Fully Scalable 3-D Overcomplete Wavelet Video Coding using Adaptive Motion Compensated Temporal Filtering

Jong Chul Ye and Mihaela van der Schaar
Philips Research USA

ABSTRACT

In this paper, we present a fully scalable 3-D overcomplete wavelet video coder that employs a new and highly efficient 3-D lifting structure for adaptive motion compensated temporal filtering (MCTF). Unlike the conventional interframe wavelet video techniques that apply MCTF on the spatial domain video data and then encode the resulting temporally filtered frames using critical sampled wavelet transforms, the scheme proposed in this paper performs first the spatial domain wavelet transform and subsequently applies MCTF for each wavelet band. To overcome the inefficiency of motion estimation in the wavelet domain, the low band shifting method (LBS) is used at both the encoder and decoder to generate an overcomplete representation of the temporal reference frames. A novel interleaving algorithm for the overcomplete wavelet coefficient is proposed that enables *optimal* sub-pixel accuracy motion estimation implementations. Furthermore, to achieve arbitrary accuracy motion estimation and compensation in the overcomplete wavelet domain with *perfect reconstruction*, a novel 3-D lifting structure is also introduced. Simulation results shows that the proposed fully scalable 3-D overcomplete wavelet video coder has comparable or better performance (up to 0.5dB) than the previously proposed interframe wavelet coders under the same coding conditions. Several techniques that can further improve the performance of the proposed overcomplete wavelet coding scheme are also discussed.

1. INTRODUCTION

A majority of the current video coding algorithms is based on motion compensated predictive coding. In such hybrid schemes, temporal correlations are removed using motion compensation and the resulting motion compensated residuals are spatially decorrelated using transform (DCT, wavelets etc.) coding techniques. Scalable compression schemes based on predictive coding structures incur drift if decoding is performed at various bit-rates or spatial resolutions and hence, they do not provide efficient spatio-temporal-SNR scalability implementations. Alternatively, 3D interframe wavelet video coding schemes [1]-[9] can provide flexible spatial, temporal, SNR and complexity scalability with fine granularity over a large range of bit rates. Early contributions to the field of wavelet and multi-resolution video coding were provided, among others, by Gharavi[1], Zhang and Zafar [2], Taubman and Zakhor [3]. Motion compensated temporal filtering (MCTF) was first proposed by Ohm [4] and later improved by Choi and Woods[6]. Unlike predictive coding, where decoded frames are used as references for the motion compensation of future frames, MCTF does not employ a temporal recursive structure. Instead, in MCTF encoding, the original frames are filtered temporally in the direction of motion and the temporally decorrelated signal is coded using 2D spatial wavelet transforms and embedded coding. In this conventional MCTF framework, as shown in Figure 1, successive pairs of frames are temporally filtered using a two-channel Haar filter-bank to create low-pass (L) and high-pass (H) frames, thereby removing the short-term dependencies between successive frames. The long-term temporal dependencies are removed by further decomposing the L-frames using a pyramidal or multi-resolution decomposition structure. In conventional MCTF, the same Haar filter-bank is used at all various temporal decomposition levels. MCTF-based wavelet coding provides many advantages over conventional motion compensation algorithms for predictive coding, such as providing flexible spatio-temporal-SNR and complexity scalabilities, as well as improved error resilience due to clear prioritization of the coded video coefficients. Hence, interframe wavelet video coding schemes are especially attractive for video transmission over heterogeneous wireless and wired networks to various devices with different capabilities. As shown in Figure 1, MCTF is performed in the spatial domain (SD), before the spatial wavelet decomposition. Throughout the paper we refer to these schemes as *spatial domain MCTF* (SDMCTF).

However, these interframe wavelet video schemes have several important limitations, as described below.

- *Limited motion-estimation efficiency.* SDMCTF are inherently limited by the quality of the matches provided by the motion estimation algorithm. Also, the discontinuities at the motion boundaries (blocking

Visual Communications and Image Processing 2003, Touradj Ebrahimi, Thomas Sikora,
Editors, Proceedings of SPIE Vol. 5150 (2003) © 2003 SPIE · 0277-786X/03/$15.00

artefacts) are represented as high-frequency content in the high-frequency wavelet subbands, thereby resulting in a reduced coding efficiency. Moreover, during the temporal filtering process, some pixels are either not referenced or are referenced multiple times due to the nature of the motion in the scene and the covering/uncovering of objects. Such pixels are termed "unconnected" pixels and they require special handling such as "Intra" block coding. However, the "Intra/Inter" mode switch for motion estimation is not very efficient in SDMCTF schemes, since the wavelet transform is applied globally and cannot encode efficiently these discontinuities.

