently used at the encoding side, mostly due to previous investments made in H.264/AVC equipment, the broadcasted video stream might remain an H.264/AVC output stream. Therefore, an H.264/AVC-to-SVC transcoding step can be applied on the convergence point of a broadband and an access network, as demonstrated in Fig. 1. So, scalability is added to the bitstream, allowing straightforward adaptation further down the network by requiring only one transcoding step. Finally, the total power consumption for the network and the end-user device is reduced, resulting in decreased operating costs and prolonged end-user connectivity.

The H.264/AVC-to-SVC transcoding in itself is a complex and thus energy consuming operation. To be able to reduce the overall costs for using SVC, the transcoding step has to be a low-complexity operation. Therefore, we propose an H.264/AVC-to-SVC transcoder which is capable of reducing the complexity with more than 90% compared to a cas-

[*]The research activities as described in this paper were funded by Ghent University, the Interdisciplinary Institute for Broadband Technology (IBBT), Ph.D. and post-doctoral fellow grants of the Agency for Innovation by Science and Technology (IWT), the Fund for Scientific Research-Flanders (FWO-Flanders), and the European Union.

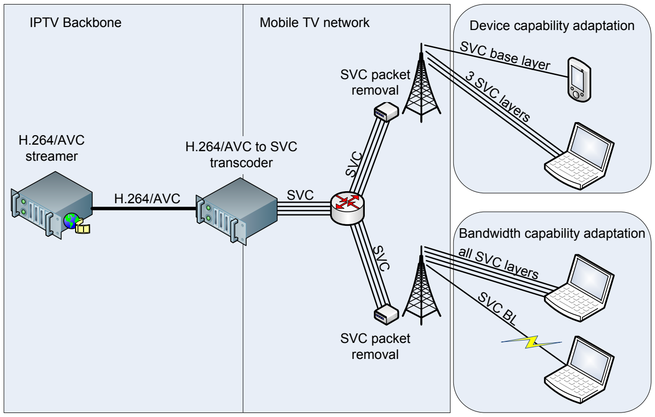


Fig. 1. Scalable video network example with both variable bandwidth and multiple end-user device illustrations

caded decoder-encoder solution, while maintaining the same bandwidth and quality. The proposed system transcodes an H.264/AVC bitstream to an SVC bitstream with quality scalability. Each network component is now able to efficiently adjust the bit rate with a fine granularity by reducing the quality, the frame rate, or a combination thereof to meet the network conditions.

The following section gives an overview of the related work in H.264/AVC-to-SVC transcoding. Section 3 elaborates on our proposed system, while Section 4 shows results of the system. Finally, Section 5 concludes this paper.

2. RELATED WORK

H.264/AVC-to-SVC transcoding schemes have been proposed in the past for the different types of scalability (temporal, spatial, and quality). Temporal H.264/AVC-to-SVC transcoding has been considered in [2], while spatial transcoding has been investigated in [3]. For quality scalability, coarse-grain scalability (CGS) and medium-grain scalability (MGS) can be used. CGS exploits the concepts of spatial scalability, by using for each layer the same resolution but a different quantisation parameter (QP). CGS allows to switch to a different quality only on pre-defined points in the bitstream, while MGS allows switching on a per-frame basis by tolerating drift between so-called key pictures. Since quality is of utmost importance in broadcasting, we focus on CGS quality layers. In [4] an H.264/AVC-to-SVC transcoder applying CGS by using open-loop architectures has been presented. The open-loop architectures imply that motion compensated macroblocks are not adapted if the referenced macroblocks are modified, which will result in drift errors. Consequently, these drift errors yield a lower QoE for end-users. Since a low QoE is unacceptable in broadcast environments, a closed-loop architecture is suggested.

