Scalable Interframe Wavelet Coding with Low Complex Spatial Wavelet Transform

Wonha Kim, Seyoon Jeong, and Kyuheon Kim

In the decoding process associated with interframe wavelet coding, the inverse wavelet transform requires high computational complexity. However, as video technology starts to pervade all aspects of our lives, decoders are becoming required in various devices such as PDAs, notebooks, PCs, and set-top boxes. Therefore, a decoder's complexity needs to be adapted to the processor's computational power, and consequently a low-complexity codec is also required for scalable video coding. In this paper, we propose a method of controlling and lowering the complexity of the spatial wavelet transform while sustaining the same coding efficiency as that currently afforded. In addition, the proposed method may alleviate the ringing effect for slowly changing image sequences.

Keywords: Interframe wavelet coding, scalable video coding.

Manuscript received May 19, 2005; revised Jan. 23, 2006.

This work was supported by Grant No. R01-2005-000-11054 from Korea Science and Engineering Foundation in the Ministry of Science & Technology, Korea.

Wonha Kim (phone: + 82 31 201 2030, email: wonha@khu.ac.kr) is with Department of Electronics Engineering, Kyunghee University, Gyeonggi-do, Korea.

Seyoon Jeong (email: jsy@etri.re.kr) is with Broadcasting Media Research Group, ETRI, Daejeon, Korea.

Kyuheon Kim (email: kyuheonkim@khu.ac.kr) is with Department of Electronics and Information, Kyunghee University, Gyeonggi-do, Korea.

I. Introduction

Today's communication environment is becoming heterogeneous, using various types of communication devices and simultaneously accessing different types of communication networks. Furthermore, this trend is to intensify as communication and broadcasting merge. To achieve digital convergence on these heterogeneous networks, the next-generation video codecs must be scalable so as not only to produce maximum coding performance, but also to adaptively cope with such heterogeneous network environments [1].

Interframe wavelet coding is a scalable video coding method. In this method, video frames are first temporally filtered, a procedure called motion compensated temporal filtering (MCTF), and then spatially transformed by means of a wavelet, with temporal filtering and the wavelet transform providing temporal scalability and spatial scalability, respectively [2]-[5]. The wavelet transform coefficients of each frame are quantized and entropy coded to produce a coded bit stream. Conventional schemes apply the same wavelet filter to all temporally filtered frames.

In this paper, we propose a scheme in which different wavelet filters are applied to temporally filtered frames, depending on characteristics of the temporally filtered frames. For this purpose, we use coding gain to analyze the properties of the temporally filtered frames [5], [6]. We derive the unified coding gain, which works for any structure of subband decompositions and any kind of finite impulse response (FIR) filter. Based on this unified coding gain, we determine the characteristics of each temporally filtered frame in order to select the optimal wavelet filters. By investigating the correlation between coding gain and motion behaviors, we also propose a practical method of determining the characteristics of the temporally filtered frames that obviates the need to measure the coding gain. Compared to the

conventional schemes, the proposed method greatly reduces the computational complexity of the spatial wavelet transform while maintaining the coding efficiency and alleviating the ringing effect for slowly changing image sequences.

The remainder of this paper is organized as follows: In section II, the procedure of MCTF is briefly explained and the properties of MCTF frames are investigated. In section III, a unified coding gain is derived and the MCTF frames are analyzed in terms of coding gain. In section IV, the scheme utilized adaptively to choose the wavelet filters according to the characteristics of MCTF frames is described. Experiments on the proposed schemes are described in section V. Finally, the conclusions are drawn in section VI.

II. Motion Compensated Temporal Filtering (MCTF)

1. Procedure of MCTF

The goal of MCTF is to reduce the temporal redundancy among image sequences within a group of pictures (GOP) by condensing the overall information into one frame called the L-frame and distributing the residual images among the other frames, which are referred to as H-frames. MCTF creates low-pass frames, or L-frames, by averaging the best matched pixels among consecutive frames, while simultaneously creating the high-pass frame, or H-frame, by filtering out matched pixels among consecutive frames. This procedure can be performed repeatedly on the L-frames in the form of a pyramidal decomposition in order to enhance the coding efficiency. Thus, H-frames correspond to B- or P-frames of the MPEG format, whereas L-frames correspond to I-frames and store most of the image information [7], [8].

The MCTF schemes mainly used are Haar-MCTF and 5/3-MCTF. Haar-MCTF performs unidirectional motion estimation and produces one L-frame and one H-frame from each pair of input frames. The 5/3-MCTF employs bidirectional motion estimation and thus produces one H-frame from three input frames and one L-frame from five input frames. Haar-MCTF and 5/3-MCTF can be used together in the same GOP. Figure 1 depicts the procedure used in 5/3-MCTF.

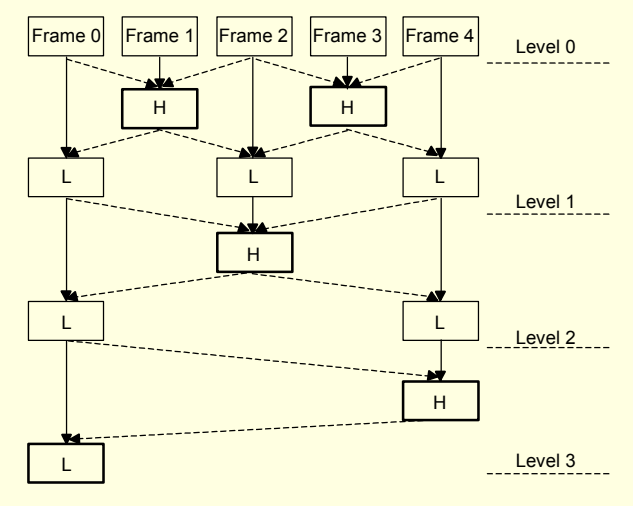


Fig. 1. Procedure of 3-level 5/3-MCTF. The solid line indicates the frames are passed on directly without motion compensation, and the dashed line means the frames are motion compensated. The frames enclosed by bold lines are wavelet transformed.