- *Quality of spatial scalability.* Spatial scalability is not very efficient for the lower resolutions, since the motion estimation accuracy is fixed for all the spatial resolution (e.g. HD to QCIF) resulting in inefficient implementations. More specifically, the motion reference during MCTF is full reference image (say CIF), which may not be available for decoding the low resolution decoding (QCIF). Hence, this incurs the mismatch of the motion reference, which results in "drift" in decoded low resolution video.

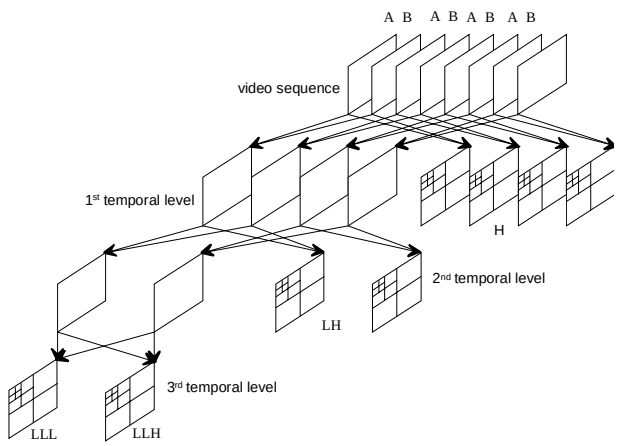

Figure 1: Interframe wavelet coder.

To cope with these inefficiencies, in [8][9] we proposed an alternative 3D wavelet video coding scheme, denominated inband motion compensated temporal filtering (IBMCTF). In IBMCTF, each video frame is first spatially decomposed into multiple bands using the wavelet filtering, and subsequently the temporal correlation for each band is removed using MCTF (see Figure 2). Note that since the critical-sampled wavelet decomposition is only periodically shift-invariant, the motion estimation and compensation in wavelet domain is not efficient and the coding penalty are observed [10]. Hence, a new procedures need to compensate the lost coding efficiency. In [10], a novel scheme called low band shifting method (LBS) is presented that can efficiently generate the overcomplete representation of the wavelet coefficients, which is now shift invariant. Additional advantage of LBS is that the overcomplete expansion is generated at the encoder and decoder using a similar procedure, and hence no additional information needs to be encoded and transmitted as compared to conventional interframe wavelet coding schemes. Consequently, IBMCTF coding schemes [8] employ the LBS algorithm for the motion compensated temporal filtering in the wavelet domain. The residual signal after IBMCTF can be coded band-by-band using any optimized texture coding technique.

Note that this structure enables improved spatial scalability performance as compared with SDMCTF [8]. This is because the temporal filtering is performed per sub-band (resolution) and hence, loss of information from the finer resolution bands does not incur any drift in the temporal direction.

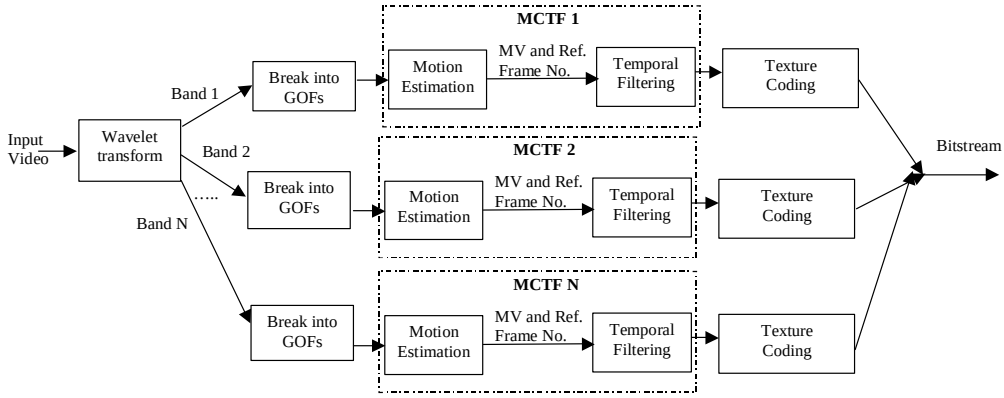

Figure 2: Block diagram of IBMCTF coder.

One of the drawbacks of [8] is the increase of motion vectors bits compared to SDMCTF. Several alternative schemes for efficient motion vector coding are provided in [11], but the motion vector coding overhead is still not negligible especially at the finest spatial resolutions. Furthermore, in [8] and [11], the phase component of the overcomplete wavelet expansion is encoded separately, thereby adding additional overhead. Another limitation of the method proposed in [8] is that the motion estimation is limited to pixel-accuracy implementations.