In [5], the authors present a simple closed-loop architecture. This architecture is based on an analysis of the macroblock modes in the original input H.264/AVC bitstream and the corresponding SVC bitstream. The analysis results in a fast mode decision model, which optimizes the encoder of a cascaded decoder-encoder. The presented results show a complexity reduction of 57% while only a small rate distortion (RD) loss of 6.7% Bjøntegaard Delta bit rate (BDRate) [6]. Since this technique does not extensively exploits all information from the input H.264/AVC bitstream, our proposed method is able to further reduce the complexity.

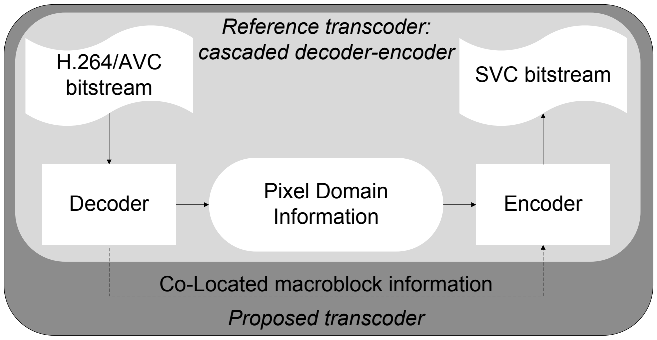


Fig. 2. Cascaded decoder-encoder scenario

3. PROPOSED SYSTEM

From an H.264/AVC encoded bitstream, an SVC CGS version is created. The quality of the enhancement layer is given by the maximum available quality of the H.264/AVC bitstream, i.e. the same quantisation as the H.264/AVC bitstream is applied. To scale to lower rate points, the quantisation of lower layers is increased in the SVC bitstream. Drift errors are avoided by applying closed-loop transcoding, based on a cascaded decoder-encoder scenario (Fig. 2). A signaling path from decoder to encoder with co-located macroblock information of the H.264/AVC bitstream is proposed to optimise the encoding. This optimises both the mode decision and submode decision by reducing the number of mode evaluations. Additionally, the prediction direction and the motion vector search range are reduced.

3.1. Base layer mode decision

Since the base layer is H.264/AVC compatible, the mode decision process of the base layer is the same as for H.264/AVC. During 'normal' re-encoding, the macroblock is encoded with the rate-distortion (RD) optimal mode, after evaluating all modes. To reduce complexity, the number of evaluated modes is reduced. Therefore, the mode selected for the H.264/AVC macroblock (hereafter referred to as $MODE_{AVC}$) can be used as prior knowledge to bias the mode decision process. Since intra-coding is typically low-complex, it is still evaluated for all macroblocks. Consequently, only (bi-) predictive modes evaluations are reduced.

