SCALABLE VIDEO CODING USING ALLPASS-BASED WAVELET FILTERS

Xi Zhang *Takuya Suzuki*

Department of Communication Engineering and Informatics
The University of Electro-Communications
Chofu-shi, Tokyo, 182-8585 Japan
E-mail: zhangxi@uec.ac.jp, gtr-taku@uec.ac.jp

ABSTRACT

Scalable video coding (SVC) allows to encode videos in a unique bit stream at several operating points corresponding to different quality, spatial resolution and temporal frame rate. In this paper, we apply the allpass-based wavelet filters to wavelet-based scalable video coding (WSVC) for improving the coding performance, in particular, at reduced spatial resolutions. The allpass-based wavelet filters used in this paper are orthogonal and symmetric, and have a better frequency response than the D-5/3 and D-9/7 biorthogonal wavelet filters. Therefore, some spatial aliasing effects can be suppressed, when the allpass-based wavelet filters are used as a spatial transform in WSVC. Moreover, the inverse MCTF transform is performed at full spatial resolution to avoid the imprecision of the deduced MVs. Finally, a coding performance comparison between the proposed allpass-based and conventional D-9/7 biorthogonal wavelet filters is shown in terms of coding efficiency.

***Index Terms*—** Scalable video coding, wavelets, allpass filter, IIR filter, MCTF, lifting scheme

1. INTRODUCTION

Scalable video coding (SVC) has a wide range of applications, such as internet video, mobile wireless video, and so on [7]∼[12]. By using SVC, a number of desirable streams can be extracted from a single bit stream corresponding to various operating points in terms of image quality, spatial resolution and temporal frame rate. SVC techniques can be classified mostly as wavelet-based SVC (WSVC) and hybrid SVC used in MPEG-x or H.26x video coding standards [9]. Although hybrid SVC based on spatial DCT and temporal ME/MC (Motion Estimation/Motion Compensation) has provided a coding performance improvement, WSVC is an attractive technique too, because wavelet transform naturally provides a scalable multidimensional (space-time-quality) signal representation. There are two schemes in WSVC; t+2D and 2D+t, depending on the space-time order in which the wavelet transforms operate. t+2D scheme performs a motion compensated temporal filtering (MCTF) in the original spatial domain followed by a spatial transform on each temporal subband, while the spatial transform is applied before the temporal one in 2D+t scheme. 2D+t scheme had demonstrated lower coding efficiency compared to t+2D one, especially at higher resolutions, due to the reduced efficiency of MCTF when applied in the spatial highpass subband domain [9]. Therefore, we will discuss only t+2D scheme in this paper. t+2D scheme is the most intuitive way to build a WSVC system, however, it has some relevant drawbacks, especially for spatial scalability performance, due to spatial aliasing introduced by non-ideal wavelet filters, e.g., the D-5/3 and D-9/7 biorthogonal wavelet filters. Therefore, it is required to design more selective wavelet filters to reduce the spatial aliasing [8], [11].

In this paper, we firstly introduce a class of orthogonal symmetric wavelet filters using IIR allpass filters proposed in [3] and [4]. The allpass-based orthogonal symmetric wavelet filters have a better (more selective) frequency response than the D-5/3 and D-9/7 biorthogonal wavelet filters, thus, aliasing introduced by the decimation can be reduced. The allpass-based wavelet filters have been applied to lossy and lossless image coding, and demonstrated a superior coding performance in [5] and [6]. We then use the allpass-based wavelet filters as a spatial transform in WSVC to suppress some spatial aliasing effects for improving the coding efficiency, in particular, at reduced spatial resolutions. Moreover, we perform the inverse MCTF transform at full spatial resolution to avoid the imprecision of the deduced MVs. Finally, a coding performance comparison between the allpass-based and D-9/7 wavelet filters is given in terms of coding efficiency. It is shown from the experimental results that the allpass-based orthogonal symmetric wavelet filters outperform the conventional D-9/7 biorthogonal wavelet filter.