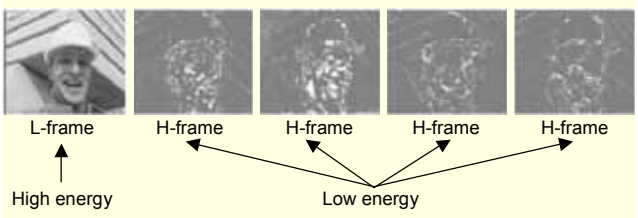


Fig. 2. Images of temporally filtered frames.

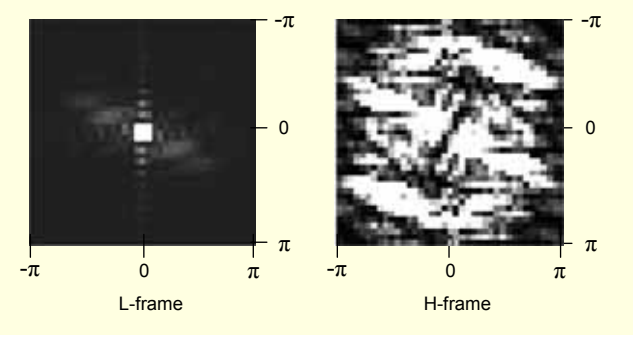


Fig. 3. Power spectrum density of L- and H-frames.

2. Image Signal Analysis of MCTF Frames

The signals in the L-frames are the average of image signals for consecutive frames. Therefore, the image signals in the L-frames are similar to those of the original image frames. Since the H-frames are obtained from the differences among consecutive frames, they contain the residual image signals that are similar to the edge signals or noise signals. Figure 2 shows the temporally filtered frames within a GOP. Figure 3 compares the power spectral densities of the L-frame and H-frame. It can been seen in the figure that the energy of an L-frame signal is concentrated in the low frequency range, whereas the energy of H-frame signals is spread out over the entire frequency range. This implies that the wavelet-based coding method exploiting energy compaction is efficient for the L-frames, but not for the H-frames. In the following sections, we analyze the coding performances of wavelet filters applied to various temporally filtered frames.

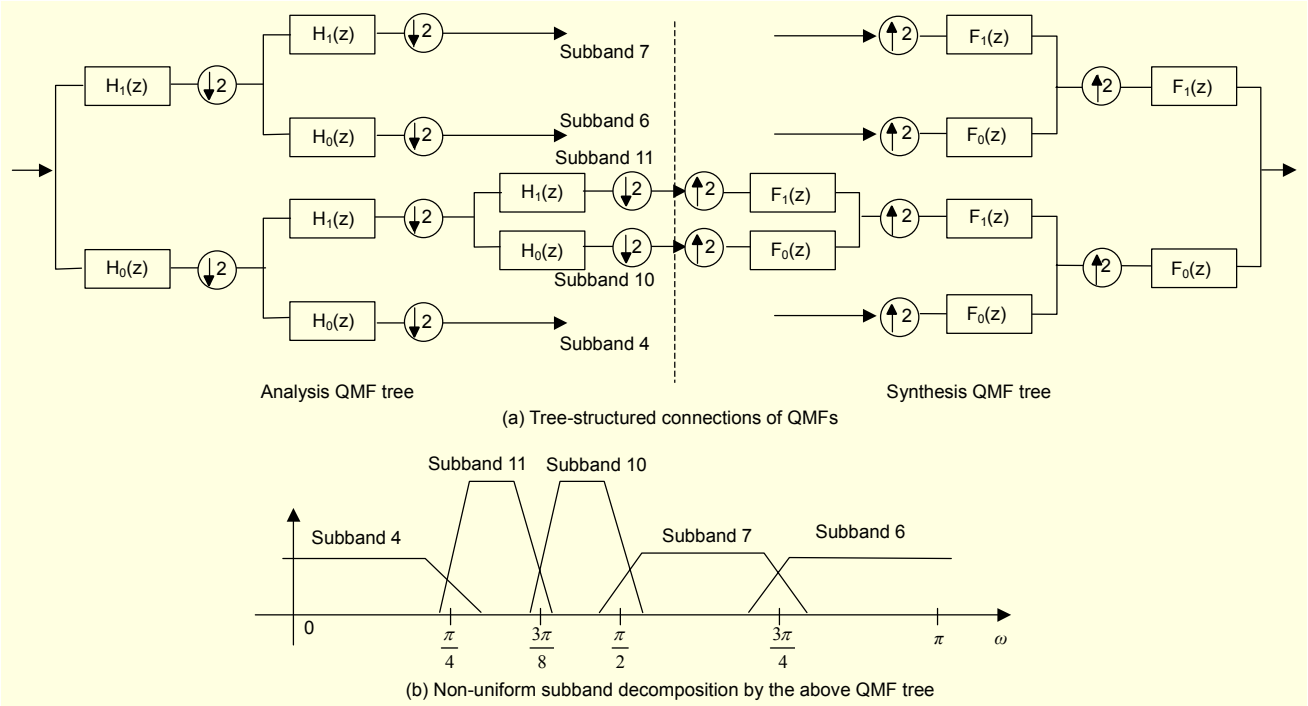


Fig. 4. An example of tree-structured quadrature mirror filter banks (QMFs) and their nonuniform subband decomposition. The leaf node set of this QMF tree is L(S) = {4, 6, 7 , 10, 11}. (2): downsampling of 2, (2): upsampling of 2.

