AN EFFICIENT FGS CODING SCHEME FOR INTERLACED SCALABLE VIDEO CODING

Yanyan Zheng[1], *Xiangyang Ji*[1], *Feng Wu*[2], *Debin Zhao*[1], *Wen Gao*[1]

[1]Institute of Computing Technology, Chinese Academy of Sciences, Beijing, China
{yyzheng, xyji, dbzhao, wgao}@jdl.ac.cn
[2]Microsoft Research Asia, Beijing, China
fengwu@microsoft.com

ABSTRACT

JVT Scalable Video Coding (SVC) provides high coding efficiency for progressive video sequences with combined scalability. However, interlaced SVC is only the straightforward extension of H.264/AVC interlaced coding with the similar FGS coding technique for progressive coding. Based on the particular temporal correlation of interlaced video sequences, this paper presents a novel and efficient FGS coding scheme for interlaced SVC, which is able to achieve higher compression efficiency and further temporal scalability by introducing additional temporal decomposition stage when coding key pictures. The advantages of our proposed interlaced FGS coding scheme are verified by integrating it into JVT-SVC reference software.

***Index Terms*—**FGS, SVC, interlace

1. INTRODUCTION

The interlaced scanning technology has persisted in many camera and display designs, and the interlaced video contents are still widely used in digital TV broadcast and storage applications. It is well known that the advanced coding techniques adopted by SVC can bring higher coding performance and more flexible scalability than the other existing coding standards [1]. However, so far, there has not been efficient solution for interlaced scalable video coding with SNR enhancement layer, except for the straightforward implementation of progressive refinement coding technique as used for coding progressive sequences [3]. While H.264/AVC has adopted some technologies, such as field coding enhanced picture-adaptive frame/field (PAFF) and MB-adaptive frame/field (MBAFF) coding into the standard to meet corresponding requirements, SVC as the scalable extension of H.264/AVC also needs to be further investigated to better support spatial, temporal and SNR scalabilities for coding interlaced video sequences with high coding performance.

For scalable video coding, one of the main reasons to be able to achieve high coding performance is the conjunction of the high efficient motion model of H.264/AVC and block-wise adaptive temporal decomposition. And the others are hierarchical B picture coding structure, inter-layer prediction and fine granularity scalable coding which together make the combined temporal, spatial and SNR scalabilities of SVC possible and efficient. Therefore, in order to efficiently implement interlaced coding in SVC, the special characters of interlaced video sequences should be investigated and combined with the key features of SVC that contribute most to the high coding efficiency.

In this paper, we propose a novel FGS coding scheme that can efficiently compress interlaced video sequences, and smartly support temporal and SNR scalabilities in the basic framework of SVC. We use the reference software provided in [2] as the basis for interlaced coding scheme, in which the temporal correlations among high-pass and low-pass field pictures in both forward and backward directions are well-exploited by adopting multiple references prediction in hierarchical B picture coding structure. Benefited from this technique, higher compression ratio can be attained compared with only one temporally most recent reference available.

Nevertheless, only efficient high-pass and low-pass field coding of base layer without the competent integration of SNR and Spatial enhancement layer can not guarantee to improve the whole interlaced coding performance with flexibly combined scalability. In particular, for the case of coding at least one SNR enhancement layer, there is a potential possibility to be able to more efficiently exploit temporal correlations among adjacent top and bottom field pictures in the same frames with FGS technique adopted to encode their corresponding residual. In this paper, we mainly focus on the improvement of SNR enhancement layer and temporal scalability by introducing an additional temporal decomposition stage when coding temporal low-pass sub bands, customarily called key pictures in SVC.

The rest of this paper is organized as follows: Section 2 introduces the current interlaced coding scheme of base layer. Section 3 outlines and explains our proposed FGS coding approaches for temporal low-pass field pictures. In Section 4, experimental results are displayed and analyzed to verify the high efficiency of our proposed FGS coding scheme for interlaced SVC. Finally, conclusions are given.

2. INTERLACED CODING FRAMEWORK OF BASE LAYER

In SVC framework, hierarchical B picture coding structure is applied to reduce the temporal redundancy among the adjacent frames, and it has been verified to be able to efficiently compress the progressive sequences. However, the current hierarchical B decomposition structure is not quite suitable for interlaced coding. Some new coding techniques can be added into the framework of SVC to improve the interlaced coding performance.