In this paper, we present a novel fully scalable 3-D overcomplete wavelet video coder that overcomes the previous drawbacks of the IBMCTF framework [8], and demonstrates coding efficiency comparable or better than conventional interframe wavelet coding schemes based on SDMCTF [7]. The proposed 3-D overcomplete wavelet video coder is based on a 3-D lifting structure that enables arbitrary accuracy motion estimation and compensation for the various sub-bands. The proposed IBMCTF framework does not incur any motion vector overhead compared to SDMCTF. We also introduce a novel interleaving scheme which combines the different phase information of the overcomplete wavelet coefficients, such that there is no need to encode the phase information separately as in [8][11]. In our proposed IBMCTF scheme, due to the interleaving, the phase information is coded inherently as part of the higher accuracy motion vectors. Moreover, the implementation difference between SDMCTF and IBMCTF coders is minimal, such that the adaptive motion estimation techniques used for SDMCTF - hierarchical variable size block matching (HVSBM)[6], backward motion compensation, and adaptive insertion of intra blocks - can easily be incorporated in IBMCTF.
The paper is organized as follows. Section 2 briefly reviews the LBS method for generating the overcomplete wavelet expansion and proposes a novel interleaving scheme for the overcomplete wavelet coefficients that is essential for enabling sub-pixel accurate motion estimation. Section 3 introduces the proposed 3-D lifting structure in the overcomplete domain, followed by a description of our IBMCTF-based wavelet video coder in Section 4. Section 5 presents the simulation results. The conclusions are presented in Section 6, together with a discussion of further research.

2. MOTION ESTIMATION AND COMPENSATION IN THE OVERCOMPLETE WAVELET DOMAIN

In this section, we briefly review the low band shifting method and wavelet block concept proposed in [10], and subsequently we introduce our interleaving scheme for the overcomplete wavelet coefficients.

Low Band Shifting Method (LBS)

Note that due to the decimation procedure in wavelet transform, the wavelet coefficient is no more shift-invariant. Hence, translation motion in spatial domain cannot be accurately estimated from the wavelet coefficients. This incurs significant coding efficiency loss. Recently, the LBS method [10] has been proposed to overcome the shift-variant property of wavelet transform. At the first level, the original and shifted signals are decomposed into low-band and high-band signals. Subsequently, the low-band signal is further decomposed in the same way as for the first level. The 1-D formulation can be easily expanded to wavelet decompositions having multiple levels and also

to 2-D image signals as shown in Figure 3. In Figure 3, the pair (m,n) indicates that the wavelet coefficients within that band were generated by shifts of m-pixels in the x-direction and n-pixels in the y-direction, respectively. As shown in Figure 3, LBS generates the full-set of wavelet coefficients for all the possible shifts of the input band. Hence, the representation now accurately conveys any shift in spatial domain. In the sequel, we refer to the different shifted wavelet coefficients corresponding to the same decomposition level at a specific spatial location, as "cross-phase" wavelet coefficients. For further illustration, a "real" video example of a 2-level overcomplete wavelet transform obtained using the LBS method is shown in Figure 4. Note that for an n-level decomposition, the overcomplete wavelet representation requires a storage space that is 3n+1 larger than that of the original image.

Interleaving of Wavelet Coefficients

In this paper, we introduce a novel interleaving technique for storing the overcomplete wavelet coefficients that differs from that depicted in Figure 3 and Figure 4. Figure 5 portrays the proposed interleaving method for the 1-D case of a one level decomposition. The interleaving is performed such that the new coordinates in the overcomplete domain correspond to the associated shift in the original spatial domain.

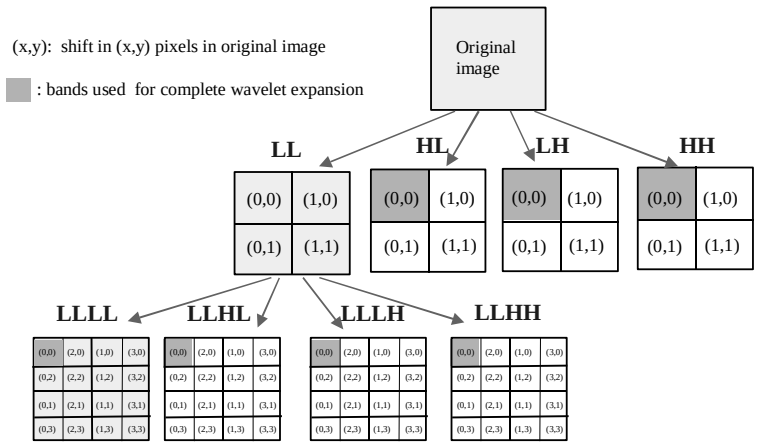

Figure 3: Over-complete wavelet expansion using LBS algorithm for two level decomposition.