2. WAVELET-BASED SCALABLE VIDEO CODING

The general framework of WSVC is shown in Fig.1, which is usually referred as to 2D+t+2D scheme. A video signal is firstly decomposed by pre-spatial 2D DWT (Discrete Wavelet Transform), then by temporal DWT including ME/MC (Motion Estimation/Motion Compensation), that is, MCTF (Motion Compensated Temporal Filtering), followed by post-spatial 2D DWT. t+2D and 2D+t schemes can be implemented without pre- or post-spatial 2D DWT. Temporal transform implements a framewise motion compensated lifting wavelet transform, and generally uses the Haar or D-5/3 wavelet filter depending whether motion modes are unidirectional (forward or backward) or bidirectional [12]. The lowpass temporal filtering updates the corresponding temporal subbands with ghosting artifacts due to unavoidable local MC failures. The update step in the lifting scheme can be also omitted to eliminate ghosting artifacts. A popular wavelet filter for lossy image coding is the D-9/7 biorthogonal wavelet filter. This FIR filter is generally used as a spatial transform

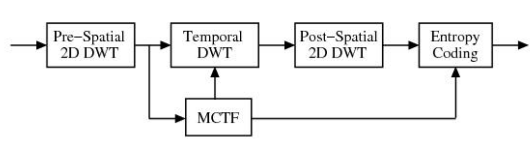

Fig. 1. General framework of WSVC.

This work was supported by JSPS KAKENHI Grant Number 22500083.

in WSVC. The D-9/7 biorthogonal wavelet filter has a symmetric impulse response (exactly linear phase), but its magnitude response is not good enough as a downsampling filter [8], [11]. Thus, spatial aliasing is introduced by the decimation and non-ideal frequency response of wavelet filters. Spatial aliasing degrades spatial scalability performance in t+2D scheme and reduces the efficiency of MCTF when applied in the spatial highpass subband domain of 2D+t scheme [9]. Therefore, a more selective wavelet filter is needed to reduce these spatial aliasing effects, as described in [8] and [11].

3. T+2D SCHEME USING THE ALLPASS-BASED WAVELET FILTERS

In this section, we propose the t+2D scheme of WSVC using the allpass-based orthogonal symmetric wavelet filters. The allpass-based wavelet filters can be also used to improve the coding efficiency of 2D+t scheme, although it had demonstrated lower coding efficiency compared with t+2D one, especially at higher resolutions, due to the reduced efficiency of MCTF [9]. In the following, we firstly describe the allpass-based orthogonal symmetric wavelet filters proposed in [4], and then used the wavelet filters as a spatial transform in WSVC.

3.1. Allpass-based Orthogonal Symmetric Wavelet Filters

A class of orthogonal symmetric wavelet filters has proposed in [4] by using allpass filters as follows;

$$\begin{cases} H(z) = \frac{1}{2}[A(z^2) + z^{-2K-1}A(z^{-2})] \\ G(z) = \frac{1}{2}[A(z^2) - z^{-2K-1}A(z^{-2})] \end{cases}, \tag{1}$$

where K is integer, and $A(z)$ is an IIR allpass filter of order N defined by

$$A(z) = z^{-N}\frac{\sum_{n=0}^{N} a_n z^n}{\sum_{n=0}^{N} a_n z^{-n}}, \tag{2}$$

where a_n are a set of real-valued coefficients and $a_0 = 1$.

Assume that $\theta(\omega)$ is the phase response of $A(z)$, then we have the frequency responses of $H(z)$ and $G(z)$ as

$$\begin{cases} H(e^{j\omega}) = e^{-j(K+\frac{1}{2})\omega}\cos\{\theta(2\omega) + (K+\frac{1}{2})\omega\} \\ G(e^{j\omega}) = je^{-j(K+\frac{1}{2})\omega}\sin\{\theta(2\omega) + (K+\frac{1}{2})\omega\} \end{cases}. \tag{3}$$

It is clear in Eq.(3) that $H(z)$ and $G(z)$ have an exactly linear phase response and satisfy the following power-complementary relation;

$$|H(e^{j\omega})|^2 + |G(e^{j\omega})|^2 = 1. \tag{4}$$

That is, this class of wavelet filters is orthogonal and symmetric.