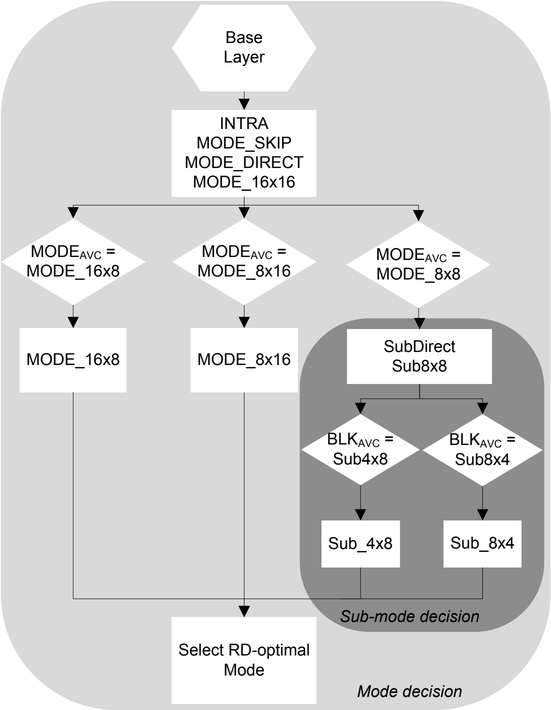


Fig. 3. Flowchart for base layer (sub-)mode selection process

The complete flowchart of the propsed base layer mode decision process is shown in Fig. 3. Due to the higher quantisation of the base layer, the probability for larger (sub)macroblock partitions will typically increase. Therefore, *MODE_16×16*, *MODE_Skip* and *MODE_Direct* are always evaluated in addition to $MODE_{AVC}$. The latter is evaluated because the lower quantisation does not guarantee that the most optimal H.264/AVC mode changes. For sub-macroblock modes the same principles apply; when the H.264/AVC mode is *MODE_8×8*, the low-complexity submodes *sub_Direct* and *sub_8×8* are always evaluated. Modes *sub_4×8* and *sub_8×4* are only evaluated when these correspond to the H.264/AVC sub-macroblock types (BLK_{AVC} in Fig. 3). Note that *sub_4×4* is never evaluated in the base layer because of the high complexity and the reduced probability due to the increased partitioning size.

3.2. Enhancement layer mode decision

The enhancement layer encoding process can use the base layer information as a prediction by using inter-layer prediction (ILP) [7]. Therefore, the enhancement layer evaluates each mode both with and without ILP during encoding. Consequently, for a CGS scenario, approximately 66% of the complexity is spent for the enhancement layer. This complexity can be reduced, since a relation between the $MODE_{AVC}$ and the enhancement layer mode ($MODE_{EL}$) is established in [5]. This relation shows that typically $MODE_{AVC}$ or a non-partitioned macroblock mode is selected, for the enhancement layer. Therefore, the evaluated modes are limited to either the input macroblock mode ($MODE_{AVC}$), or the base layer mode ($MODE_{BL}$), which might be non-partitioned. Additionally, *MODE_Skip* is evaluated because the enhancement layer reference picture might have been changed due to ILP. Therefore, using *MODE_Skip* might yield a better RD. Additional complexity reduction is obtained, by evaluating $MODE_{BL}$ only with ILP, while a normal encoding (without ILP) is applied for the $MODE_{AVC}$. Consequently, if $MODE_{BL} = MODE_{AVC}$, only one macroblock mode is evaluated compared to a standard enhancement layer encoding.

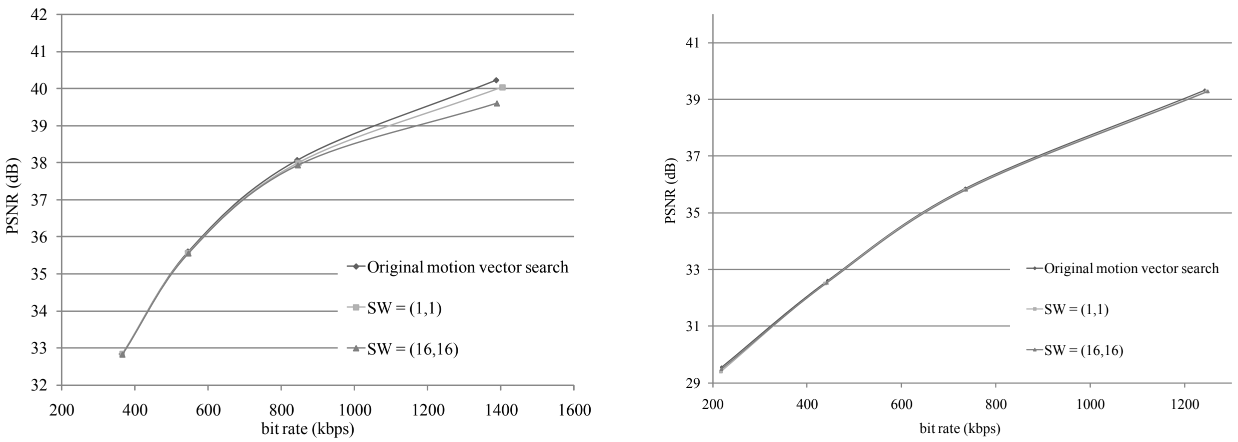
Sub-macroblock modes are evaluated with or without ILP, according to the evaluation of *MODE_8×8*. The complexity of this process is reduced by only evaluating *sub_Direct*, *sub_8×8*, and the co-located block size of the H.264/AVC bitstream. Note that *sub_4×4* might be evaluated, due to the increase in quality.