Considering the high efficiency of hierarchical B picture coding structure for progressive sequences, it is a straightforward way to imitate its original prediction process in the way that neighboring frames with the same or lower temporal level are used as references for current frame to be encoded [1]. Thereupon, in current interlaced coding scheme, in order to generate the high-pass sub bands with lower energy, the top fields employ at least a pair of top and bottom field pictures as references in both forward and backward directions at the same temporal level, and the bottom fields perform nearly in the same way, except that the top field in the same frame picture can also be referenced in forward direction by the bottom field. Since this additional top field reference is the spatial- and temporal-closest picture for the bottom field, the redundancy between them may be better reduced.

Similarly but differently, when MCTF (motion-compensation temporal filtering) is adopted, the decomposition process is as illustrated in Figure 1, with multiple references prediction employed for each decomposition stage (just shown in concept).

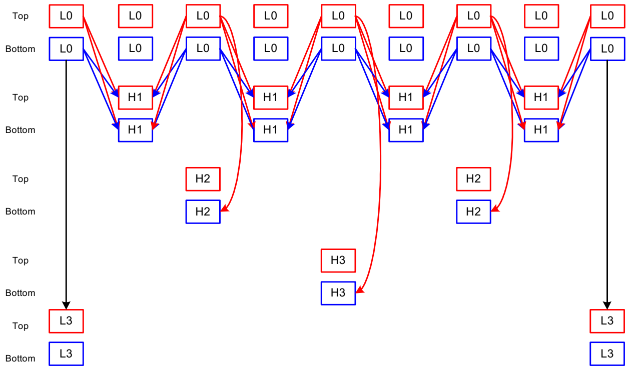


Fig.1: MCTF decomposition structure for interlaced SVC coding.

3. 3. INTERLACED CODING FRAMEWORK OF SNR ENHANCEMENT LAYER

For progressive source coding in the framework of SVC, after several hierarchical B decomposition stages, the temporal correlations among temporally adjacent low-pass pictures always become poor especially when the large GOP size is used. These pictures, namely key frames, will be either coded independently from each other as H.264/AVC intra pictures or inter coded as H.264/AVC inter-P pictures. Usually they are coded in IPPP… structure with close-loop motion compensation and one forward reference so as to not only control the propagation of drifting error, but also improve their coding performance.

3.1. Straightforward Implementation of Key Fields Coding

In current interlaced SVC software [2], the adopted reference prediction rule of coding temporal low-pass field pictures is similar as explained above and shown in figure 1 and figure 2 for two cases of reference number being set to one and two respectively and noted as Anchor1 and Anchor2. The remarkable benefit of these two methods is that both of them can efficiently control the drifting error even at very low bit rate. It is because they are coded as inter-P pictures with close-loop MC and thus the references for them are limited to be the reconstructions of base layer which can be obtained whenever at decoder. However, the well-performed error resilience associated with this basic FGS coding scheme is achieved with the expense of low coding efficiency. As to further comparing the coding performance of key fields, Anchor2 is obviously better than Anchor1.

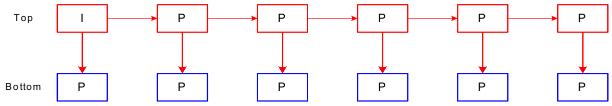


Fig.2: close-loop MC for key fields coding with one forward reference—Anchor1.

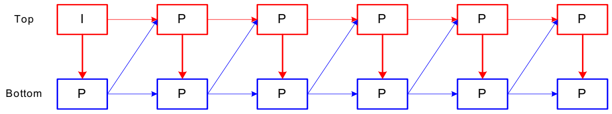


Fig.3: close-loop MC for key fields coding with two forward references—Anchor2.

3.2. Improved FGS Coding for Key Fields

Nevertheless, for interlaced source coding, there are still temporal correlations that have not been exploited even after several-level temporal decompositions. Illustratively speaking, in figure 1, L3s are the temporal low-pass pictures (key fields), although the temporal redundancy among L3s in different GOPs is small, the temporal correlations between the top and bottom fields that belong to the same frame are still strong. Therefore, it seems reasonable to introduce additional open-loop motion compensation between the top and bottom key fields in the same frame to further improve the coding performance of key fields.

3.2.1. TL_Haar1

We first implement this idea as shown in figure 4: the top key fields are coded as inter-P pictures with close-loop MC (solid line), while the remaining bottom fields are coded also as inter-P pictures but with open-loop MC (dashed line). In this way, the low-pass sub bands are actually made up of top key fields, and the bottom fields in key frames have become high-pass sub bands.

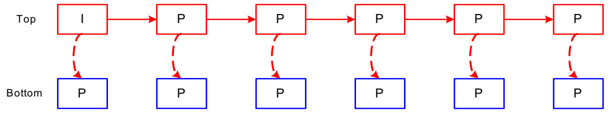


Fig.4: open-loop MC for key fields coding with one forward reference—TL_Haar1.