For $H(z)$ and $G(z)$ to be a pair of lowpass and highpass filters, the phase response of $A(z)$ must satisfy

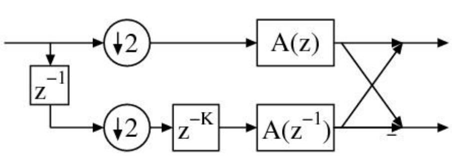

Fig. 2. Orthogonal symmetric wavelet filters using allpass filters.

$$\theta(2\omega) + (K+\frac{1}{2})\omega = \begin{cases} 0 & (0 \le \omega \le \omega_p) \\ \pm\frac{\pi}{2} & (\omega_s \le \omega \le \pi) \end{cases}, \tag{5}$$

where ω_p and ω_s are the cutoff frequencies in passband and stopband of $H(z)$, respectively, and $\omega_p + \omega_s = \pi$.

Due to the antisymmetry property of the phase response, the desired phase response of $A(z)$ can be reduced to

$$\theta_d(\omega) = -(\frac{K}{2} + \frac{1}{4})\omega \qquad (0 \le \omega \le 2\omega_p). \tag{6}$$

Therefore, the design problem of the allpass-based orthogonal symmetric wavelet filters shown in Fig.2 becomes the phase approximation of $A(z)$ to the desired phase response in Eq.(6). The existing design methods of allpass filters, for example, the maximally flat and equiripple approximations, can be used to design this class of orthogonal symmetric wavelet filters. In [4], a design method has been proposed for the allpass-based orthogonal symmetric wavelet filters with the specified degree of flatness, thus $H(z)$ and $G(z)$ have $2L+1$ zeros at $z=-1$ and $z=1$ respectively, where L is a parameter that controls the degree of flatness in the magnitude response, and $0 \le L \le N$. If $L = N$, then the filter is the maximally flat (Maxflat) one. When $L < N$, the remaining degree of freedom in the design is used to minimize the magnitude error in the approximation band.

3.2. MC-EZBC Scalable Video Coder

In this paper, we use the MC-EZBC scalable video coder proposed in [12] to implement the t+2D scheme of WSVC. The t+2D scheme performs MCTF directly in the original spatial domain, followed by a spatial transform on each temporal subband. The block-based ME/MC is implemented via a quadtree structure. To reduce the spatial aliasing effects, we use the allpass-based orthogonal symmetric wavelet filters as the spatial transform, instead of the conventional D-9/7 wavelet filter.

We have used the design method proposed in [4] to examine the performance of these wavelet filters. The magnitude responses of the obtained wavelet filters with $N=2$ and $K=3$ are shown in Fig.3. It is seen in Fig.3 that the Maxflat wavelet filter with $L=N=2$ has a better magnitude response than the D-9/7 biorthogonal wavelet filter. The magnitude responses of other two filters with $L=1$ are also shown in Fig.3, where the passband cutoff frequency is set to $\omega_p = 0.360\pi$ and $\omega_p = 0.414\pi$, respectively. The wavelet filter with $\omega_p = 0.414\pi$ has the sharpest transition band compared with other filters, but has a larger ripple in the stopband.

It is known that aliasing is caused by non-ideal magnitude response of wavelet filters. Spatial aliasing degrades spatial scalability

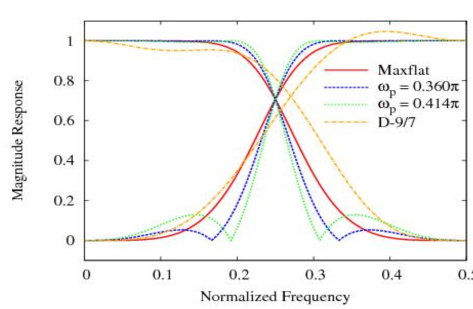

Fig. 3. Magnitude responses of wavelet filters with N=2.