III. Frame Analysis via Coding Gain

In this section, we analyze the coding performances of wavelet filters applied to various temporally filtered frames. Subband coding (SBC) methods such as those based on the wavelet transforms decomposes a signal frequency space into groups of subbands in order to increase its energy compactness [9], [10]. Due to this compact energy distribution, SBC is able to optimally allocate a given bit source to each subband.

Since the coding gain can act as a barometer of the energy compactness [6], we use it to evaluate the coding performances of the different wavelet filters. For this purpose, we build a mathematical model of the subband decompositions constructed by a wavelet transform and then drive the unified coding gain applicable to any structure of subband decompositions and any kind of FIR filters.

1. Mathematical Model of Subband Decomposition

Tree-structured quadrature mirror filter banks (QMF) realize non-uniform subband decompositions [9]. An example of a tree-structured QMF is shown in Fig. 4, where $H_0(z)$, $H_1(z)$ are the transfer functions of low and high analysis filters and $F_0(z)$, $F_1(z)$ are the low and high synthesis filters. If the non-uniform subband decomposition is a perfect reconstruction (PR) system, $H_0(z) = -F_1(-z)$, and $H_1(z) = F_0(-z)$. The QMF tree can be represented by a binary tree. For mathematical analysis, it is necessary to formulate the tree in association with the filter bank theory as follows:

- S: The set of nodes of a QMF tree. The root node is numbered as 1. Node $2i$ and node $(2i+1)$ are child nodes of node i. The branch from node i to node $2i$ performs lowpass filtering either by $H_0(z)$ followed by a down sampler in analysis filter banks, or by $F_0(z)$ following an upsampler in synthesis banks. The branch from node i to node $(2i+1)$ performs highpass filtering by either $H_1(z)$ followed by a downsampler in analysis filter banks, or by $F_1(z)$ following an upsampler in synthesis banks. A subband number is labeled as the corresponding node number.
- $L(S)$: The set of leaf nodes of tree S. Subbands corresponding to leaf nodes determine a subband decomposition. We use $|L(S)|$ to denote the number of channels (that is, the number of subbands).
- d_i: Depth of node i. That is, $d_i = \lfloor \log_2 i \rfloor$, where $\lfloor x \rfloor$ is the greatest integer such that $\lfloor x \rfloor \le x$. The subband size of node i is $\pi / 2^{d_i}$, and $\sum_{i \in L(S)} 1/2^{d_i} = 1$.
- $h^i(n)$, $H^i(z)$: The analysis filter and its response, which constructs subband i. This filter bank is equivalent to the d_i connections of QMFs along the path from the root node to node i.
- $f^i(n)$, $F^i(z)$: The synthesis filter and its response, which are counter parts of $h^i(n)$, $H^i(z)$.

$H^i(z)$, $F^i(z)$ are recursively calculated as [9]

$$H^i(z) = H_{(i \bmod 2)}(z^{d_i}) \cdot H^{\lfloor i/2 \rfloor}(z)$$
$$F^i(z) = F_{(i \bmod 2)}(z^{d_i}) \cdot F^{\lfloor i/2 \rfloor}(z) \quad ,$$

where $i \geq 2$, $H^1(z)$=1, and $F^1(z)$=1. Thus, the signals at each sides are obtained as follows.

At the analysis filter,

$$x_i(n) = \sum_{m=-\infty}^{\infty} x(m) h^i(2^{d_i} n - m) \, . \tag{1a}$$

At the synthesis filter,

$$\hat{x}_i(n) = \sum_{m=-\infty}^{\infty} \tilde{x}_i(m) f^i(n - 2^{d_i} m) \, , \tag{1b}$$

where $x(n)$ is an input signal at the analysis filter bank and $\tilde{x}_i(n)$ is the i-th subband input signal at the synthesis filter bank.

2. Unified Coding Gain

Without loss of generality, we can consider that a subband decomposition of the SBC is S, where the subband size of the subband channel i is $\pi / 2^{d_i}$, $i \in L(S)$. Figure 5 depicts the coding process at the subband channel i. The quantization noise $q_i(n)$ embedded into the subband i can be reasonably assumed to be an additive white process and uncorrelated with the subband signal, $x_i(n)$. From this uncorrelation, the channel reconstruction error is $e_i(n)$ produced by only $q_i(n)$ and is also uncorrelated with the channel reconstruction signal $\hat{x}_i(n)$ and other channel reconstruction error $e_j(n)$, $i \neq j$. Therefore, the reconstructed signal of the SBC is $\hat{x}(n) = \sum_{i \in L(S)} \{\hat{x}_i(n) + e_i(n)\} = x(n) + e(n)$, where $x(n)$ is the original signal and $e(n)$ is the reconstruction error of the SBC and $E\{|e(n)|^2\} = \sum_{i \in L(S)} E\{|e_i(n)|^2\}$.

Through the following analysis, the coding gain for a nonuniform subband coder is obtained in terms of the subband signal variances.

Theorem 1. When an input signal is wide-sense stationary (WSS), the unified coding gain, $G(S)$, for subband decomposition S is obtained as follows:

$$G(S) = \frac{\sigma_x^2}{\prod_{i \in L(S)} (\beta_i \sigma_{x_i}^2)^{1/2^{d_i}}} \, , \tag{2}$$

where $\beta_i = \sum_m |f^i(m)|^2, \sigma_x^2$ is the input signal variance, and $\sigma_{x_i}^2$ is the signal variance on subband i.

