H.264-Compatible Spatially Scalable Video Coding with In-band Prediction

Xin Jin*[1], Xiaoyan Sun[2], Feng Wu[2], Guangxi Zhu[1] and Shipeng Li[2]
[1]Dept. of Electronics & Inf., Huazhong Univ. of Sci. and Technol., Wuhan, China 430074
[2]Microsoft Research Asia, Beijing, China 100080
goldcamel2004@yahoo.com, {t-xysun, fengwu}@Microsoft.com, gxzhu@mail.hust.edu.cn, spli@Microsoft.com

Abstract—**In this paper, a H.264 compatible spatially scalable video coding method with in-band prediction is proposed which taking advantages from both the high coding efficiency of H.264 coding scheme and the attractive performance of in-band overcomplete discrete wavelet transform (ODWT) in wavelet-domain motion estimation and motion compensation. Four MV prediction modes are proposed for INTER prediction of high frequency subbands. The intra prediction modes of H.264 are also simplified for each high band according to the directional features inherited inside. Finally, a H.264 compatible scheme based on one of the MV prediction modes is presented to provide better tradeoff among standard compatibility, low complexity and high performance.**

Keywords-in-band prediction; spatial scalability; H.264 compatibility; video coding

I. INTRODUCTION

With the rapid development in heterogeneous end devices and networks, more and more users show great demands on enjoying multimedia services through various PC and non-PC devices over Internet or wireless connections. Such kind of ubiquitous multimedia services post great challenges to traditional coding techniques, such as H.264 [1] coding scheme. Responding to the new requirements, scalable coding schemes are emerging in multitude, which have drawn great attention in both industry and research areas.

Among several scalabilities provided by scalable coding methods, spatial scalability is one of the key features required in video streaming and sharing through various devices over heterogeneous network. It is feasible for spatially scalable coding to readily provide efficient video representation with different resolutions and different decoding complexity according to the actual connection speed and device capability.

Several wavelet-based video coding techniques have been developed in the past years [2-7]. Generally, these schemes can be classified into two categories: One is the 3D wavelet coding methods based on spatial domain motion compensation (MC) [2][3]; the other is the 3D wavelet coding or hybrid coding approaches based on in-band motion estimation (ME) and MC in wavelet domain [4-7]. Though the video coding schemes belonging to the first group are competitive in higher resolutions coding, it is hard for them to achieve good performance when coding lower resolutions. However, this problem can be solved by the in-band approaches. In [6], it is proved that, for spatially scalable coding, the in-band motion compensation based on overcomplete wavelet transform (ODWT) can yield similar performance to the full resolution MC approaches under comparable MC accuracy. Moreover it is much easier for in-band schemes to take advantages from and be compatible with the traditional coding methods.

In this paper, a H.264-compatible spatially scalable video coding method based on in-band prediction is proposed. The overcomplete ME and MC methods presented in [8] are utilized in the proposed scheme to avoid the shift-variant property of the wavelet transform for subband conversion from complete to overcomplete. The low frequency subband forms our low-resolution base layer, which is fully compatible with H.264. To the enhancement layer formed by high frequency subbands, adaptive and jointed INTER prediction with personalized INTRA prediction are investigated and presented for high coding efficiency and syntax compatibility.

This paper is organized as follows. The architecture of the proposed spatially scalable video coding system is first presented in section 2 with the related techniques focusing on INTER prediction and INTRA prediction. The compatibility with H.264 is discussed in section 3. The experimental results are shown in section 4 and section 5 concludes this paper.

II. THE PROPOSED VIDEO CODING SYSTEM

In this section, the proposed in-band spatially scalable video coding (SSVC) method is described in details. The architecture of the coding method is first presented. Then some improvements in the coding techniques of high frequency subbands are proposed.

A. System architecture and realization

Fig. 1 shows the architecture of proposed SSVC system with in-band prediction. As shown in the figure, every image of input video sequence is firstly down-sampled by spatial analysis employing one level discrete wavelet transform (DWT) of 5/3 linear phase FIR low pass and high pass filters. Four spatial subbands, one low pass subband LL and three high pass subbands LH, HL and HH, are generated. The block-based ME and MC are performed in the wavelet domain of each subband. Similar to H.264 coding method, the residuals of the four subbands are afterwards coded separately using 4x4