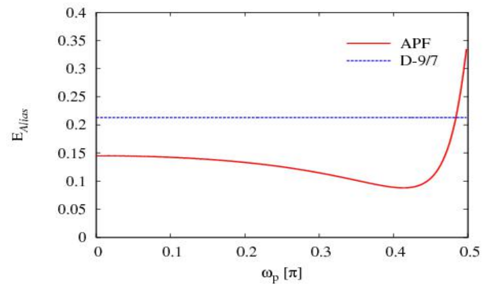

Fig. 4. Relation between cutoff frequency and aliasing energy.

performance and reduces the efficiency of MCTF at lower spatial resolutions. An desirable characteristic is that $H(z)$ and $G(z)$ have no overlap in the magnitude response, but it is impossible to realize it. To evaluate the aliasing effect, we define an aliasing energy of wavelet filters as follows;

$$E_{Alias} = \int_0^{\pi} |H(e^{j\omega})G(e^{j\omega})|^2 d\omega, \tag{7}$$

where E_{Alias} is zero for an ideal wavelet filter pair without a magnitude overlap. We have calculated the aliasing energy E_{Alias} for the wavelet filters with $N = 2$, $L = 1$ and $K = 3$, where the passband cutoff frequency is varied from $\omega_p = 0$ to $\omega_p = 0.5\pi$. Note that $\omega_p = 0$ is correspondent to the Maxflat wavelet filter. The result is shown in Fig.4, and it is seen that if $\omega_p < 0.48\pi$, the allpass-based wavelet filter has a smaller aliasing energy E_{Alias} than the D-9/7 biorthogonal wavelet filter. When $\omega_p = 0.414\pi$, we can obtain a wavelet filter with the smallest aliasing energy, whose E_{Alias} is less than half that of the D-9/7 wavelet filter.

In this paper, we use the allpass-based wavelet filter with $N = 2$ and $K = 3$. This filter requires only 4 multipliers and 10 adders, whereas the D-9/7 wavelet filter implemented by the lifting scheme has 6 multipliers and 8 adders.

3.3. The Inverse MCTF at Full Resolution in Decoder

The t+2D scheme performs MCTF directly in the original spatial domain, thus, the estimated motion vectors (MVs) are optimal at full spatial resolution. However, it has some relevant drawbacks, especially for spatial scalability performance. When a lower spatial resolution is needed, the inverse MCTF transform is usually implemented at reduced spatial resolution in decoder, where MVs are typically deduced from the MVs estimated at full resolution by the down-conversion, and an interpolation filter is needed to calculate some subpixels. Due to spatial aliasing introduced by the non-ideal wavelet filters, the deduced MVs cannot be guaranteed to be optimal at the target resolution, resulting in the reduced coding efficiency at lower spatial resolutions.

Although the allpass-based wavelet filter has been used to reduce some spatial aliasing effects, it is still needed to avoid the imprecision of the deduced MVs. In this paper, we do not perform the inverse MCTF transform at reduced spatial resolution. Firstly, we do up-sampling at reduced spatial resolution and filtering it by a lowpass filter to get full spatial resolution. Then we perform the inverse MCTF transform at the full spatial resolution where the MVs are estimated. Finally, the target spatial resolution is obtained by decimation (lowpass-filtering and down-sampling). Note that those MVs at reduced spatial resolution are not needed. In this paper, the lowpass filter $H(z)$ in the allpass-based wavelet filters has been used for down-sampling and up-sampling.

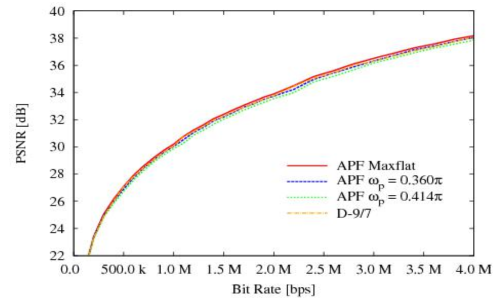

Fig. 5. PSNR comparison of test sequence Bus at CIF resolution with different wavelet filters.