Proof. Let $\sigma_{q_i}^2$ be the quantization noise variance of the subband i whose size is $\frac{\pi}{M_i} = \frac{\pi}{2^{d_i}}$, and let f^i(m) be the synthesis filter to construct subband i, as shown in Fig. 5. The WSS signal after an interpolator becomes a cyclostationary process with the period of the interpolation rate [11]. Therefore, the channel reconstruction error $e_i(n)$ is a cyclostationary process with period M_i. From (1), the autocorrelation of $e_i(n)$ is obtained as follows:

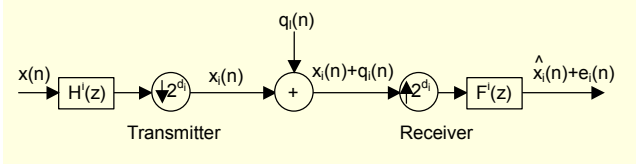


Fig. 5. One-channel system whose subband size is $\pi / 2^{d_i}$: $x(n)$ is the input signal, $q_i(n)$ is the channel quantization noise, $x_i(n)$ is the subband signal to be transmitted, $e_i(n)$ is the channel reconstruction error, and $\hat{x}_i(n)$ is the reconstruction signal at the channel.

$$\begin{aligned} R_e(n, n+m) &= E\{e_i(n) e_i(n+m)\} \\ &= \sum_r \sum_s E\{q_i(r) q_i(s)\} f^i(n - r M_i) f^i(m + n - s M_i) \\ &= \sum_r \sum_s R_{qq}(r-s) f^i(n - r M_i) f^i(m + n - s M_i). \end{aligned}$$

The average autocorrelation of a cyclostationary process is computed by taking the time average with respect to cyclo-period [11]. Thus, we have

$$\begin{aligned} \overline{R}_e(n) &= \frac{1}{M_i} \sum_{l=0}^{M_i - 1} R_e(l, n+l) \\ &= \frac{1}{M_i} \sum_{l=0}^{M_i - 1} \sum_r \sum_s f^i(l - r M_i) f^i(l + n - s M_i) \cdot R_{qq}(r-s) \end{aligned}$$

Because the quantization noises are white processes, $R_{qq}(r-s) = \sigma_{q_i}^2 \cdot \delta(r-s)$. Therefore,

$$\overline{R}_e(n) = \frac{\sigma_{q_i}^2}{M_i} \sum_{l=0}^{M_i - 1} \sum_r f^i(l - r M_i) f^i(l + n - r M_i).$$

From $E\{|e_i(n)|^2\} = \overline{R}_e(0)$, $E\{|e_i(n)|^2\}$ is calculated such that

$$E\{|e_i(n)|^2\} = \frac{\sigma_{q_i}^2}{M_i} \sum_{l=0}^{M_i - 1} \sum_r |f^i(l - r M_i)|^2 = \frac{\sigma_{q_i}^2}{M_i} \sum_m |f^i(m)|^2 \, .$$

Since $\sum_{i \in L(S)} 1/2^{d_i} = 1$, the arithmetic and geometric mean inequality leads to a reconstruction error variance bound such as

$$\begin{aligned} E\{|e(n)|^2\} &= \sum_{i \in L(S)} E\{|e_i(n)|^2\} \\ &\geq \prod_{i \in L(S)} \left(\sigma_{q_i}^2 \sum_m |f^i(m)|^2 \right)^{1/d_i} \end{aligned} \tag{3}$$

Let R_i be the number of bits allocated to the quantizer encoding the signal at subband i. The coding bit rate of this subband is $R_i / 2^{d_i}$ bits per second because the subband signals are sampled at a rate of $1/2^{d_i}$ samples per second. Thus, the total coding bit rate is $R = \sum_{i \in L(S)} R_i / 2^{d_i}$. According to the Shannon rate-distortion bound, the quantization error variance of quantizer q_i must satisfy $\sigma_{q_i}^2 > 2^{-2R_i} \sigma_{x_i}^2$, where $\sigma_{x_i}^2$ is the signal variance of subband i [9]. Thus, (3) becomes

$$E\{|e(n)|^2\} > \prod_{i \in L(S)} \left(2^{-2R_i} \sigma_{x_i}^2 \sum_n |f^i(n)|^2 \right)^{1/2^{d_i}} = 2^{-2R} \prod_{i \in L(S)} \left(|f^i(m)|^2 \sigma_{x_i}^2 \right)^{1/2^{d_i}} \quad (4)$$

The reconstruction error lower bound in (4) indicates the degree of signal energy compaction, which is equal to the denominator of the coding gain. And the direct quantization of the input signal through pulse code modulation generates a quantization error variance larger than $2^{-2R_i} \sigma_x^2$, where σ_x^2 is the signal variance of the input signal. Therefore, we obtain (2) as the coding gain for subband decomposition S. □

The normalized filter energy $\sum_m |f^i(m)|^2$ represents the factor of the distribution of energy by the subband filter and may be regarded as a weight factor for the subband. This weight factor takes into account the quantization noise leakage to other channels. In practically used biorthogonal filters, energy leakage is not desirable and not greatly allowed. The normalized filter energy for practical biorthogonal filters should be $\sum_m |f^i(m)|^2 \approx 1$. As an biorthogonal example, Table 1 lists the filter coefficients of the 9/7 filter that is popularly used for image coding [12]. And the normalized filter energies at each subband i are calculated in Table 2.

Table 1. Coefficients of the biorthogonal 9/7 wavelet filter.

n	0	±1	±2	±3	±4
$2^{-1/2}h_0(n)$	0.602949	0.266864	-0.078223	-0.016864	0.026749
$2^{-1/2}f_0(n)$	0.557543	0.295636	-0.028772	-0.045636	

3. Analysis of Coding Performance

In this section, we evaluate and analyze the coding performances of various wavelet filters in terms of coding gain. We extract an L-frame at the 4th level MCTF and H-frames at the 1st level MCTF, and then apply the 9/7 filter to the L-frame and Haar filter to the H-frames. The level of wavelet decomposition is 3 and thus the number of subbands is 10.