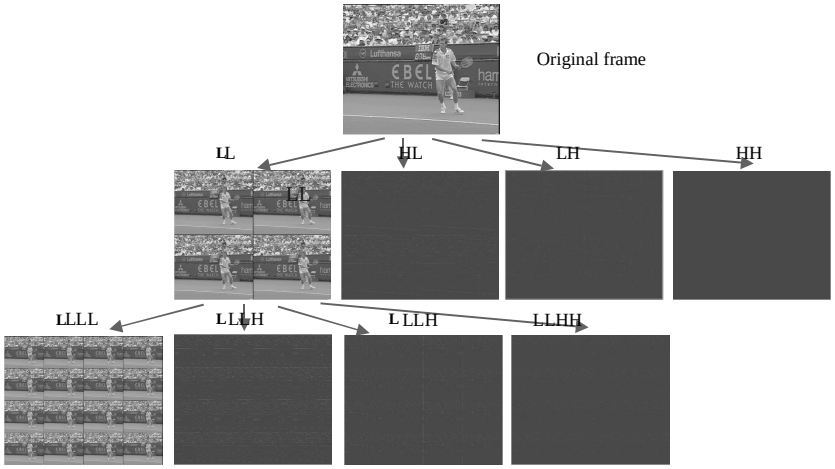

Figure 4: Wavelet coefficients for the first frame of the "Stefan" sequence using LBS algorithm.

The interleaving structure can be used recursively at each decomposition level and can be directly extended for 2-D signals. The overcomplete wavelet coefficients after the proposed interleaving step are depicted in Figure 6. As observed from Figure 6, the interleaved low band signal is a low-pass filtered version of the original frame using the overcomplete wavelet low-pass filter. The interleaving algorithm enables the *optimal method* for providing sub-pixel accuracy motion estimation and compensation in IBMCTF. Without the interleaving procedure, previously

proposed IBMCTF schemes [8] cannot provide the optimal sub-pixel accuracy motion estimation and compensation, because they do not take into consideration cross-phase dependencies between neighbouring wavelet coefficients. Furthermore, the proposed interleaving method allows the usage of the same adaptive motion estimation techniques in IBMCTF as those used for SDMCTF - hierarchical variable size block matching, backward motion compensation, and adaptive insertion of intra blocks.

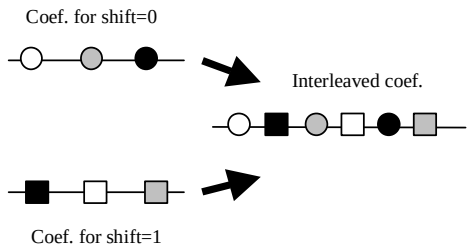

Figure 5: Interleaving of overcomplete wavelet coefficients for 1-level decomposition.

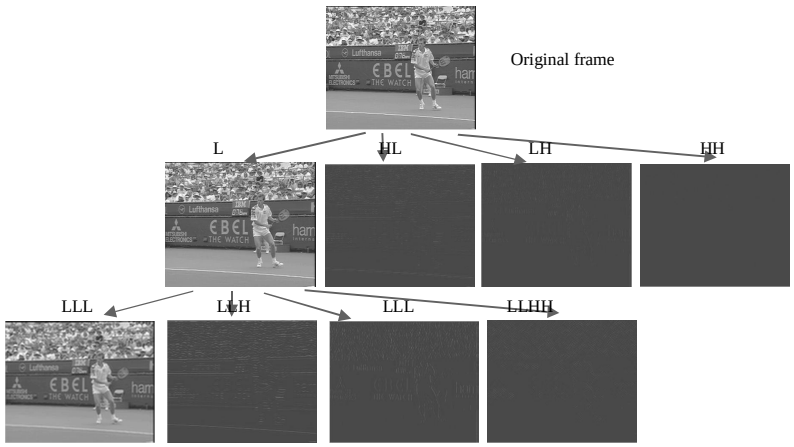

Figure 6: Overcomplete wavelet coefficients of the first frame of the "Stefan" sequence using the proposed interleaving algorithm.

Generation of Wavelet Block [10]

In the wavelet decomposition, every coefficient at a given scale, with the exception of those in the highest frequency subbands, can be related to a set of coefficients of the same orientation at finer scales. In many wavelet coders, this relationship is exploited by representing the coefficients as a data structure called a wavelet tree. In the LBS algorithm, the coefficients of each wavelet tree rooted in the lowest band are rearranged to form a wavelet block, as shown in Figure 7. The purpose of the wavelet block is to provide a direct association between the wavelet coefficients and what they represent spatially in the image. Related coefficients at all scales and orientations are included in each block.