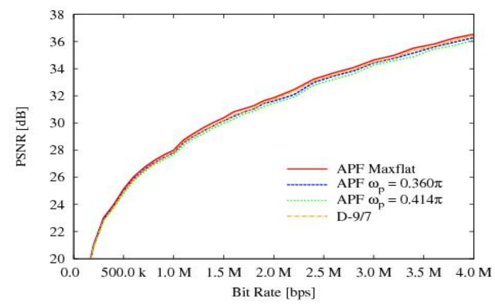

Fig. 6. PSNR comparison of test sequence Mobile at CIF resolution with different wavelet filters.

4. EXPERIMENTAL RESULTS

In this section, we present some experimental results and comparison with the D-9/7 biorthogonal wavelet filter. Two test sequences Bus and Mobile (CIF, 30fps) have been used to evaluate the coding performance. Three allpass-based wavelet filters of $N = 2, L = 1, K = 3$ with $\omega_p = 0$, $\omega_p = 0.360\pi$, and $\omega_p = 0.414\pi$, as shown in Fig.3, have been used in the t+2D scheme. The filter with $\omega_p = 0$ is the Maxflat filter, and that with $\omega_p = 0.414\pi$ has the smallest aliasing energy E_{Alias}. In addition, the filter with $\omega_p = 0.360\pi$ is chosen as a trade-off between the aliasing energy and stopband ripple, having a slightly larger E_{Alias} and smaller stopband ripple, compared with that of $\omega_p = 0.414\pi$.

Firstly, the PSNR comparison at full CIF resolution for different wavelet filters is given in Fig.5 and Fig.6, respectively. It is seen that the Maxflat filter has the best coding performance at full resolution. Then, the PSNR comparison of different MVs at lower QCIF resolution is given in Fig.7 and Fig.8. It is clear that the coding performance when MVs were implemented at full resolution is much better than when MVs were deduced at lower resolution for all wavelet filters. Finally, the PSNR comparison at lower QCIF resolution for different wavelet filters is shown in Fig.9 and Fig.10, where the inverse MCTF was imlemented at full resolution. The coding performance of the allpass-based wavelet filters are better than the D-9/7 wavelet filter at lower resolution, and the filter with $\omega_p = 0.414\pi$ has the best coding performance.

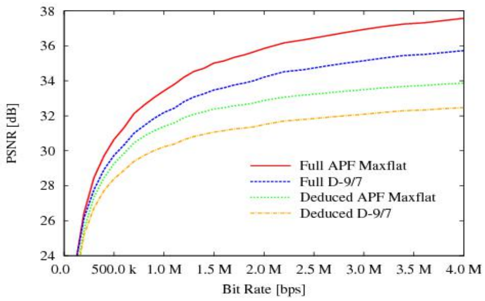

Fig. 7. PSNR comparison of test sequence Bus at QCIF resolution with different MVs.

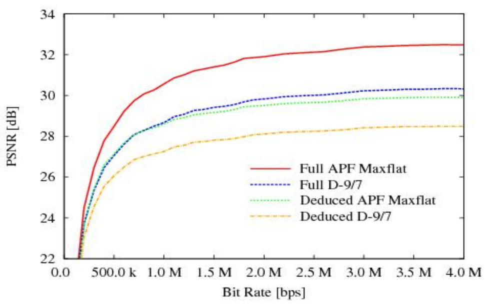

Fig. 8. PSNR comparison of test sequence Mobile at QCIF resolution with different MVs.

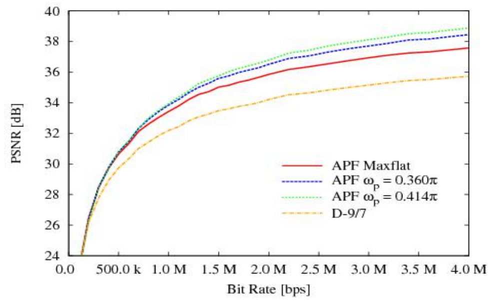

Fig. 9. PSNR comparison of test sequence Bus at QCIF resolution with different wavelet filters.