3.3. Prediction direction

In B pictures (bi-)predictive or intra prediction modes can be used. When using intra prediction or bi-predictive coding, no further optimizations are applied. However, a numerous number of macroblocks in B pictures are predictively coded by using only one of both prediction lists. It can be assumed that the same prediction list as the H.264/AVC macroblock is used for the SVC macroblock. Therefore, the (sub-) macroblock mode decision process only has to be performed for the corresponding prediction list. Consequently, for macroblocks which use only one prediction list, this yields to only a third of the complexity.

3.4. Motion vector estimation

Since the motion information is known from the H.264/AVC bitstream, it can be reused for the SVC bitstream. However, for both base and enhancement layer a motion refinement is proposed for two reasons. First, the base layer has a reduced quality, which might result in a different motion vector. Second, the enhancement layer motion vector could be different due to the ILP. The H.264/AVC motion vector is used as a starting point for the motion vector refinement, which defines a search window (SW) size for both base (SW_{BL}) and enhancement layer (SW_{EL}). For both layers, we have evaluated multiple SW combinations $(SW_{BL}, SW_{EL}) \in \{(1,1),(4,2),(8,1),(8,8),(16,8),(16,16)\}$ using six test sequences: *Harbour, Ice, Rushhour, Soccer, Station,* and *Tractor*. Unexpectedly, the combinations yielding a higher complexity $(16,16)$ do not necessarily result in a better RD. Fig. 4 shows the RD-curves for the extrema $SW = (1,1)$ and $SW = (16,16)$ compared to the RD of an unmodified cascaded decoder-encoder scenario. Two sequences (*Ice* and *Station*) are shown, which have respectively the worst



(a) Sequence *Ice* (ΔQP= 5)

(b) Sequence *Station* (ΔQP= 5)

Fig. 4. RD-curves for two sequences showing the impact of different search window sizes.

and best RD performance. As can be seen in Figs. 4(a) and 4(b), $SW = (1,1)$ outperforms or equals the RD of $SW = (16,16)$. This is because no RD optimisation is performed to limit the complexity while transcoding, therefore the effective bit rate impact of large motion vectors is not taken into account for the RD calculations. All following results are discussed with a single pixel (i.e., four quarter-pixel) search window size, which eliminates the need for a fast motion estimation algorithm.

4. RESULTS

The proposed system is evaluated for content with different characteristics, using six test sequences (*Harbour, Ice, Rushhour, Soccer, Station*, and *Tractor*) with a 4CIF resolution. Each sequence is encoded as an H.264/AVC bitstream with the H.264/AVC QP (QP_{AVC}): $QP_{AVC} \in \{27, 32, 37, 42\}$. The input bitstream is transcoded to an SVC CGS bitstream, with a base layer QP (QP_{BL}): $QP_{BL} = QP_{AVC} + \Delta QP$ and enhancement layer QP (QP_{EL}): $QP_{EL} = QP_{AVC}$. To evaluate the system, different ΔQPs have been evaluated: $\Delta QP \in \{5, 6, 8, 10\}$. Finally, also a scenario with a constant base layer quality ($QP_{BL} = 47$) for all rate points is applied.

The proposed system is based on the JSVM reference software (JSVM_9_19_9) [8] and is compared against an unmodified cascaded decoder-encoder of the same software (reference transcoder). This reference transcoder is used to evaluate the proposed system for both RD and complexity.