In order to compare the coding gains achieved by the 9/7 and Haar filters, we define the ratio of those coding gains such that

$$G_{comp} = \frac{Gain_{9/7}}{Gain_{Haar}} = \frac{\prod_{i \in L(S)} (\beta_i^{Haar} \cdot \sigma_{x_i^{Haar}}^2)^{1/2^{d_i}}}{\prod_{i \in L(S)} (\beta_i^{9/7} \cdot \sigma_{x_i^{9/7}}^2)^{1/2^{d_i}}},$$

where $\sigma_{x_i^{Haar}}^2$ and $\sigma_{x_i^{9/7}}^2$ are the signal variances of the subband i transformed by the Haar and 9/7 filters, respectively, $L(S)=\{5,6,7,9,10,11,16,17,18,19\}$, $d_i=\{2,2,2,3,3,3,4,4,4,4\}$, and $\beta_i^{Haar} = 1$ for all i, and the values of $\beta_i^{9/7}$ are listed in Table 2.

When G_{comp} >1, the coding gain of the 9/7 filter is greater than that of the Haar filter, and when G_{comp} <1, the coding gain of the Haar filter is greater than that of the 9/7 filter. Table 3 shows the values of G_{comp} for the image sequences used in the core experiment of the MPEG committee [13]. The values in Table 3 are measured at each frame on the first temporal decomposition level. When G_{comp} is close to 1, the coding performance differences between the 9/7 filter and Haar filter are very trivial. We learned that it is better to use the Haar filter when $|G_{comp}| < 1.1$ since the filter reduces complexity without degrading the coding efficiency.

Most of the energy of L-frames is concentrated in the low

Table 2. Subband weight factors $\sum_m |f^i(m)|^2$.

Subband i	2	3	4	5	6	7	8	9	10	11	12	13	14	15	…
β_i	0.98	1.04	0.97	1.02	1.02	1.08	0.95	1.07	0.98	0.97	1.04	1.08	0.99	0.96	…

Table 3. The values of G_{comp} for various image sequences.

	Mobile	City	Foreman	Soccer	Crew	Football	Harbor	Bus
L-frame	1.5680	1.11956	1.8977	1.2928	1.3232	2.2070	1.6640	2.0175
H-frame	1.0651	0.98400	1.0814	1.0441	1.2260	1.4652	1.1303	1.3039

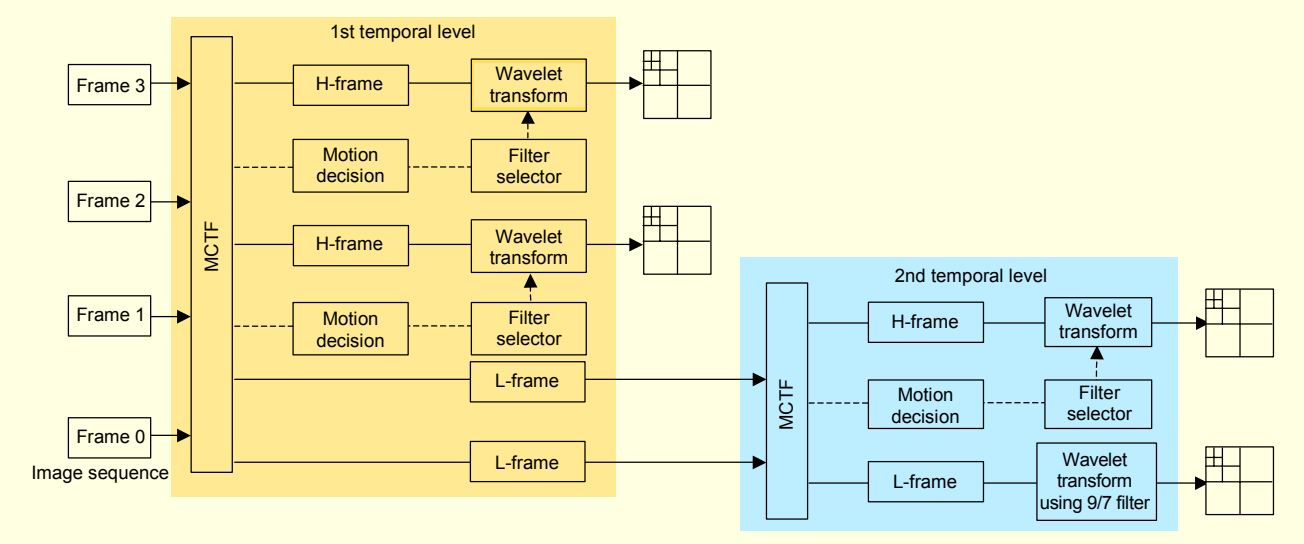


Fig. 6. Proposed frame-adaptive spatial wavelet transform.

frequency range. Therefore, the subband decomposition capability of wavelet filters applied to L-frames is essential to achieve efficient energy compactness [6]. In this respect, since it provides superior energy compactness, the performance of the 9/7 filter is superior to that of the Haar filter for L-frames [10].

In the case of H-frames obtained from slowly varying images sequences such as 'mobile' and 'city', the motions are well predicted, and the H-frames of these sequences mostly contain inter-blocks with residual image signals as shown in Fig. 2. The spectrums of these H-frames are also widely spread as seen in Fig. 3. Consequently, they cannot be efficiently compacted. Therefore, in this case the subband decomposition capability of the wavelet filters does not have a significant effect on their coding performances. Thus, for such H-frames, the performances of the 9/7 filter is almost equal to that of the Haar filter. On the contrary, in image sequences with fast motions and large movements such as 'football' and 'bus', it is difficult to find matched blocks among consecutive frames. Therefore, H-frames obtained from fast-changing image sequences should contain many intra-blocks whose image signals are original image signals and would be expected to have spectrums that are not widely spread. Thus, the energy compactness of these H-frames is effective, and therefore the 9/7 filter shows better performances than the Haar filter in this case.