Structure of Motion Estimation

In the spatial domain, the block-based motion estimation usually divides an image into small blocks and then finds the block of the reference frame that minimizes the MAD to each block of the current frame. The motion estimation of the LBS algorithm finds the motion vector (dx, dy) that generates the minimum MAD between the current wavelet block and the reference wavelet block. As an example, if an input image is decomposed up to the third level (i.e. the input image can be decomposed to a total of ten subbands), and the displacement vector is (dx,dy), then the MAD of the k-th wavelet block in Figure 7 is computed as follows:

$$MAD_k(dx,dy) = \sum_{i=1}^{3} \sum_{x_i=x_{i,k}}^{x_{i,k}+M/2^i} \sum_{y_i=y_{i,k}}^{y_{i,k}+N/2^i} \{$$

$$\left| HL_{cur}^{(i)}(x_i,y_i) - HL_{ref}^{(i)}(\, dx\%2^i, dy\%2^i; x_i + \left\lfloor \frac{dx}{2^i} \right\rfloor, y_i + \left\lfloor \frac{dy}{2^i} \right\rfloor) \right|$$

$$+ \left| LH_{cur}^{(i)}(x_i,y_i) - LH_{ref}^{(i)}(\, dx\%2^i, dy\%2^i; x_i + \left\lfloor \frac{dx}{2^i} \right\rfloor, y_i + \left\lfloor \frac{dy}{2^i} \right\rfloor) \right|$$

$$+ \left| HH_{cur}^{(i)}(x_i,y_i) - HH_{ref}^{(i)}(\, dx\%2^i, dy\%2^i; x_i + \left\lfloor \frac{dx}{2^i} \right\rfloor, y_i + \left\lfloor \frac{dy}{2^i} \right\rfloor) \right| \}$$

$$+ \sum_{x_i=x_{3,k}}^{x_{3,k}+M/2^i} \sum_{y_i=y_{3,k}}^{y_{3,k}+N/2^i} \left| LL_{cur}^{(3)}(x_i,y_i) - LL_{ref}^{(3)}(\, dx\%2^3, dy\%2^3; x_i + \left\lfloor \frac{dx}{2^3} \right\rfloor, y_i + \left\lfloor \frac{dy}{2^3} \right\rfloor) \right|$$

where $x_{i,k} = x_{0,k} / 2^i$ and $y_{i,k} = y_{0,k} / 2^i$; and $(x_{0,k}, y_{0,k})$ denotes the initial position of the k-th wavelet block in the spatial domain, as shown in Figure 7 and $\lfloor x \rfloor$ denotes largest integer not bigger than x. Here, for example, the i-th level HL band of the reference frame is represented by $HL_{ref}^{(i)}(m,n;x,y)$, where (m,n) denotes the number of shift in x- and y- direction in the spatial domain and (x,y) is the location of the subband signal. The optimization criterion for the motion estimation is now finding the optimal (dx,dy) which minimizes this MAD. Note that in the original LBS algorithm, for the non-integer value of (dx,dy), it is not possible to compute the MAD using the above formula. More specifically, the MAD of [10] is based solely on the same-phase wavelet coefficients and the resulting sub-pixel accuracy motion estimation and compensation is not optimal.

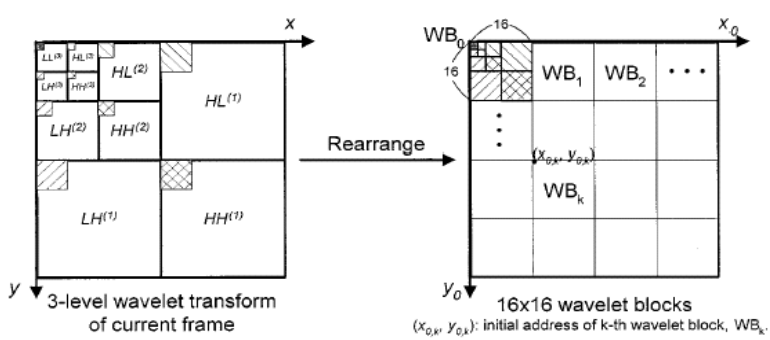

Figure 7: Generation of wavelet blocks from the 3-level wavelet transform [10].

