COMBINING OPEN- AND CLOSED-LOOP ARCHITECTURES FOR H.264/AVC-TO-SVC TRANSCODING

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

Ghent University - IBBT
ELIS Department - Multimedia Lab
B-9050 Ledeberg-Ghent, Belgium

Rosario Garrido-Cantos
José Luis Martínez, Pedro Cuenca

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

ABSTRACT

Scalable video coding (SVC) allows encoded bitstreams to be adapted. However, most bitstreams do not incorporate this scalability so bitstreams have to be adapted multiple times to accommodate for varying network conditions or end-user devices. Each adaptation incorporates an additional loss of quality due to transcoding. To overcome this issue, we propose a single transcoding step from H.264/AVC to SVC. Doing so, the resulting bitstream can be freely adapted without any additional quality reduction. Open-loop transcoding architectures can be used for H.264/AVC-to-SVC transcoding with a low complexity, although these architectures suffer from drift artifacts. Closed-loop transcoding, on the other hand, requires a higher complexity. To overcome the drawbacks of both systems, we propose combining both techniques.

***Index Terms*—** Scalable Video Coding (SVC), transcoding, open-loop, fast mode decision, complexity reduction

1. INTRODUCTION

To allow flexibility of video streams towards varying network conditions or end-user devices, scalable video coding (SVC) has been designed based on the H.264/AVC video coding standard [1]. SVC allows for adaptability, by removing spatial, temporal or quality information from the bitstream [2].

Despite its adaptability, SVC is not yet commonly used. Therefore, currently H.264/AVC bitstreams are transcoded to cope with bandwidth fluctuations, different network characteristics or a large variety of end-user devices. The alternative, randomly drop packets, would deteriorate the visual quality and might result in an unacceptable quality-of-experience (QoE). Transcoding reduces the bit rate and quality in a controlled way. However, compared to encoding the sequence directly at the target bit rate transcoding introduces additional quality loss. To limit the quality reduction due to multiple transcoding steps, we propose to use a single transcoding step from H.264/AVC to an SVC bitstream with coarse grain scalability (CGS). Consequently, the rate-distortion (RD) degradation is limited, while the scalability allows for the bitstreams to be adapted whenever necessary without introducing additional loss in quality. CGS allows for quality scalability by coding additional quality data in the enhancement layer. Using spatial scalability [3] does not allow to handle small bit rate fluctuations efficiently, since the rate points of the extracted stream correspond to the reduced resolution. Transcoding with temporal scalability [4] is achieved by introducing a hierarchical temporal prediction structure to the bitstream. However, this can also be applied at the H.264/AVC encoder, which is preferred over using a transcoding step.

H.264/AVC-to-SVC with CGS transcoding has already been proposed for an open-loop architecture [5]. However, this open-loop transcoding architecture implies drift artifacts. These artifacts are due to the difference of the reference frame at the encoder and decoder side, because open-loop transcoding does not perform motion compensation. The drift errors propagate through the sequence, since consecutive frames use previously encoded (and distorted) frames as a reference.

In the next section, we propose an H.264/AVC-to-SVC transcoder which limits these open-loop transcoding drawbacks. Results for the proposed system are given in Section 3. Section 4 presents the conclusions and indicates future work.

2. PROPOSED SYSTEM

To overcome drift effects, open- and closed-loop architectures are combined, as shown in in Fig. 1. We optimised the cascaded decoder-encoder by applying fast mode decision which uses information encoded in the H.264/AVC bitstream.

