Generic Techniques to Improve SVC Enhancement Layer Encoding
Digest of Technical Papers

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

[1]Department of Electronics and Information Systems – Multimedia Lab, Ghent University – IBBT, Ghent Belgium
[2]Albacete Research Institute of Informatics, University of Castilla-La Mancha, Albacete, Spain

Abstract – **Scalable video coding is an important mechanism to provide different types of end-user devices with different versions of the same encoded bitstream. However, scalable video encoding remains a computationally expensive operation. To decrease the complexity we propose generic techniques. These techniques can also be combined with existing fast mode decision modes. We show that extending these existing techniques yield an average complexity reduction of 87%.**

I. INTRODUCTION

Applications for the Scalable Video Coding (SVC) extension of H.264 have not met the market yet. One of the reasons is the significant increase of the encoding complexity over single-layer H.264/AVC video, due to the layered nature of SVC. SVC allows for three types of scalability, i.e. quality, temporal and spatial. Using quality scalability, additional quality information is transmitted to the user, while temporal scalability allows scaling of the frame rate. Both techniques slightly increase complexity, in contrast to spatial scalability.

Spatial scalability allows different resolutions to be encoded in a single bitstream. In order not to end up with a multicast scenario, where all streams are encoded independently, inter-layer prediction (ILP) can be applied. Using ILP, the lowest resolution (base layer) can be used as a predictor for higher resolutions (enhancement layers). Hence, the mode decision (including motion estimation) has to be performed twice for the enhancement layer, once using regular techniques (as in H.264/AVC) and once with the base layer as a predictor. Therefore, spatial scalability comes with a high complexity.

To reduce the enhancement layer complexity, fast mode decision models have been proposed. Most of these models are based on limiting the evaluations of macroblock partition size, but do not limit complexity on a sub-mode decision level.

A fast mode decision method which exploits neighboring macroblock statistics is proposed in [1]. Encoded base layer information, such as macroblock types, was not used for spatial enhancement layer mode decision. Since a relation between base and enhancement layer modes has been shown [2], a lower complexity is feasible. In [3], a method based on a classification mechanism for the most probable modes is suggested. A model using co-located base layer modes [4] shows significant time savings, with small bit rate and PSNR changes. While relevant methods are listed here, many techniques using base layer information have been proposed. To further reduce the complexity, we propose techniques which can be freely combined. These generic techniques can even improve fast mode models.

The research activities that have been 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.

II. PROPOSED GENERIC TECHNIQUES

The easy to implement low-complex proposed techniques maintaining high coding efficiencies. These techniques can be mutually combined and extended with fast mode decision models as they mainly operate on a sub-mode decision level.

1) Disallow orthogonal macroblock modes

Since [2] showed orthogonal modes (e.g.: MODE_8x16 vs. MODE_16x8) in both layers have low probabilities, the orthogonal mode of the base layer should not be evaluated in the enhancement layer.

2) Only evaluate sub8x8 blocks if present in base layer

Sub8x8 partitions are very computationally expensive; therefore these calculations are limited to regions where they have been selected in the base layer.

3) Only evaluate the base layer list predictions

Since both layers have a high probability for using the same prediction list, only the referenced list from the base layer is evaluated. For bi-prediction both lists are evaluated.

In the following, these techniques will be referred to as 1, 2, and 3, respectively.

III. RESULTS

The proposed techniques have been implemented both as standalone and as additional techniques to [4]. The resulting have been encoded using JSVM 9.4 [5], for test sequences with different properties (*Harbour*, *Ice*, *Rushhour* and *Soccer*), with varying combinations of QP_{BL}, $QP_{EL} \in \{18, 24, 30, 36\}$. Here QP_{BL} and QP_{EL} are the quantization parameters of the base and enhancement layer, respectively. The enhancement and base layer have a 4CIF and CIF resolution respectively.

A. Generic techniques as a standalone solution

TABLE I shows the average results for the different proposed techniques. For singular techniques, the sub8x8 reduction method (2) results in the highest time saving, while maintaining the highest coding efficiency. When small complexity reductions are sufficient, it is a good candidate. Extending this technique with the reduction by orthogonality (1) results in even better time savings. Combining all three techniques will have the highest time saving; however, a bit rate increase of about 1% has to be acceptable.

TABLE I

Standalone scenario for proposed techniques

Method	ΔBR(%)	ΔPSNR(dB)	TS (%)
1	0.60	-0.05	26.95
2	0.20	-0.03	53.98
3	0.91	-0.06	17.76
1+2	**0.53**	**-0.09**	**73.91**
1+2+3	**1.06**	**-0.13**	**77.15**

ΔBR = bit rate increase; ΔPSNR = difference in quality (negative means reduction); TS (time saving) =complexity reduction for encoding the enhancement layer, given by: $TS = (Time_{JSVM} - Time_{Fast})/Time_{JSVM}$

TABLE II

Proposed techniques in combination with Li's model [4]

Method	ΔBR(%)	ΔPSNR(dB)	TS (%)
[4]	1.40	-0.25	66.76
[4] +1	1.55	-0.27	68.37
[4] +2	1.39	-0.28	82.02
[4] +3	2.13	-0.30	71.93
[4] +1+2	**1.50**	**-0.31**	**84.47**
[4] +1+2+3	**2.14**	**-0.36**	**87.27**

B. Generic techniques with fast mode decision models

While the proposed singular techniques are useful in the standalone scenarios, they have never been combined with existing fast mode decision models. We used [4] to evaluate the effects of generic improvements for such models. Results are shown in TABLE II.