IV. Frame-Adaptive Spatial Wavelet Transform

As mentioned in section III.3, a long filter should be used when coding L-frames or H-frames obtained from fast-changing sequences, whereas a short filter can be used when coding H-frames from rarely changing image sequences without deteriorating the coding performance. Based on this observation, we propose a scheme in which customized wavelet filters are applied to the temporally filtered frames in accordance with their characteristics. In other words, a long filter is applied to the L-frames or H-frames obtained from the fast changing sequences, while a short filter is used for the H-frames obtained from rarely changing image sequences.

Figure 6 shows an example of the proposed scheme with a 2-level MCTF. The proposed scheme determines the characteristics of the H-frames through motion prediction rather than by measuring the coding gain, which is impractical in a real system. Motion prediction performed during MCTF detects unconnected pixels that do not match any pixels in the adjacent frames [3], [4]. A large number of unconnected pixels indicate that the image sequence being processed includes fast changes, and the H-frames obtained from this sequence are likely to contain many pixels whose signals are close to the original image signals. Therefore, the proposed scheme applies a long filter to an L-frame and H-frames containing many unconnected pixels. Conversely, since H-frames containing a small number of unconnected pixels are bound to be obtained from rarely changing image sequences, a short filter is sufficient for their coding. In the proposed scheme, a wavelet filter is adaptively selected during the filter selection phase according to the characteristics of the H-frames, which are determined during the motion prediction phase.

Considering that most consecutive frames do not change abruptly, a short filter is used for coding most of the H-frames. Therefore, the proposed method greatly reduces the computational complexity of the spatial wavelet transform without deteriorating the coding efficiencies. The proposed method also alleviates the ringing effect of slowly changing image sequences, since the shorter filter that is applied to the

Table 4. Spatial wavelet transform complexity by MC-EZBC for 'city' sequence.

		Conventional method		Proposed method		CI (%)
Size	Frame rate (fps)	Num. of multiplications	Num. of 9/7 filter uses	Num. of multiplications	Num. of 9/7 filter uses	
4CIF	30	290,820,000	300	959,713,920	32	303
CIF	30	727,056,000	300	239,928,480	32	303

Table 5. Spatial wavelet transform complexity by MC-EZBC for 'crew' sequence.

		Conventional method		Proposed method		CI (%)
Size	Frame rate (fps)	Num. of multiplications	Num. of 9/7 filter uses	Num. of multiplications	Num. of 9/7 filter uses	
4CIF	30	2,326,579,200	240	690,703,200	15	336
CIF	30	581,644,800	240	172,675,800	15	336

H-frames containing edge-like or residual image signals propagate less coding errors than would a longer filter.

$$CI = \frac{C_{conventional}}{C_{proposed}} \times 100\ (\%).$$

V. Experiment and Discussion

In this section, we evaluate the coding performances and complexities of the proposed scheme. For YUV (4:2:0) format sequences, the numbers of multiplications required by the proposed spatial wavelet transform scheme and the conventional scheme are calculated as follows [10]:

$$C_{conventional} = GOP_{size} \cdot \sum_{l=0}^{L} \frac{M \cdot N}{4^l}(l^{long} + h^{long}),$$

$$C_{proposed} = (GOP_{size} - H_{num}) \cdot \sum_{l=0}^{L} \frac{M \cdot N}{4^l}(l^{long} + h^{long}) + H \cdot \sum_{l=0}^{L} \frac{M \cdot N}{4^l}(l^{short} + h^{short})$$

where GOP_{size} is the number of frames in a GOP, L is the decomposition level in the spatial wavelet transform, $M \cdot N$ are the horizontal and vertical sizes of the frame, H_{num} is the number of H-frames from slowly varying image sequences, H_{num} is counted during MCTF, l^{long}, h^{long} are the lengths of the low-pass and high-pass wavelet filters used for L-frames and H-frames from fast varying image sequences, and l^{short}, h^{short} are the lengths of the low-pass and high-pass wavelet filters used for H-frames from slowly varying image sequences.

For the experiments, the 9/7 and 2/2 (Haar) filters are used as the long and short filters, respectively. We determine how much the proposed scheme reduces the multiplication complexity in comparison with the conventional scheme. For this purpose, we define the complexity improvement (CI) as follows:

For the experiments, we use MSRA developed by Microsoft Research-Asia and MC-EZBC developed by Choi and Woods [3], [8], and [14]. For all of the experiments, we used full search motion estimation with quarter-pixel resolution, and we set the GOP size to 16 frames. The spatial wavelet transform decomposition level is 4. The tests were conducted on various frame sizes, as well as different frame and bit rates.

Tables 4 and 5 show the complexity improvements, while Tables 6 and 7 show comparisons of the coding performances between the conventional and proposed methods when the latter is applied to MC-EZBC. We apply the 9/7 filter to the L-frames at the last temporal filtering level. In the case of MC-EZBC using the Haar MCTF, whenever a scene change occurs the L-frames to be temporally filtered are passed down directly without MCTF, so as to become an L-frame and an H-frame at the next temporal filtering level. Thus, the H-frames corresponding to points in the image sequence where scene changes occur should contain the natural image signals. Consequently, we also apply the 9/7 filter to these H-frames.

We apply the Haar filter to H-frames obtained from image sequences where scene changes occur. When the portion of unconnected pixels between consecutive frames exceeds 60% of the frame size, we consider that a scene change occurs. Figure 7 depicts the procedure of MCTF by MC-EZBC.

We also apply the proposed scheme to MSRA since it performs bi-directional motion compensation using the 5/3 MCTF based on overlapped block motion compensation (OBMC). Due to the OBMC-based MCTF, MSRA always treats adjacent frames as inter-frames and, consequently, no

Table 6. PSNR comparison by MC-EZBC for 'mobile' sequence.