4.1. Rate distortion analysis

The RD curves for $\Delta QP = 5$ and $\Delta QP = 6$ show a minor difference between the original and the proposed transcoder. Fig. 4(a) shows the worst situation (valid for *Ice* and *Tractor*), while four sequences behave similar to Fig. 4(b). As can be seen from Fig. 5, a larger ΔQP results in a reduced coding efficiency. On average, a bit rate increase of 0.53% and 1.41% for a PSNR decrease of 0.11 dB and 0.12 dB are obtained for $\Delta QP = 8$ and $\Delta QP = 10$, respectively. As can be expected for a constant QP_{BL}, the proposed system performs better at low rate points, because of the small ΔQP for these points. For higher rate points the impact of the larger ΔQP results in a slightly lower RD for the proposed transcoder.

Table 1. BDPSNR and BDRate for ΔQP= 5 and ΔQP= 10

	DQP = 5		DQP = 10	
	BDPSNR	BDRate	BDPSNR	BDRate
Harbour	-0,038	1,018	-0,070	1,849
Ice	-0,089	1,658	-0,192	3,242
Rushhour	-0,038	0,821	-0,091	1,983
Soccer	-0,057	1,239	-0,174	3,877
Station	-0,051	0,964	-0,204	3,704
Tractor	-0,148	2,673	-0,413	7,295
Average	-0,070	1,396	-0,191	3,659

Table 1 shows the BDRate and Bjøntegaard Delta PSNR (BDPSNR) for the best ($\Delta QP = 5$) and worst ($\Delta QP = 10$) performance of our transcoder. As can be seen, only small Bjøntegaard measures are reported. Consequently, the proposed transcoder results in only a small bit rate differences for the same quality. Note that the average bit rate for $\Delta QP = 5$ is reduced by 0.02% with a ΔPSNR = -0.085, while for $\Delta QP = 10$ a 1.41% bit rate increase and a -0.12 dB PSNR decrease is measured

4.2. Complexity

The complexity of the system is evaluated as the time saving (TS) obtained by the proposed transcoding and is given by:

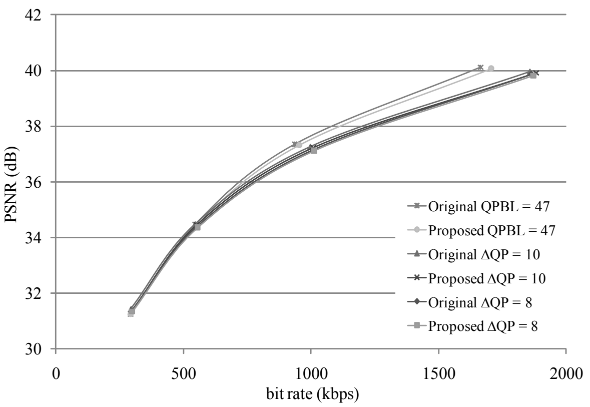


Fig. 5. RD for *Rushhour* with ΔQP= 8, ΔQP= 10 and QP_{BL}= 47.

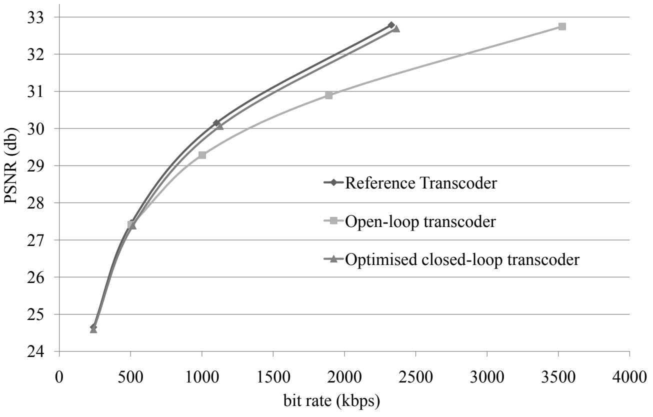


Fig. 6. RD for extracted base layer of sequence *Harbour with* ΔQP*= 5.*

$$TS\ (\%) = \frac{T_{Original}\ (ms) - T_{Fast}\ (ms)}{T_{Original}\ (ms)}\ .$$

Since complexity is hard to measure, the time saving gives an indication of the relative complexity for the proposed modifications within the same code base. The complexity reductions for base layer, enhancement layer and the full system are given in Table 2. On average, only 8.3% of the complexity of a cascaded decoder-encoder is required. Furthermore, the system is likely to be content independent, since a large set of video content is used to cover different video characteristics while similar complexity reductions are achieved. The complexity reduction will differ compared to real-world commercial solutions. However, JSVM is widely known and can be used as a common ground for comparison.