In this paper, the close-loop MC means that the motion compensation is completed with the predicted references which are reconstructed pictures from base layer, while the open-loop MC stands for employing the highest-quality reconstructed field pictures that can be attained. At encoder, "the highest-quality references" are the encoded pictures consisting of both the base layer and SNR enhancement layer data. However, at decoder, due to the limited transmission bandwidth, not all the information lies in the SNR enhancement layer can always be received by decoder, and then the references employed at decoder are different from the ones used at encoder, thus this mismatch of references will result in the drifting error. Fortunately, this kind of drifting propagation in our proposed interlaced FGS coding scheme can be greatly mitigated by performing close-loop MC in half of the temporal low-pass pictures (the other half are predicted in open-loop MC). This is because for the interlaced sequences, the temporal correlations between top and bottom fields in the same frames are fixed and larger than that in the progressive sequences, and our FGS coding method just make full use of this special character of interlaced video source.

With the close-loop MC controlling the propagation of drifting error, the adopted open-loop MC can improve the coding performance of both base layer and SNR enhancement layer. When at least one FGS layer appended on the base layer, the proposed method in figure 4 can acquire much better quality references and thus is able to more efficiently reduce the temporal redundancy between the top and bottom field pictures than Anchor1 and Anchor2 whose references are confined to base layer-quality ones. Consequently, the coding efficiency of base layer that encoded in open-loop MC is improved. Moreover, the advantage also lies in that up to 50% bits spent in coding SNR enhancement layers of key fields are saved due to the much lower energy of residual produced by open-loop MC.

The coding scenarios illustrated in figure 4, 5 and 6 are respectively noted as TL_Haar1, TL_Haar2 and TL_5/3 because it is the temporal low-pass pictures that operate open-loop motion compensation like using Haar or 5/3 wavelet kernel.

3.2.2. TL_Haar2

Extended from TL_Haar1, an alternative method named as TL_Haar2 is shown in figure 5: the top fields in even-indexed key frames and the bottom fields in odd-indexed key frames are linked together according to their time instants and coded as inter-P pictures with close-loop MC (solid line); and the remaining fields are coded as inter-P pictures but with open-loop MC (dashed line). This coding method can achieve comparable coding performance with TL_Haar1. Similarly to TL_Haar1 that once network bandwidth is limited, all of the bottom key pictures can be dropped without introducing error drifting propagation, TL_Haar2 can also improve temporal scalability by alternatively abandoning the top and bottom key fields coded in open-loop MC.

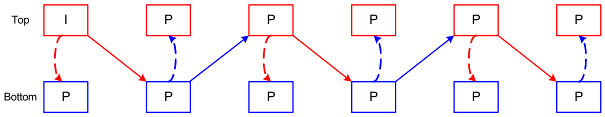


Fig.5: open-loop MC for key fields coding with one forward reference—TL_Haar2.

3.2.3. TL_5/3

So far, we have tried to reduce the temporal redundancy by introducing additional open-loop MC in TL_Haar1 and TL_Haar2. Considering the higher efficiency of Anchor2 that owns one more reference than Anchor1, we further extend TL_Haar2 to TL_5/3 which is shown in Figure 6: different from TL_Haar2, half of the P-fields are turned to be coded as inter-B pictures with open-loop MC and will become the high-pass pictures which can be abandoned when the transmission bandwidth is narrow enough. Generally speaking, replacing inter-P with B pictures will result in improved coding efficiency at the expense of increasing computation complexity. Therefore, TL_5/3 can not only efficiently improve the coding performance of base layer and SNR enhancement layer against Anchor2, but also provide an additional temporal scalability yielded by dropping B-fields in key frames.

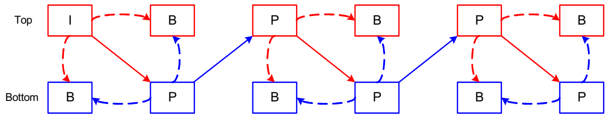


Fig.6: open-loop MC for key fields coding with two forward/backward references—TL_5/3.

However, considering that in TL_5/3, the bottom B-fields in even-indexed key frames will not be encoded or decoded until the P-fields in the following odd-indexed key frames encoded/decoded, which will result in a GOP delay unless the special bottom P-fields are marked to be unreferenced. We further simplify TL_5/3 as shown in Figure 7 that can avoid this extra delay.

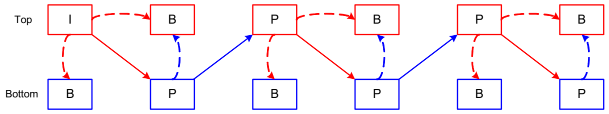


Fig.7: open-loop MC for key fields coding with two forward/backward—simplified TL_5/3.