			Conventional method PSNR (dB)			Proposed method PSNR (dB)		
Size	Frame rate (fps)	Bit rate (bps)	Y-frame	U-frame	V-frame	Y-frame	U-frame	V-frame
CIF	30	1,024	28.97	33.47	32.97	28.84	33.46	32.96
CIF	15	512	20.52	32.14	31.16	20.52	32.34	31.32
QCIF	15	256	20.01	29.27	28.33	20.00	29.53	28.57
QCIF	15	128	20.34	27.65	26.33	20.31	27.78	26.46

Table 7. PSNR comparison by MC-EZBC for 'city' sequence.

			Conventional method PSNR (dB)			Proposed method PSNR (dB)		
Size	Frame rate (fps)	Bit rate (bps)	Y-frame	U-frame	V-frame	Y-frame	U-frame	V-frame
4CIF	30	3,000	36.18	44.35	46.16	36.14	43.61	46.50
4CIF	30	1,500	33.63	42.78	44.68	33.62	43.03	44.99
CIF	30	750	31.58	42.75	44.35	31.51	43.07	44.54
CIF	30	384	29.81	40.88	42.30	29.81	40.85	42.25

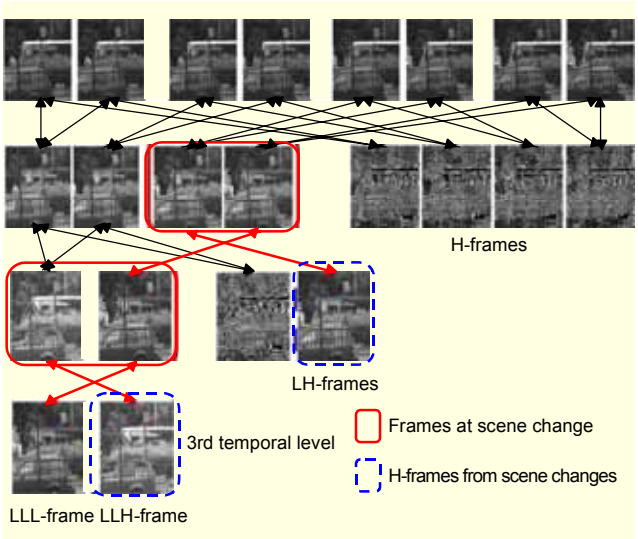


Fig. 7. Example of MCTF procedure by MC-EZBC.

scene changes occur. Thus, in this case we apply the 9/7 filter to the L-frames and the Haar filter to all of the H-frames. Table 8 shows the complexity improvements. Since all of the H-frames are transformed by the Haar filter, the CI is the same for all image sequences. Tables 9 and 10 show a comparison of the coding performances between the conventional and proposed methods.

As can be seen in the experimental results, the coding performances of the conventional and proposed methods differ by less than 0.03 dB in average peak signal-to-noise ratio. These differences in the coding efficiency are not distinguishable in terms of the subjective qualities of the resulting images. The relative performances of the two methods depend on the nature of video data being processed. Therefore, it is not possible to decide the outperforming method in terms of coding efficiency. However, the proposed method reduces the number of multiplications required in the spatial wavelet transform by more than 300%.

The ringing effect caused by the propagation of coding errors is also apparent at object boundaries or edges, which are most probably contained at H-frames obtained from slowly changing image sequences [10]. Since the shorter filter applied to the H-frames containing edge-like image signals propagates less coding error than a longer filter would, the proposed scheme applying the short filter to H-frames enables the ringing effect to be reduced for slowly changing images sequences. Figure 8 shows an example of the ringing effect reduced by the proposed method. It should be noted that the ringing effect on moving sequence is more apparent, while the effect at a still image is not clearly apparent.

VI. Conclusions

In this paper, we presented an interframe wavelet coding scheme, which adaptively chooses the spatial wavelet transform filters to apply to each temporally filtered frame depending on its characteristics. We analyzed the characteristics of the temporally filtered frames in terms of their coding gain. In order to render the proposed scheme more practical, we also proposed an efficient method of determining

Table 8. Spatial wavelet transform complexity by MSRA.

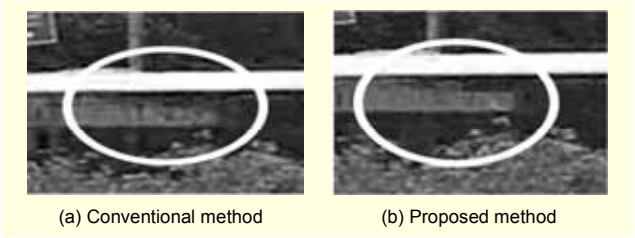
		Conventional method		Proposed method		CI (%)
Size	Frame rate (fps)	Num. of multiplications	Num. of 9/7 filter uses	Num. of multiplications	Num. of 9/7 filter uses	
4CIF	30	2,585,100,000	240	767,448,000	15	336
CIF	30	646,272,000	240	191,862,000	15	336

Table 9. PSNR comparison by MSRA for 'city' sequence.

			Conventional method PSNR (dB)			Proposed method PSNR (dB)		
Size	Frame rate (fps)	Bit rate (bps)	Y-frame	U-frame	V-frame	Y-frame	U-frame	V-frame
4CIF	30	1,024	29.66	39.04	40.08	29.48	39.06	40.11
CIF	30	512	31.78	39.33	40.84	31.73	39.09	40.80
CIF	30	256	30.32	41.99	43.99	30.10	41.89	43.66
QCIF	15	64	35.22	43.01	44.78	35.12	43.01	44.79

Table 10. PSNR comparison by MSRA for 'mobile' sequence.