4.3. Comparison with existing techniques

Since there has not been a lot of investigation in the field of H.264/AVC-to-SVC transcoding, the number of algorithms is limited. To have a common ground of comparison, only techniques are considered which are able to transcode towards a quality scalable bitstream. Consequently, we will not compare our system with [2] and [3] since these techniques do not provide such a fine granularity of the rate points of the resulting SVC bitstream.

4.3.1. Closed-loop transcoding

Only one closed-loop transcoding algorithm has been previously proposed. In Section 2, the results for [5] are given. As can be seen, both the complexity as well the RD of our proposed closed-loop model outperforms this technique.

4.3.2. Open-loop transcoding

As was pointed out in Section 2, an open-loop transcoding mechanism for H.264/AVC-to-SVC with quality scalability has already been proposed [4]. Since the open-loop transcoding only applies an entropy decoding, dequantisation and requantisation step, the required complexity is very low. Compared to the cascaded decoder-encoder, near 100% complexity reduction is achieved. Obviously, in terms of complexity, open-loop transcoding outperforms our proposed method.

The rate distortion on the other hand is strongly influenced. Open-loop transcoding results in better RD for the enhancement layer, since no decoding step is applied on the input H.264/AVC bitstream. Consequently, the original encoded quality is maintained. On the other hand, the bit rate is drastically increased, specifically for the base layer, as can be seen in Fig. 6. Mainly because all intra-coded macroblocks are encoded in the the base layer. Consequently, the degree of scalability is reduced.

4.3.3. Fast mode decision models

In the past many fast mode decision models for SVC have been proposed. None of these models are optimised for encoding with the prior knowledge of an H.264/AVC bitstream. One the most referred models in literature [9], noted as Li's model, uses base layer information to reduce the complexity of the enhancement layer encoding. Additionally, the authors have suggested generic techniques to improve SVC enhancement layer encoding [10]. We compare our proposed H.264/AVC-to-SVC closed-loop transcoding technique with the results reported for Li's model extended with these generic techniques.

The complexity of the extended Li's model is only reduced for the enhancement layer, since the base layer encoding is not optimised. The required lowest complexity for the enhancement layer encoding is still 12.73% on average. Since the base layer encoding takes approximately 33% of the total complexity (due to the ILP), a reduction of 54.27% is achieved compared to 91.69% for our proposed closed-loop approach. To compare the RD, the absolute bit rate increase

Table 2. Complexity reduction for the proposed closed-loop transcoding architecture

	Average complexity reduction (%)						
	Harbour	Ice	Rushhour	Soccer	Station	Tractor	Avg.
Base Layer	85,83	85,28	85,69	85,43	86,4	86,76	85,90
Enhancement Layer	94,48	95,03	94,87	94,72	94,9	94,44	94,74
Full System	91,53	91,65	91,75	91,52	91,98	91,73	91,69

and PSNR values are compared (since the BDRate and BDPSNR are not reported in [10]). For the worst performing close-loop scenario, $\Delta QP = 10$, on average a bit rate increase of 1,41% and a PSNR reduction of -0.12dB is reported. The RD also outperforms the extended Li's model (bit rate: +2.14%; PSNR: -0.36dB).

This low RD for fast mode decision models is because the existing models exploit information from the low quality signal, which yields less optimal macroblock modes in the enhancement layer. Consequently, this result in a low RD performance. On the other hand, exploiting also H.264/AVC information will reduce the complexity but also improves significantly the overall RD. Since the best prediction for the high quality signal is known from the H.264/AVC bitstream, the base layer macroblock might be less efficient. However, this is greatly compensated by selecting the best macroblock mode for the enhancement layer. This is in line with the ideas and results for cross-layer optimisation.