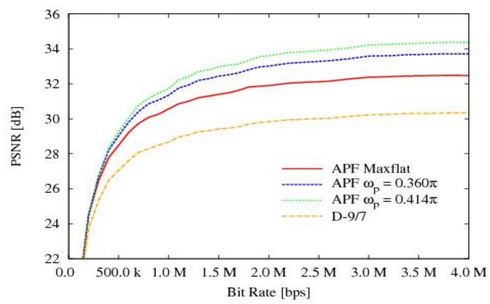

Fig. 10. PSNR comparison of test sequence Mobile at QCIF resolution with different wavelet filters.

5. CONCLUSION

In this paper, we have introduced a class of orthogonal symmetric wavelet filters using allpass filters. The allpass-based orthogonal symmetric wavelet filters have a better magnitude response than the D-9/7 biorthogonal wavelet filter, thus, some spatial aliasing effects can be reduced. The allpass-based wavelet filters have been applied to lossy and lossless image coding, and demonstrated a superior coding performance. In this paper, we have used the allpass-based wavelet filters as the spatial transform in WSVC to improve the coding efficiency, in particular, at reduced spatial resolutions. Moreover, we have performed the inverse MCTF transform at full spatial resolution to avoid the imprecision of MVs. Finally, a comparison in the coding efficiency has been given between the allpass-based and D-9/7 wavelet filters. It is shown from the experimental results that the allpass-based orthogonal symmetric wavelet filters outperform the conventional D-9/7 biorthogonal wavelet filter.

6. REFERENCES

[1] I. Daubechies, *Ten Lectures on Wavelets*, Philadelphia, PA: SIAM, 1992.

[2] M. Vetterli and J. Kovacevic, *Wavelets and Subband Coding*, Englewood Cliffs, NJ: Prentice Hall, 1995.

[3] C. Herley and M. Vetterli, “Wavelets and recursive filter banks,” IEEE Trans. Signal Process., vol.41, no.8, pp.2536–2556, Aug. 1993.

[4] X. Zhang, T. Muguruma and T. Yoshikawa, “Design of orthonormal symmetric wavelet filters using real allpass filters,” Signal Processing, vol.80, no.8, pp.1551–1559, Aug. 2000.

[5] S. Kamimura, X. Zhang and T. Yoshikawa, “Wavelet-based image coding using allpass filters,” IEICE Trans.(A), vol.J84-A, no.1, pp.13–21, Jan. 2001, (in Japanese); Electronics and Communications in Japan, parts 3, vol.85, no.2, pp.13–21, Feb. 2002. (in English).

[6] X. Zhang, K. Kawai, T. Yoshikawa and Y. Takei, “Lossy to lossless image compression using allpass filters,” in Proc. Int. Conf. Image Process. (ICIP), vol.1, pp.557–560, Genova, Italy, Sep. 2005.

[7] G. Pau, C. Tillier, B. Pesquet-Popescu and H. Heijmans, “Motion compensation and scalability in lifting-based video coding,” Signal Process.: Image Commun., vol.19, no.7, pp.577–600, Aug. 2004.

[8] M. Li and T. Nguyen, “Optimal wavelet filter design in scalable video coding,” in Proc. Int. Conf. Image Process. (ICIP), vol.1, pp.473–476, Genova, Italy, Sep. 2005.

[9] N. Adami, A. Signoroni and R. Leonardi, “State-of-the-art and trends in scalable video compression with wavelet-based approaches,” IEEE Trans. Circuits Syst. Video Technol., vol.17, no.9, pp.1238–1255, Sep. 2007.

[10] R. Q. Xiong, J. Z. Xu, F. Wu and S. P. Li, “Barbell-lifting based 3-D wavelet coding scheme,” IEEE Trans. Circuits Syst. Video Technol., vol.17, no.9, pp.1256–1269, Sep. 2007.

[11] Y. J. Wu and J. W. Woods, “Aliasing reduction via frequency roll-off for scalable image/video coding,” IEEE Trans. Circuits Syst. Video Technol., vol.18, no.1, pp.48–58, Jan. 2008.

[12] Y. J. Wu, K. Hanke, T. Rusert and J. W. Woods, “Enhanced MC-EZBC scalable video coder,” IEEE Trans. Circuits Syst. Video Technol., vol.18, no.10, pp.1432–1436, Oct. 2008.