2.1. Open-loop transcoder

Requantisation transcoding is used for the open-loop architecture [5]. A requantisation step is applied after entropy decoding and dequantisation. This results in the base layer of the SVC bitstream. Since only coefficients are adjusted, all information such as motion vectors and macroblock partitioning is

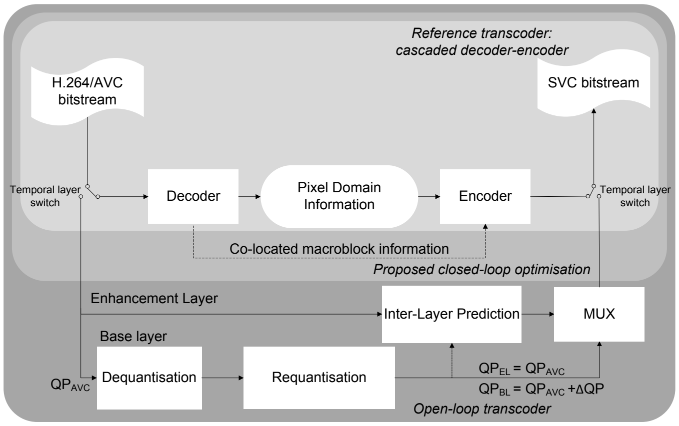

Fig. 1. Overview of the proposed combined open- and closed loop architecture for H.264/AVC-to-SVC transcoding.

maintained. Consequently, a sub-optimal base layer is generated, since for the lower quality different partition sizes might be preferred [6]. Additionally, all intra-coded macroblocks are encoded in the base layer, to reduce drift artifacts.

2.2. Closed-loop transcoder

The closed-loop transcoder introduces no drift errors because a complete decoding step is executed. Since information of the H.264/AVC bitstream can be reused, we propose a fast mode decision to reduce the complexity.

Our proposed *fast mode decision* which is applied to determine the base layer macroblock mode (μ_{BL}) is given in Fig. 2. Four macroblock modes ($MODE\_Direct$, $MODE\_Skip$, $MODE\_16 \times 16$, $MODE\_Intra$) are always evaluated because of a low complexity and high probability [6]. Additionally, the co-located H.264/AVC macroblock mode (μ_{AVC}) is evaluated, if it has not already been evaluated. For $MODE\_8 \times 8$ macroblocks, the same principles apply for the sub-macroblock partitioning. Furthermore, $sub\_4 \times 4$ is not evaluated for the base layer; due to the reduced quality, small partition sizes have a low probability while they yield a high complexity.

For the enhancement layer, a fast mode decision method is applied as well. Here, only μ_{AVC} and μ_{BL} are evaluated next to $MODE\_Skip$, while ILP is only applied for μ_{BL}. When $\mu_{BL} = \mu_{AVC}$, the mode is evaluated both with and without ILP, which corresponds to the normal mode evaluation process. Applying ILP when $\mu_{AVC} \neq \mu_{BL}$ would yield a high RD-cost, since the resemblance between different types is small, and is unlikely to be selected. Hence, to reduce the complexity, ILP is not evaluated for μ_{AVC}.

Furthermore, the complexity for evaluating the *prediction direction* (forward, backward, or bi-predictive) for B-frames is reduced. Both base and enhancement layer use the H.264/AVC list prediction, since it is unlikely that the prediction direction of the base and enhancement layer are different compared to the H.264/AVC bitstream.