However, in our proposed scheme, due to the interleaving, the MAD calculation is performed similarly as in SDMCTF, even for the sub-pixel accuracy. More specifically, the MAD for the displacement vector (dx,dy) for the proposed IBMCTF is computed as follows:

$$MAD_k(dx,dy) = \sum_{i=1}^{3} \sum_{x_i=x_{i,k}}^{x_{i,k}+M/2^i} \sum_{y_i=y_{i,k}}^{y_{i,k}+N/2^i} \{$$

$$\left| HL_{cur}^{(i)}(x_i,y_i) - LBS\_HL_{ref}^{(i)}(2^i x_i + dx,\, 2^i y_i + dy) \right| + \left| LH_{cur}^{(i)}(x_i,y_i) - LBS\_LH_{ref}^{(i)}(2^i x_i + dx,\, 2^i y_i + dy) \right|$$

$$+ \left| HH_{cur}^{(i)}(x_i,y_i) - LBS\_HH_{ref}^{(i)}(2^i x_i + dx,\, 2^i y_i + dy) \right| \}$$

$$+ \sum_{x_i=x_{3,k}}^{x_{3,k}+M/2^i} \sum_{y_i=y_{3,k}}^{y_{3,k}+N/2^i} \left| LL_{cur}^{(i)}(x_i,y_i) - LBS\_LL_{ref}^{(i)}(2^i x_i + dx,\, 2^i y_i + dy) \right|$$

where, for example, $LBS\_HL_{ref}^{(i)}(x,y)$ denotes the extended HL band of reference frame using proposed interleaving algorithm. Note that even if (dx,dy) are non-integer values, the same interpolation technique used for SDMCTF can be easily used for each extended subband to generate the MAD for the non-integer displacement. Therefore, the proposed scheme provides more efficient and indeed *optimal* sub-pixel motion estimation algorithm compared to the existing ones [8][9]. Also, using the proposed coding scheme with the wavelet block structure does not incur any motion vector overhead because the number of the motion vector to be coded is the same as that of SDMCTF. Since the motion estimation is closely aligned with the residual coding, a more sophisticated motion estimation criterion (such as the entropy of the residual signal) could be used to improve the coding performance. This is an interesting research topic that will be investigated in the future.

3. SUB-PIXEL ACCURACY IBMCTF USING 3-D LIFTING

Sub-pixel accuracy SDMCTF using Lifting [5]

To improve the quality of the matches provided by the motion estimation algorithm, sub-pixel accuracy motion estimation has been employed for SDMCTF using the lifting implementation [5]. This implementation guarantees perfect reconstruction for motion compensated temporal filtering in arbitrary sub-pixel accuracy. In the following, we first review this implementation for connected pixels. In the next section, this concept is extended for sub-pixel accuracy IBMCTF. (An example of unconnected and connected pixels and their handling is shown in Figure 8.)

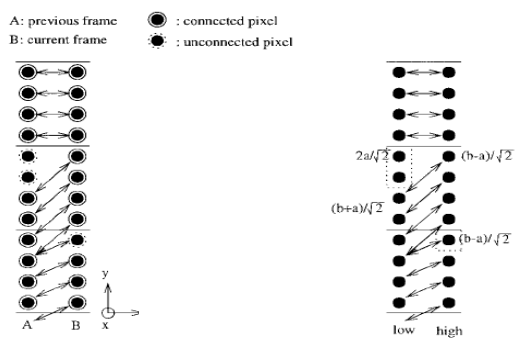

Figure 8: An example connected and unconnected pixels [6].

In Figure 8, if motion vectors have sub-pixel accuracy, the lifting scheme leads to the following expression for the temporal high band $H[m,n]$,

$$H[m,n] = \left(B[m,n] - \tilde{A}[m-d_m, n-d_n]\right)/\sqrt{2} \qquad (1)$$

where $\tilde{A}[x,y]$ denotes the interpolated pixel values of A frame at (x,y) location and B[m,n] denotes the (m,n)-th pixel in B frame. For the low pass filtered frame, we have

$$L\left[m-\bar{d}_m, n-\bar{d}_n\right] = \tilde{H}\left[m-\bar{d}_m+d_m, n-\bar{d}_n+d_n\right] + \sqrt{2}A\left[m-\bar{d}_m, n-\bar{d}_n\right] \qquad (2)$$

where (d_m, d_n) denotes a sub-pixel accuracy motion vector and $(\bar{d}_m, \bar{d}_n)$ denotes approximation to the nearest integer value lattice. At the decoder, by using L and H, we can do the same interpolation on H and reconstruct A exactly if there is no quantization error,

$$A\left[m-\bar{d}_m, n-\bar{d}_n\right] = \left(L\left[m-\bar{d}_m, n-\bar{d}_n\right] - \tilde{H}\left[m-\bar{d}_m+d_m, n-\bar{d}_n+d_n\right]\right)/\sqrt{2} \qquad (3)$$