[*] This work has been done while the author is with Microsoft Research Asia

0-7803-9134-9/05/$20.00 ©2005 IEEE

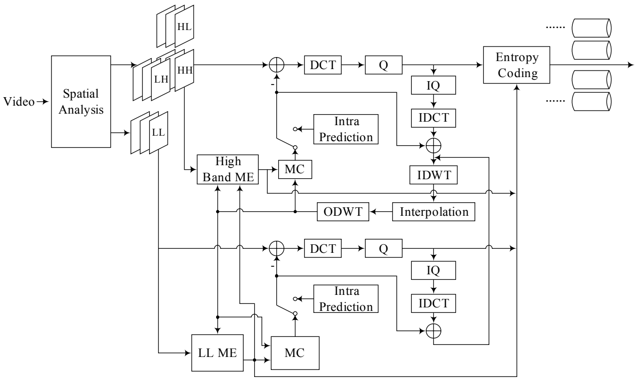


Fig. 1. Architecture of proposed SSVC scheme.

discrete cosine transform (DCT) and quantization. In order to improve the efficiency of entropy coding for discrete cosine transformed wavelet coefficients, context-adaptive binary arithmetic coding (CABAC) is used to generate the coded bit stream. The usage of adaptive codes of CABAC permits adaptation to nonstationary symbol statistics to match the variant distribution of high frequency subband coefficients [9].

In the proposed SSVC system, both INTER prediction and INTRA prediction are enabled for every subband. The prediction mode of each macroblock is determined by rate distortion optimized selection. The INTER prediction is performed in the wavelet domain using in-band block-matching ME and MC to achieve the spatial scalability. For higher accuracy in ME and MC, the reconstructed references of the four subbands are inversely DWT transformed and resulted in a reconstruction at full resolution on which the half-pixel interpolation with six-tap weiner filter [1] is performed. Then, the generated reconstruction at four times of full resolution is separated into four full-resolution reference images including one integer samples image and three fractional samples images.

The overcomplete discrete wavelet transform described in [7] is utilized in the proposed scheme. It is performed on those four full-resolution reference images of which the optimum interpolations are achieved at half-pixel positions. Finally, we interleave the interpolated overcomplete images and achieve four reference frames in quarter-pixel accuracy for the four subbands respectively. Based on these reference frames, the in-band ME and MC are performed in both low frequency band and three high frequency subbands. In the subbands' conversion from complete to overcomplete, low-band-shift (LBS) method proposed in [8] is used to overcome the shift-variant property, so that the efficiency of ME and MC is improved.

The coding techniques specified in H.264 can be readily introduced into each subband coding process, including the ME, MC, mode decision and so on. Among all the subbands, the similarity between LL band and the original image, together with the good performance of H.264, guarantee the efficiency of LL subband coding. On the other hand, for high frequency subbands which are quite different from the original image, it is hard to say that the techniques provided by H.264 are very suitable.

The INTRA prediction in H.264 was first investigated in high frequency subband coding. Though the correlation among the pixels inside is much weaker than that of LL band, test results still proved that enabling all nine intra prediction modes would benefit the overall performance of high frequency subbands.

To the efficiency of INTER prediction of high frequency subbands, evaluations were carried out by comparing the coding performance of entirely INTRA coding without INTER prediction with entirely INTER prediction. Results showed that, though there are kinds of randomness in pixel distribution in high frequency subbands, INTER prediction still could exploit the potential relativity among pixels between frames resulting in great contributions to coding efficiency. Furthermore, enabling INTRA prediction mode for INTER prediction frame (P frame) will achieve better performance by overcoming the prediction inaccuracy caused by vast moving and irregular distribution features.

Moreover, the mode decision method based on rate distortion optimized selection of H.264 is also an effective method for high bands coding. Experimental results demonstrated that multi-partition and multi-block mode selection do contribute to the coding efficiency.

Beside the above investigations in INTER and INTRA prediction, techniques improvements focusing on these two aspects will be discussed in the following based on the distribution properties analysis for high frequency subbands.

B. Improvements in high frequency subbands coding

1) Inter prediction

Four coding modes are proposed for high frequency subbands INTER prediction, as shown in TABLE I.

In the *spatial prediction* mode, the in-band ME and MC are preformed. The spatially motion vector prediction same as the one used in H.264 is enabled. This mode works well in case there is still some meaningful motion remaining in the high frequency bands.

In the second mode, namely *zero prediction*, ME and MC are performed in each subband independently. Moreover, motion vector (MV) prediction is disabled in the ME process which means that the motion search is centered on the relative {0, 0} position. It is based on the assumption that for some blocks, it is lack of motion correlation among adjacent blocks in the high frequency subbands.