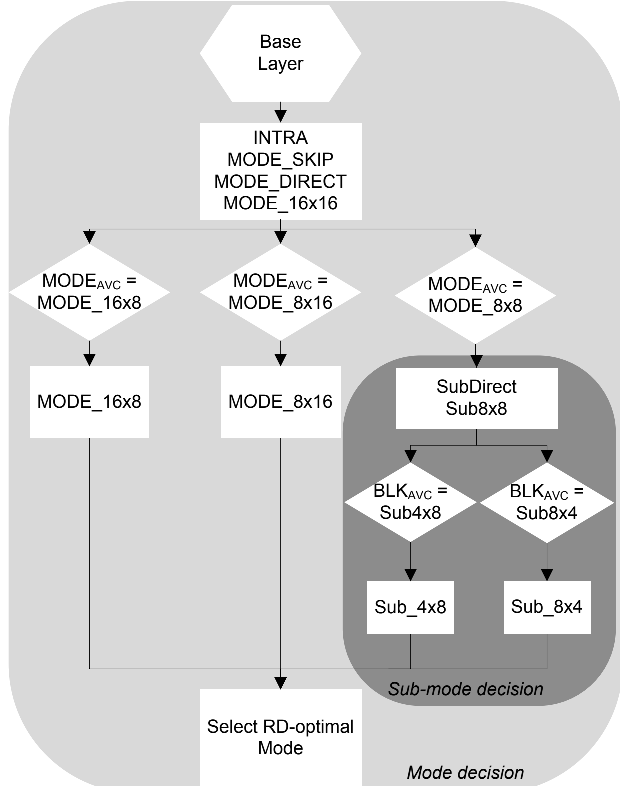

Fig. 2. Flowchart for base layer (sub-)mode selection process.

Lastly, the *motion estimation* complexity is reduced. For the base and enhancement layer, the motion vector of the H.264/AVC input bitstream (MV_{AVC}) is used as a predictor. To ensure the optimal motion vector is found, a one pixel wide motion refinement step is applied to this predictor. A full motion vector search is only performed when $MODE\_16 \times 16$ is evaluated for the base layer and $\mu_{AVC} \neq MODE\_16 \times 16$, since there might not be a linear mapping between the (partitioned) MV_{AVC} and $MODE\_16 \times 16$.

2.3. Combined open- and closed-loop transcoding architecture

Combining the open-and closed-loop architectures reduces the complexity of closed-loop transcoding, while the quality and scalability of open-loop transcoding is improved. This is done by either transcoding a frame open- or closed-loop, depending on the temporal ID (Tid) of the frame in the GOP, as indicated by the temporal switch in Fig. 1. Since open-loop transcoding results in drift artifacts, non-referenced frames (highest Tid) are open-loop transcoded, while all other frames are closed-loop transcoded.

Table 1. Complexity reduction for open-loop (OL), combined open- and closed-loop, and optimised closed-loop (CL).

Type	Complexity Reduction	Frames/GOP OL
Open-loop	∼100%	8
Combined OL/CL	95.73%	4
Optimised CL	91.52%	0

3. RESULTS

The proposed combined open- and closed-loop architecture, the optimised closed-loop transcoder and an open-loop transcoder are evaluated against a reference transcoder. The latter is a cascaded decoder-encoder based on the Joint Scalable Video Model reference software (JSVM_9_19_9) [7] as shown in Fig. 1. Six commonly used test sequences with 4CIF resolution were evaluated (*Harbour, Ice, Rushhour, Soccer, Station, Tractor*). These sequences have been encoded as H.264/AVC bitstreams with a hierarchical GOP of 8 frames. Different quantisation parameters (QP) were used: $QP_{AVC} \in \{27, 32, 37, 42\}$. These input bitstreams are transcoded to SVC with CGS, such that the base layer has a reduced quality, and the enhancement layer quantisation is maintained: $QP_{BL} = QP_{AVC} + \Delta QP; QP_{EL} = QP_{AVC}$. Here ΔQP is the difference in quantisation between base layer and enhancement layer and given by $\Delta QP \in \{5, 6, 8\}$. These numbers correspond to the range of ΔQP for which CGS is beneficial to use.

3.1. Complexity

Reduction in complexity is expressed as the time saving (TS) for transcoding with an optimised transcoder (T_{Fast}) compared to the reference transcoder ($T_{Original}$), and is given by:

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

Results for complexity measurements can be found in Table 1. The optimized closed-loop system has an average time saving of 91.5% compared to the reference transcoder. By open-loop encoding frames with: $Tid \geq 3$ (the highest Tid for a GOP with 8 frames) the complexity reduces by 95.73%. Since the complexity for open-loop transcoding is negligible (0.26% compared to re-encoding [5]) and the highest temporal resolution includes half the frames of a GOP, the combined open- and closed-loop transcoding roughly halves the required complexity compared to the optimised closed-loop.