After A is available, B can be reconstructed as

$$B[m,n] = \sqrt{2}H[m,n] + \tilde{A}[m-d_m, n-d_n]. \qquad (4)$$

Unconnected pixels in B are processed like (1), and unconnected (unreferred) pixels in A are processed as

$$L[m,n] = \sqrt{2}A[m,n] \qquad (5)$$

3-D Lifting Structure for IBMCTF

In order to derive the lifting structure for IBMCTF, let us consider a simple two level decomposition of a B frame as shown in Figure 9. The straightforward extension of Eq.(1) for IBMCTF leads to:

$$H_j^i[m,n] = \left(B_j^i[m,n] - \tilde{A}_j^i\left[m-d_j^i(m), n-d_j^i(n)\right]\right)\Big/\sqrt{2} \quad ,i = 0,...,3 \qquad (6),$$

where $d_j^i(m) = d_m / 2^j$, $d_j^i(n) = d_n / 2^j$ and (d_m, d_n) denotes the motion vector in spatial domain. However, in this structure, the interpolation operation for A_j^i frame is not optimal because it does not incorporate the dependencies of the cross-phase wavelet coefficients. Instead, our interleaving structure described in the previous section provides a *simple and optimal* motion compensated filtering structure:

$$H_j^i[m,n] = \left(B_j^i[m,n] - LBS\_\tilde{A}_j^i\left[2^j m - d_m, 2^j n - d_n\right]\right)\Big/\sqrt{2} \quad ,i = 0,...,3 \qquad (7)$$

where $LBS\_A_j^i$ denotes the interleaved overcomplete wavelet coefficients as depicted in Figure 6 and $LBS\_\tilde{A}_j^i\left[2^j m - d_m, 2^j n - d_n\right]$ denotes its interpolated pixel value at location $\left[2^j m - d_m, 2^j n - d_n\right]$. After interleaving, the interpolation operation is a simple spatial domain interpolation of the neighboring wavelet coefficients, similar to that used in SDMCTF. For the temporally low-pass filtered frame, we have

$$L_j^i\left[m-\bar{d}_j^i(m), n-\bar{d}_j^i(n)\right] = LBS\_\tilde{H}_j^i\left[2^j m - \bar{d}_m + d_m, n - \bar{d}_n + d_n\right] + \sqrt{2}A_j^i\left[m-\bar{d}_j^i(m), n-\bar{d}_j^i(n)\right] \qquad i = 0,..,3$$

(8)

where $d_j^i(m) = d_m / 2^j$ and $d_j^i(n) = d_n / 2^j$ and $LBS\_H_j^i$ denotes the interleaved overcomplete wavelet coefficients of the H_j^i frame.

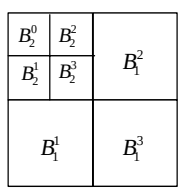

Figure 9: Two level wavelet decomposition example.

At the decoder side, a perfect reconstruction is still guaranteed:

$$A_j^i\left[m-\bar{d}_j^i(m), n-\bar{d}_j^i(n)\right] = L_j^i\left[m-\bar{d}_j^i(m), n-\bar{d}_j^i(n)\right]\Big/\sqrt{2} - LBS\_\tilde{H}_j^i\left[2^j m - \bar{d}_m + d_m, n - \bar{d}_n + d_n\right]\Big/\sqrt{2}$$

(9)

and $B_j^i[m,n] = \sqrt{2}H_j^i[m,n] + LBS\_\tilde{A}_j^i\left[2^j m - d_m, 2^j n - d_n\right]$ (10)

Note that perfect reconstruction can be realized independent on the interpolation method used, as long as the same method is employed at both the encoder and the decoder. Unconnected pixels in B are processed as in (8), and unconnected (unreferred) pixels in A_j^i are processed as

$$L_j^i[m,n] = \sqrt{2}A_j^i[m,n] \qquad (10)$$

4. SIMULATION RESULTS

The IBMCTF coder has been developed based on the SDMCTF interframe wavelet codec proposed in [7]. In the current implementation of the IBMCTF, the only difference from the SDMCTF is the motion estimation and filtering step. Hence, the residual images from the IBMCTF are coded using the EZBC intraframe texture coding algorithm of [7].

In order to verify that the ME/MC in the overcomplete wavelet domain yields lower residual energy in the wavelet domain, we use a one level temporal decomposition and compute the MAD for both IBMCTF and SDMCTF. Note that in interframe wavelet coding, the MAD is computed in the spatial-domain, but actually what needs to be minimized is the residual energy in the wavelet domain. Table 1 illustrates the MAD in wavelet domain for temporal high band frames. The MAD values are averaged over the first 50 frames of temporal high bands. For the SDMCTF cases, the corresponding MAD values in wavelet domains are computed after the wavelet transform of the residual signal. Note that the MAD for the IBMCTF is always smaller than for SDMCTF, which indicates the possible coding gain of the IBMCTF over SDMCTF.

Sequence		SDMCTF	IMBCTF
Stefan	Integer pixel	4.0825	4.0143
	Half pixel	3.3402	3.2732
Foreman	Integer pixel	1.8435	1.8414
	Half pixel	1.6400	1.6090
Mobile	Integer pixel	5.7399	5.7214
	Half pixel	4.1163	4.1035

Table 1: Comparison of average MAD in the wavelet domain for SDMCTF and IBMCTF from the first 50 temporal high band frames.