The third mode is *LL prediction*. Instead of using spatial or zero as MV predictor, this mode performs the motion search by referring to the MV of LL band which is regarded as the MV predictor. The correlations between high bands and LL band are exploited in this mode.

TABLE I. MV PREDICTION MODES FOR HIGH BANDS CODING

Mode	Predictor	Refinement	Final MV
Spatial Pred.	Neighboring	Quarter Pel	Refined MV
Zero Pred.	0,0		
LL Pred.	LL MV		
LL MV	No prediction	No refine	LL MV

The last one is *LL MV*, which means that the MV used in the high frequency subband coding is inherited from the LL band rather than obtained by in-band ME. This mode can greatly save the bits in MV coding.

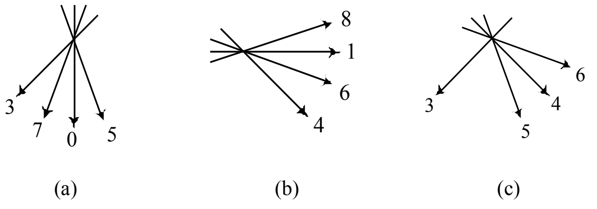
In other words, *Spatial prediction* mode represents the moving relevance inside each high frequency band. The *zero prediction* mode reflects the irregular moving characteristics of high bands. The *LL prediction* mode and the *LL MV* mode show the strong motion correlation between high bands and low band.

The mode decision method presented in H.264 is utilized to select the proper coding mode. Using these four kinds of motion modes adaptively in the high bands motion estimation, the proposed scheme can achieve average 0.6 dB gain in PSNR.

By further investigating on the four MV prediction modes shown in TABLE I, we found that it is possible to reduce the encoding complexity of the proposed INTER prediction method. Adaptation between merely two modes, *LL MV* and *Spatial prediction*, can maintain close performance to the 4-mode scheme. Moreover, if the encoding complexity is really an issue for certain application, the *LL MV* mode only used can still provide 0.3 dB gain averagely in PSNR, meanwhile enable the H.264 syntax compatibility.

2) Intra prediction

Since the wavelet decomposition can be interpreted as signal decomposition in a set of independent, spatially oriented frequency channels, the high frequency wavelet subbands can be treated as directed ones. That means LH band will represent the horizontal high frequencies, HL band will represent the vertical high frequencies and HH band will represent the high frequencies in diagonal directions. Correspondingly, they show the vertical, the horizontal and the diagonal edges of one frame. So, based on statistical analysis on the directional representation property of the three high bands, we reduced the nine INTRA prediction directions originally in H.264 into three subsets of 5 directions. Except of the DC direction, other directions used in INTRA prediction for each high subband are shown in Fig. 2. The proposed INTRA prediction scheme saves 30% intra mode bits in the case of similar coding performance.



(a) (b) (c)

Fig. 2. Intra prediction direction subset for: (a) LH band; (b) HL band and (c) HH band.

III. H.264 Compatibility

In our spatially scalable coding system, the low frequency subband forms our low-resolution base layer which using the original techniques specified in H.264 for predicting and coding. The low-resolution base layer of our system is completely compatible with H.264 standard and could be decoded independently.

To enhancement layer, which refers to three high frequency subbands, the compatibility could be exploited with good balancing between performance and complexity. As mentioned before, motion estimation and compensation performed independently in each high frequency subband possesses standard compatibility while unsatisfied performance. Adaptive motion estimation and compensation based on four MV prediction modes could achieve the higher gain in PSNR while loss the compatibility with H.264 because of the syntax modification for each partition of every block mode. Fortunately, we can use only *LL MV* mode to enable the compatibility in INTER prediction for high frequency subbands which still shows benefit to the coding performance.

For INTRA prediction, since the subset of the prediction modes is utilized, it can be readily compatible with H.264 INTRA mode coding.

The base layer and enhancement layer are arranged in four slices including one low frequency slice and three high frequency slices. We use slice ID from zero to three to represent LL, LH, HL and HH subbands.

IV. Experimental Results

Fig. 3 shows the coding performance of the proposed spatially scalable video coding system. The test results presented below are performed on CIF “Bus” video sequence. A single bit stream has been encoded at 30 frame/s (fps). The quantization parameters allocated to LH, HL and HH are three, four and five lager than LL’s. To demonstrate the efficiency of our improved techniques in high frequency subbands, the system performances of CIF are presented in Fig. 3.

The representations of curves shown on the figure are explained in TABLE II. The curve order is from the lowest performance to the highest corresponding to scheme 1 with worst performance to scheme 5 achieved the best gain.