3.3. Adaptive Quantization Parameter Selection

Usually, the quantization parameters (QP) are fixed to the same for both fields of a field pair. Whereas, in our proposed FGS coding scheme, the top or bottom key fields that to be low-pass pictures are coded with lower QP, and the fields to be high-pass pictures with higher QP than the initial QP. Experimental results show that this adaptive selection of QP serves well for helping improve the coding performance of SNR enhancement layer.

4. EXPERIMENTAL RESULTS AND ANALYSES

To evaluate the performance of our proposed FGS coding scheme for interlaced SVC, all the proposed coding scenarios, including TL_Haar1, TL_Haar2 and TL_5/3, have been integrated into interlaced SVC software [2]. And the results of Anchor1 and Anchor2 are tested with [2] under the same conditions. Simulations were carried out for various interlaced video sequences including the six common sequences for interlaced testing [4]. In this paper, we display the results of the following sequences including interlaced SD frames (720x576) Rugby and Formula1 @25Hz, and mobile @30HZ. The initial QP for base layer is set to 38, and for enhancement layer QP is decreased by step of 6. One bit-plane is appended for FGS enhancement layer coding. Only the first key field is coded as intra-picture, and the remaining key fields are coded either as P or B pictures with close-loop or open-loop MC decided by adopted methods. In order to fairly and clearly verify the advantages of our proposed FGS coding scenarios, we first encode the original sequences, for instance @30Hz with GOP size fixed to 16, and then extract and decode the encoded sequences separately @15Hz, 7.5Hz, 3.75Hz at different bit rates including their corresponding base layer's bit rate. In this case, the encoded sequences @3.75Hz comprise only the temporal low-pass (L3) and the lowest high-pass (H3) sub bands. Since our proposed improved FGS frameworks focus on the coding of temporal low-pass sub bands, for clarity, we mainly compare our proposed methods—TL_Haar1, TL_Haar2 and TL_5/3, with the straightforward implementations of FGS coding—Anchor1 and Anchor2 at low frequency.

As shown in figure 8(a), by adopting close-loop MC to prevent error drifting, around base layer bit-rate, TL_Haar1 and TL_Haar2 can achieve comparable or slightly worse coding performance against with Anchor1; and by incorporating open-loop MC, at middle and high bit-rate, they obviously outperform Anchor1 in PSNR performance. While figure 8(a) illustrates better performance for the case of one forward reference limited for coding key fields; figure 8(b) shows similar gain for the case of two forward/backward references allowed, that is TL_5/3 versus Anchor2. Moreover, in figure 9, all the coding methods are drawn together for overall comparisons. And interlaced SD sequence Rugby @25Hz encoded with GOP size set to 8 is also present in figure 10 to show the influence to the whole coding performance of both temporal low- and high-pass sub bands. More simulation results under different coding conditions can be found in [6].

It can be observed that in these curves, there is a common trend that the gain of our proposed FGS coding methods versus the original implementation in [2] is steadily increasing from low bit-rate to high bit-rate. Especially around the highest bit-rate, indicating that all the SNR enhancement layer data has been received, nearly 1 dB gain is achieved. As to the little loss near lowest bit-rate, which implies that except for the base layer data all other enhancement layer data is lost during the transmission, it is resulted from the extra error drifting generated by integrating SNR enhancement layer data into the prediction loop of base layer. Nevertheless, the loss is too small to be easily noticed subjectively; moreover, what the practical applications usually care about is the performance at the middle or high bit-rate instead of the lowest one. Another factor that influences the PSNR performance and the error resilience is the GOP size in the way that a small GOP size can better reduce the drifting errors that are easily accumulated around the base layer bit-rate than a big GOP size [5].

5. CONCLUSION

This paper proposes a novel FGS coding scheme for compressing interlaced sequences in the framework of SVC, which can efficiently improve the coding performance of both base layer and SNR enhancement layer, especially at middle and high bit rate, compared with the straightforward implementation of interlaced SVC coding. Moreover, the proposed TL_Haar1, TL_Haar2 and TL_5/3 can also provide an additional temporal scalability with low complexity to be suitable for practical applications.

6. REFERENCES

[1] J. Reichel, H. Schwarz, M. Wien, eds., "Scalable Video Coding – Working Draft 4," Joint Video Team, Doc. JVT-Q201, Poznan, Poland, July 2005.