When using multiple generic techniques, only the results for [4]+1+2 are shown, because [4]+2+3 and [4]+1+3 yield a higher complexity and lower coding efficiencies. Note that while adding a single technique only seems to yield small time savings, the absolute gain is comparable to those shown in TABLE I. As can be seen, [4]+1+2 has only 2.6% less time gain compared to [4]+1+2+3, although the absolute complexity of the latter is 17% lower compared to the former.

Comparing TABLE I with TABLE II shows that generic techniques yield better rate distortion (RD) for comparable time savings. From this observation, it can be concluded that singular generic techniques are preferred for small complexity reductions (< 80%), while the criterion for combinations with fast mode decisions lies with very low complexity solutions.

Fig. 1 shows the coding efficiency for the proposed generic techniques. It can be seen that only the total combined solution has a slightly lower RD performance. Such a small decrease justifies the use of low complex generic techniques. In Fig. 2, the RD-curves for combinations with fast mode decision techniques are shown. The mutual combination of generic techniques outperforms Li's model. However, the proposed combination with fast mode decision models results in negligible small RD loss compared to these state of the art fast mode decision models, while further halving the required complexity.

Based on the available complexity of the designed system, one of the proposed techniques can be chosen, while the highest possible RD efficiency is guaranteed.

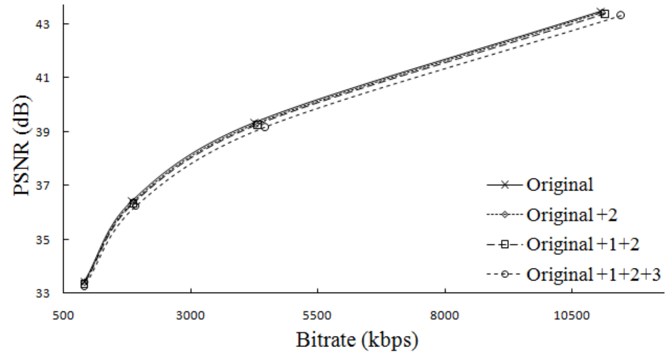

Fig. 1. RD comparison for generic techniques in a standalone scenario

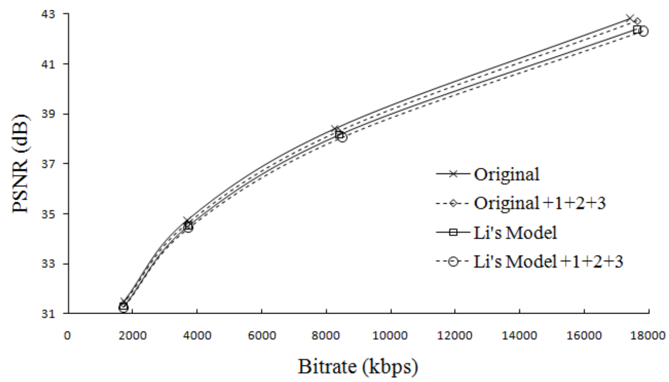

Fig. 2. RD comparison for *Harbour* @ QPBL = 30 using the combination of all techniques for both the original and Li's Model.

IV. CONCLUSION

The proposed generic techniques are usable in a standalone scenario where a complexity reduction is required, while a high coding efficiency is important. When combining these generic techniques with existing fast mode decision models, a system that needs only 12% of the complexity compared to a normal SVC encoder can be built. Furthermore, the presented techniques are applicable to future improved fast mode decision models. This opens the path for the introduction of SVC encoders to allow efficient transport systems to deliver one single bitstream, carrying multimedia content for different types of end-user terminals over different types of networks.

REFERENCES

[1] G.Goh, J. Kang, M.Cho, and C.-S. Cho, "Fast mode decsision for scalable video coding based on neighbouring macroblock analysis", in Proceedings of the 2009 ACM symposium on Applied Computing, pp 1845-1846, March 2009

[2] S. Van Leuven, K. De Wolf, P. Lambert, and R. Van de Walle, 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.

[3] S.-T. Kim, K. R. Konda, C. Su Park, C.-S. Cho, and S.-J. Ko, "Fast mode decision algorithm for inter-layer coding in scalable video coding," IEEE Trans. on Consumer Electronics, vol. 55, no. 3, pp. 1572–1580, Aug. 2009.

[4] H. Li, Z. Li, C. Wen, and L.-P. Chau, "Fast mode decision for spatial scalable video coding," in International Symposium on Circuits and Systems (ISCAS), May 2006

[5] Joint Video Team (JVT) of ISO/IEC MPEG & ITU-T VCEG, "Doc. JVT-W203: Joint Scalable Video Model 10," MPEG / ITU-T, Tech. Rep., April 2007.