			Conventional method PSNR (dB)			Proposed method PSNR (dB)		
Size	Frame rate (fps)	Bit rate (bps)	Y-frame	U-frame	V-frame	Y-frame	U-frame	V-frame
CIF	30	1024	31.45	36.91	36.21	31.59	37.10	36.30
CIF	15	512	27.53	32.09	31.46	27.43	32.09	31.46
CIF	15	384	25.01	29.33	28.73	25.06	29.34	28.73
QCIF	15	192	25.57	29.76	29.36	25.57	29.77	29.36



(a) Conventional method (b) Proposed method

Fig. 8. The proposed method produces less ringing distortion than the conventional method. White circles point to the ringing effects for each method. The test sequence is the 'bus' sequence whose size and frame rate are CIF and 30 fps respectively. The coding bit rate is 512 kbps. It should be noted that the ringing effect on a moving sequence is more apparent than that on still images.

the characteristics of the temporally filtered frames by counting the number of unconnected pixels. The experimental results confirmed that the coding performances afforded by the proposed scheme are equivalent to those of the conventional method, while the proposed method reduces the spatial transform complexity by about 300% when the GOP size is 16. The proposed scheme also reduces the ringing effect for slowly changing image sequences.

References

[1] Gwang Hoon Park and Kyuheon Kim, "Adaptive Scanning Methods for Fine Granularity Scalable Video Coding," *ETRI Journal*, vol. 26, no. 3, Aug. 2004, pp. 332-343.

[2] J.R. Ohm, "Complexity and Delay Analysis of MCTF Interframe Wavelet Structures," ISO/IEC JTC1/SC29/WG11 M8520, Klagenfurt, July 2002.

[3] S.-J Choi and J.W. Woods, "Motion Compensated 3-D Subband Coding of Video," *IEEE Trans. Image Proc.*, vol. 8, no. 2, Feb. 1999, pp. 155-167.

[4] S.-Ta Hsiang and J.W. Woods, "Embedded Video Coding Using Invertible Motion Compensated 3-D Subband/Wavelet Filter Bank," *Signal Processing: Image Communication*, vol. 16, May 2001, pp. 705-724.

[5] Bong-Keun Choi and Won-Ha Kim, "Dynamic Algorithm for Constructing the Optimal Subband Decomposition," *IEICE Trans. Information and Systems*, vol. E86-D, no. 3, Mar. 2003, pp. 633-640.

[6] A.K. Soman and P.P. Vaidyanathan, "Coding Gain in Paraunitary Analysis/Synthesis System," *IEEE Trans. Signal Processing*, vol. 41, no. 5, May 1993, pp. 1824-1835.

[7] J.R. Ohm, "Three-Dimensional Subband Coding with Motion Compensation," *IEEE Trans. Image Proc.*, vol. 3, no. 5, Sept. 1994, pp. 559-571.

[8] Jizheng Xu, Ruiqin Xiong, Bo Feng, Gary Sullivan, Ming-Chieh Lee, Feng Wu, and Shipeng Li, "3D Sub-band Video Coding Using Barbell Lifting," ISO/IEC JCT1/SC29/WG11 M10569/S05, Mar. 2004.

[9] P.P. Vaidyanathan, *Multirate Systems and Filter Banks*, Prentice-Hall, 1993.

[10] G. Strang and Truong Nguyen, *Wavelets and Filter Banks*, Wellesy-Cambridge Press, 1997.

[11] V.P. Sathe and P.P. Vaidyanathan, "Effects of Multirate Systems on Statistical Properties of Random Signal," *IEEE Trans. Acoust., Speech Signal Process.*, vol. 41, no. 1, Jan. 1989, pp. 131-146.

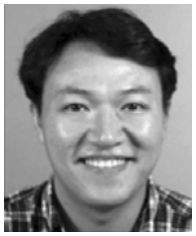
[12] M. Antonimi, M. Barlaud, P. Mathiu, and I. Daubeechies, "Image Coding Using Wavelet Transform," *IEEE Trans. Image Process.*, vol. 1, no. 2, Apr. 1992, pp. 205-220.

[13] M. van der Schaar and J. Ridge, "Description of Core Experiments in SVC," ISO/IEC JTC 1/SC 29/WG 11/N6521, July 2004.

[14] Software package "MPEG-CVS," http://mpeg.nist.gov

Wonha Kim received the BS degree in electronics engineering in 1995 from Yonsei University, Seoul, Korea, and the MS and PhD degrees in electrical engineering from the University of Wisconsin-Madison, USA, in 1988 and 1997. He worked with Motorola at Schamburg, Illinois, USA, from Jan. 1996 to Dec. 1996. He was with Los Alamos National Laboratory at Los Alamos, New Mexico, USA from August 1997 to February 2000 as a Post Doctor. From March 2000 to August 2003, he was with Myongji University, Yongin, Korea, where he was an Assistant Professor of the Department of Electronics, Information and Communication Engineering. Since September 2003, he has been with the College of Electronics and Information Engineering, Kyunghee University, Suwon, Korea, as an Associate Professor. His research interests involve multimedia signal processing and communications.

Seyoon Jeong received the BS degree in electronics engineering in 1995 from Inha University, Incheon, Korea, and the MS degree in electronic engineering from Inha University in 1997. Since 1996, he has been a member of research staff in Electronics and Telecommunications Research Institute, Korea. His current research activities are in the development of digital mobile broadcasting. He is interested in video coding, digital mobile broadcasting, and interactive multimedia broadcasting systems.



Kyuheon Kim received the BS in electronics engineering in 1989 from Hanyang University, Seoul, Korea, and the M.Phil and PhD degrees in electrical and electronic engineering from University of Newcastle upon Tyne UK, in 1992 and 1996. From 1996 to 1997, he worked as a Research Fellow for University of Sheffield, UK. From 1997, he was the Leader of the Interactive Media Research Team, Broadcasting Media Research Department in ETRI, and is currently working for Kyunghee University as an Associate Professor. He is interested in digital signal and video processing, and interactive multimedia broadcasting systems.