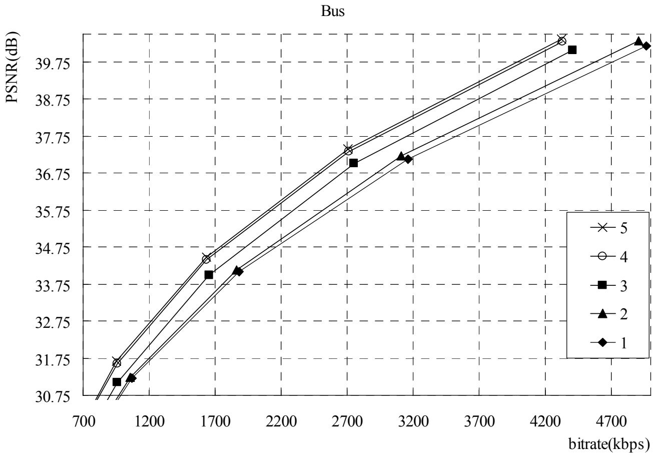


Fig. 3. Performance comparison of several attempts.

TABLE II. CURVE REPRESENTATIONS IN FIG.3

Curve	**Representation**		
	INTER prediction		***INTRA prediction***
	Refinement	***Motion Mode***	
1	Quarter-pel refine	*Spatial*	disable
2			enable
3	No refine	*LL MV*	
4	Quarter-pel refine	*Spatial and LL MV*	
5		*Four modes adapt*	

By making a good tradeoff among H.264 compatibility, low encoding complexity and high coding performance, the scheme 3 is recommended. Compared with the second approach, scheme three has over 0.3dB average gain in PSNR.

V. CONCLUSION

In this paper, a H.264 compatible spatially scalable video coding system is proposed based on in-band overcomplete motion estimation and motion compensation. Both block-based inter prediction and intra prediction are enabled for low resolution base layer and three high frequency subband enhancement layers. By investigating the pixel distribution properties and correlations among and inside blocks, four MV prediction modes are proposed for INTER prediction of high frequency subbands. Moreover, the intra prediction modes of H.264 are simplified for each high band according to the directional features inherited inside, which results in 30% saving in intra mode bits. Finally, a scheme based on one of the MV prediction mode, *LL MV* is recommended to enable H.264 compatibility while provide better tradeoff among standard compatibility, low complexity and high performance.

REFERENCES

[1] Draft ITU-T Recommendation and Final Draft International Standard of Joint Video Specification (ITU-T Rec. H.264/ISO/IEC 14 496-10 AVC), Mar. 2003.

[2] Jens-Rainer Ohm, “Three-dimensional subband coding with motion compensation,” IEEE Trans. Image Processing, vol. 3, pp.559-571, September 1994.

[3] Seung-Jong Choi and John W. Woods, “Motion-compensated 3-D subband coding of video,” IEEE Trans. Image Processing, vol.8, pp.155-167, February 1999.

[4] Marek Domanski, Adam Luczak, and Slawomir Mackowiak, “Spatio-temporal scalability for MPEG video coding,” IEEE Trans. Circuits Syst. Video Technol., vol.10, pp. 1088-1093, October 2000.

[5] Y. Andreopoulos, A. Munteanu, G. Van der Auwera, P. Schelkens and J. Cornelis, “Scalable wavelet video-coding with in-band prediction-implementation and experimental results,” in Proc. IEEE Int. Conf. Image Processing, 2002, pp. III-729-732.

[6] Y. Andreopoulos, M. van der Schaar, A. Munteanu, etal, “Fully-scalable wavelet video coding using in-band motion compensated temporal filtering,” in Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing, 6-10 April 2003, vol.3, pp. III - 417-20.

[7] Claudia Mayer, “Motion compensation in-band prediction for wavelet-based spatially scalable video coding,” in Proc. IEEE Int. Conf. Acoustics Speech Signal Processing, 2003, pp. III - 73-6.

[8] Hyun-Wook Park and Hyung-Sum Kim, “Motion estimation using low-band-shift method for wavelet-based moving-picture coding,” IEEE Trans. Image Processing., vol. 9, pp. 577-587, April, 2000.

[9] Thomas Wiegand, Gary J. Sullivan, Gisle Bjontegaard, and Ajay Luthra, “Overview of the H.264/AVC video coding standard,” IEEE Trans. Circuits Syst. Video Technol., vol.13, pp. 560-576, July 2003.