5. CONCLUSIONS

Only a single transcoding step has to be applied, to cope with several different devices and heterogeneous networks. To reduce the complexity of this transcoding step, an optimised closed-loop transcoding scheme is proposed. By reducing the number of modes and optimizing the mode decision process, a low complex closed-loop transcoder is obtained. Only 8.3% of the complexity is required compared to a cascaded decoder-encoder scenario, while bit rate and quality remains stable. This complexity reduction will result either in more bitstreams being processed or less energy consumption with the same equipment. Compared to the existing optimised closed-loop transcoder, we further reduce the complexity, while improving the RD. Additionally, the drawbacks of an open-loop encoder are tackled. No drift artificats are introduced, the bit rate is reduced and due to the lower base layer bit rate the degree of scalability is increased.

6. REFERENCES

[1] Joint Video Team (JVT) of ISO/IEC MPEG & ITU-T VCEG, "Advanced Video Coding for Generic Audiovisual Services, ITU-T Rec. H.264 and ISO/IEC 14496-10 Advanced Video Coding, Edition 5.0 (incl. SVC extension)," Tech. Rep., MPEG / ITU-T, March 2010.

[2] Rosario Garrido-Cantos, Jan De Cock, José Luis Martínez, Sebastiaan Van Leuven, Pedro Cuenca, Antonio Garrido, and Rik Van de Walle, "Video Adaptation for Mobile Digital Television," in *3rd Joint IFIP Wireless Mobile Networking Conference (WMNC)*, Oct. 2010, pp. 1–6.

[3] Ravin Sachdeva, Sumit Johar, and Emiliano Mario Piccinelli, "Adding SVC Spatial Scalability to Existing H.264/AVC Video," in *ACIS-ICIS*, Huaikou Miao and Gongzhu Hu, Eds. 2009, pp. 1090–1095, IEEE Computer Society.

[4] Jan De Cock, Stijn Notebaert, Peter Lambert, and Rik Van de Walle, "Architectures for Fast Transcoding of H.264/AVC to Quality-Scalable SVC Streams," *IEEE Transactions on Multimedia*, vol. 11, no. 7, pp. 1209–1224, 2009.

[5] Glenn Van Wallendael, Sebastiaan Van Leuven, Rosario Garrido-Cantos, Jan De Cock, José Luis Martinez, Peter Lambert, Pedra Cuenca, and Rik Van de Walle, "Fast H.264/AVC-to-SVC Transcoding in a Mobile Television Environment," in *6th International Mobile Multimedia Communications Conference*, 2010.

[6] Gisle Bjøntegaard, "Doc. VCEG-M33: Calculation of average PSNR differences between RD-curves," Tech. Rep., MPEG / ITU-T, USA, 2-4 April. 2001.

[7] H. Schwarz, D. Marpe, and T. Wiegand, "Overview of the Scalable Video Coding Extension of the H.264/AVC Standard," *IEEE Trans. Circuits and Systems for Video Technology*, vol. 17, no. 9, pp. 1103–1120, Sept. 2007.

[8] Joint Video Team (JVT) of ISO/IEC MPEG & ITU-T VCEG, "Joint Scalable Video Model," Tech. Rep., MPEG / ITU-T, Jan. 2010.

[9] He Li, Zhengguo Li, Changyun Wen, and Lap-Pui Chau, "Fast mode decision for spatial scalable video coding," in *International Symposium on Circuits and Systems (ISCAS)*, May 2006.

[10] Sebastiaan Van Leuven, Glenn Van Wallendael, Jan De Cock, Rosario Garrido-Cantos, José Luis Martínez, Pedro Cuenca, and Rik Van de Walle, "Generic Techniques to Improve SVC Enhancement Layer Encoding," in *IEEE International Conference on Consumer Electronics (ICCE)*, Jan. 2011, pp. 135 –136.