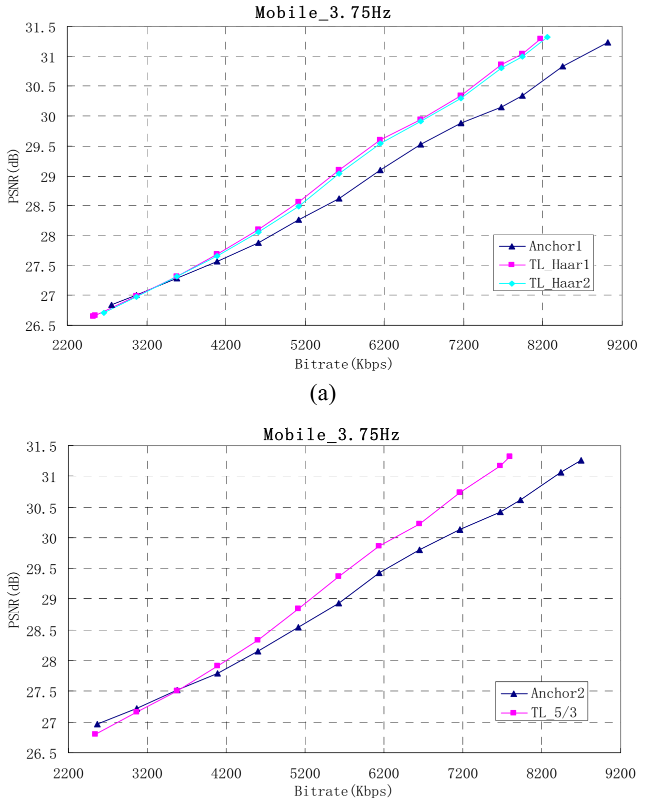
[2] V. Bottreau, J. Vieron, E. François, and Y. Chen, "Thomson contribution to the AHG SVC interlaced coding," Joint Video Team, Doc. JVT-R015, Bangkok, Thailand, January 2005.

[3] T. Hinz, H. Schwarz, and T. Wiegand, "FGS for field pictures and MBAFF frames," Joint Video Team, Doc. JVT-R061, Bangkok, Thailand, January 2006.

[4] ITU-Video Coding Experts Group (VCEG), Document VCEG-O59, "Core Experiment on Interlaced Video Coding," Pattaya, Thailand, 4-6 Dec., 2001.

[5] M. Schaar and H.Radha, "Adaptive Motion-Compensation Fine-Granular-Scalability for Wireless Video," IEEE Trans. Circuits Syst. Video Technol., vol. 12, pp. 360–371, June, 2002.

[6] Y. Zheng, X. Ji, and F. Wu, "FGS coding for interlaced SVC," Joint Video Team, Doc. JVT-R083r1, Bangkok, Thailand, January 2006.



(a)

(b)

Fig.8: RD curves of mobile @3.75Hz: (a) with one forward reference and (b) with two forward/backward references.

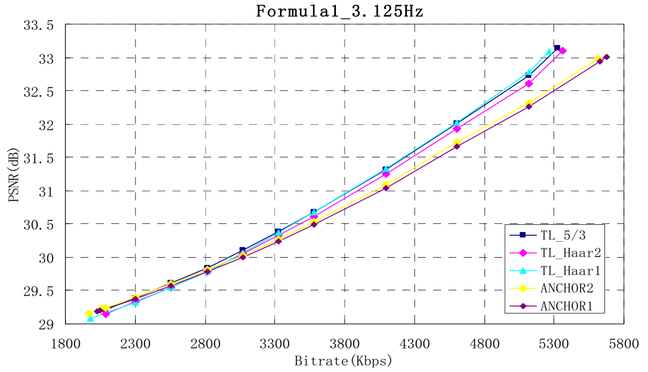


Fig.9: RD curves of Formula1@3.125Hz.

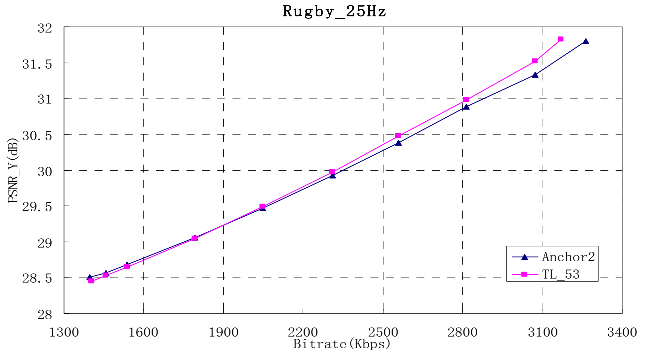


Fig.10: RD curves of rugby @25Hz.