3.2. Rate Distortion

RD results for a complete bitstream are shown in Fig. 3. Bjøntegaard Delta bit rate (BDrate) and PSNR (BDPSNR) [8] results are given in Table 2. The RD performance for the optimised closed-loop transcoder is close to the reference transcoder. However, the open-loop transcoder has a significantly improved RD performance. This is because the quality of the H.264/AVC bitstream is maintained, while the other methods first perform a decoding step. Nevertheless, the open-loop scenario introduces drift errors and increases the bit rate for the base layer. The combined architecture slightly increases the RD towards the open-loop scenario, while further halving the complexity. The combined open- and closed-loop transcoder achieves an increase of 7.11% BDrate and -0.33 dB BDPSNR.

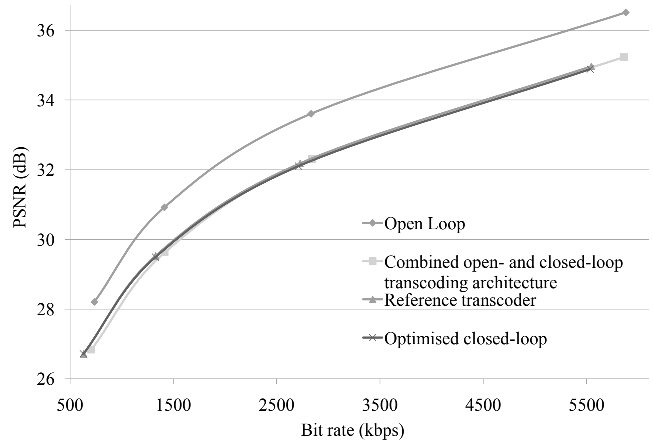

Fig. 3. RD-curves for *Harbour* with $\Delta QP = 5$.

RD curves of extracted base layer bitstreams are presented in Fig. 4. As can be seen, the RD performance for the base layer for both the reference transcoder and optimised closed-loop transcoder are close to each other. On the other hand, the RD curve for the open-loop scenario shows significantly higher bit rates. The PSNR for these rate points is higher, but does not justify the (approximately) doubled bit rates. As pointed out in Section 2, this is because of the sub-optimal base layer. By combining open- and closed-loop transcoding architectures, the base layer bit rate slightly increases compared to the optimised closed-loop transcoder.

4. CONCLUSION AND FUTURE WORK

Combining open- and closed-loop transcoding architectures requires only half of the complexity compared to an optimised closed-loop transcoder. The system transcodes an H.264/AVC input bitstream to an SVC bitstream with quality scalability. Compared to open-loop transcoding, drift artifacts are avoided, while the bit rate of base and enhancement layer is improved. By increasing the number of open-loop transcoded frames, the complexity can even further be reduced. This eventually leads to complexity scalable transcoding, where for each GOP the number of open-loop transcoded frames is evaluated based on the available resources.

Table 2. BDPSNR and BDrate for ΔQP= 5, ΔQP= 6, and ΔQP= 8.

		DQP = 5		DQP = 6		DQP = 8		Average	
		BDPSNR	BDrate	BDPSNR	BDrate	BDPSNR	BDrate	BDPSNR	BDrate
OL	Harbour	1.19	-26.70	1.18	-26.25	1.09	-24.62		
	Ice	0.75	-11.84	0.80	-12.73	0.84	-13.48		
	Rushhour	1.02	-17.40	1.07	-18.02	1.03	-17.50		
	Soccer	1.17	-22.51	1.16	-22.24	1.16	-22.14		
	Station	1.23	-13.77	1.23	-13.83	1.14	-12.54		
	Tractor	1.53	-23.78	1.59	-24.62	1.54	-24.13		
	Avg.	1.15	-19.33	1.17	-19.62	1.13	-19.07	1.15	-19.34
OL + CL	Harbour	-0.10	2.69	-0.08	2.25	-0.09	2.36		
	Ice	-0.48	9.52	-0.46	9.12	-0.49	9.35		
	Rushhour	-0.46	10.90	-0.46	10.87	-0.47	11.27		
	Soccer	-0.17	3.85	-0.15	3.37	-0.18	4.13		
	Station	-0.55	11.41	-0.60	12.28	-0.69	14.09		
	Tractor	-0.20	3.86	-0.14	2.76	-0.21	3.94		
	Avg.	-0.33	7.04	-0.31	6.77	-0.35	7.52	-0.33	7.33
CL	Harbour	-0.04	1.02	-0.05	1.20	-0.04	1.17		
	Ice	-0.09	1.66	-0.10	1.83	-0.16	2.84		
	Rushhour	-0.04	0.82	-0.05	1.02	-0.06	1.39		
	Soccer	-0.06	1.24	-0.06	1.33	-0.13	2.86		
	Station	-0.05	0.96	-0.08	1.58	-0.14	2.58		
	Tractor	-0.15	2.67	-0.12	2.05	-0.20	3.46		
	Avg.	-0.07	1.40	-0.08	1.50	-0.12	2.38	-0.09	1.76

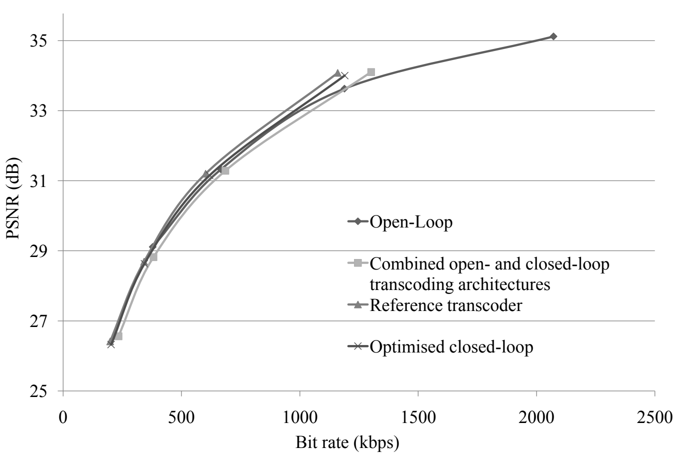

Fig. 4. RD-curves for the base layer of *Soccer* with $\Delta QP = 6$.

5. ACKNOWLEDGEMENT

The research activities as described in this paper were funded by Ghent University, IBBT, Ph.D. and post-doctoral fellow grants of IWT, FWO-Flanders, and the European Union.

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] 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.

[3] R. Sachdeva, S. Johar, and E. M. Piccinelli, "Adding SVC spatial scalability to existing H.264/AVC video," in *8th IEEE/ACIS Int. Conf. on Comp. and Information Science, Shanghai, China*, June. 2009, pp. 1090–1095.

[4] R. Garrido-Cantos, J.-L. Martínez, P. Cuenca, and A. Garrido, "An Approach for an AVC to SVC Transcoder with Temporal Scalability," in *HAIS (2)*, 2010, vol. 6077 of *LNCS*, pp. 225–232.

[5] J. De Cock, S. Notebaert, P. Lambert, and R. Van de Walle, "Architectures for fast transcoding of H.264/AVC to quality-scalable SVC streams," *IEEE Trans. on Multimedia*, vol. 11, no. 7, pp. 1209–1224, 2009.

[6] S. Van Leuven, et al., "Probability analysis for macroblock types in spatial enhancement layers for SVC," in *Proceedings of the 11th IASTED International Conference on Signal and Image Processing*, Aug. 2009.

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

[8] G. Bjøntegaard, "Doc. VCEG-M33: Calculation of average PSNR differences between RD-curves," Tech. Rep., MPEG / ITU-T, Austin, Texas, USA, April. 2001.