Figures 10-14 plot the R-D performance of the IBMCTF and the SDMCTF for several test sequence for integer and 1/8-pel accurate motion estimation. The inband structure for MCTF was computed using two level spatial decomposition using Daubechies 9/7 filter, and the four levels of decomposition was used for temporal direction. The texture coding is done using EZBC algorithm in [7]. Similar to SDMCTF, the sub-pixel motion estimation using 1/8 pel greatly improves the coding performance of the IBMCTF. The overall coding performance of the IBMCTF and SDMCTF is comparable. However, some sequences such as “Coastguard”, “Silent” and “Stefan” exhibit a performance gain of up to 0.5dB, while for the “Mobile” sequence a 0.3dB performance degradation can be observed. Visually, the IBMCTF algorithm is free of blocking artefacts of the motion estimation since the motion estimation and filtering is done in each subband and the boundary of the motion is filtered out using wavelet recomposition filter.

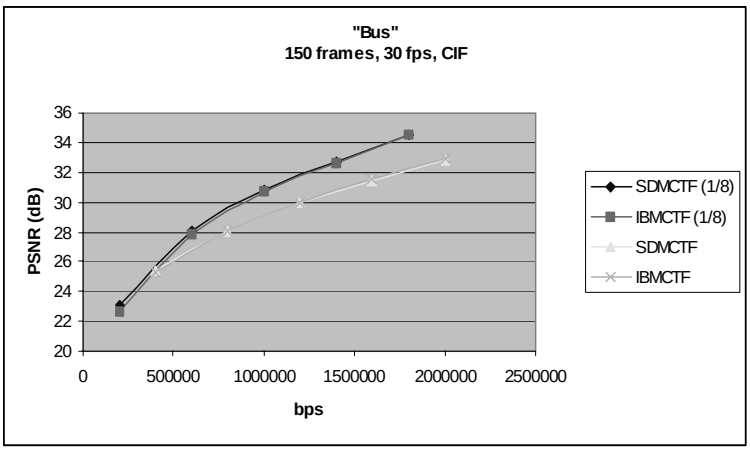

Figure 10: R-D performance of “Bus” sequence for IBMCTF and SDMCTF.

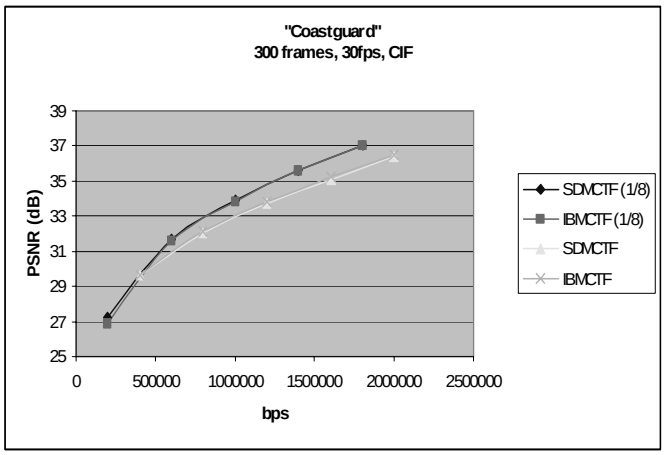

Figure 11: R-D performance of "Coastguard" sequence for IBMCTF and SDMCTF.

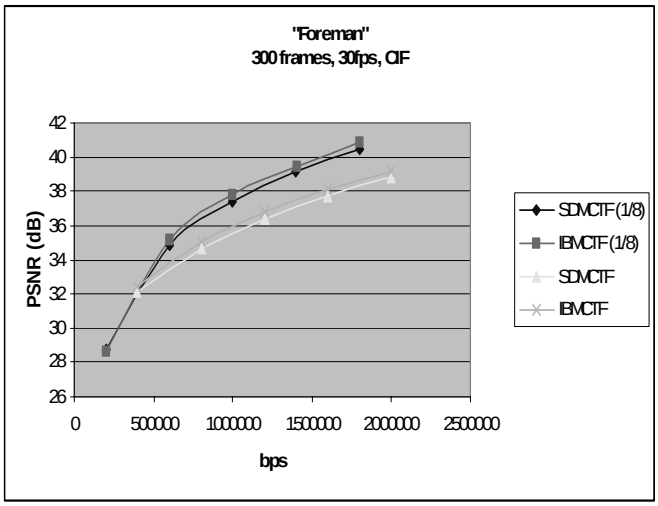

Figure 12: R-D performance of "Foreman" sequence for IBMCTF and SDMCTF.

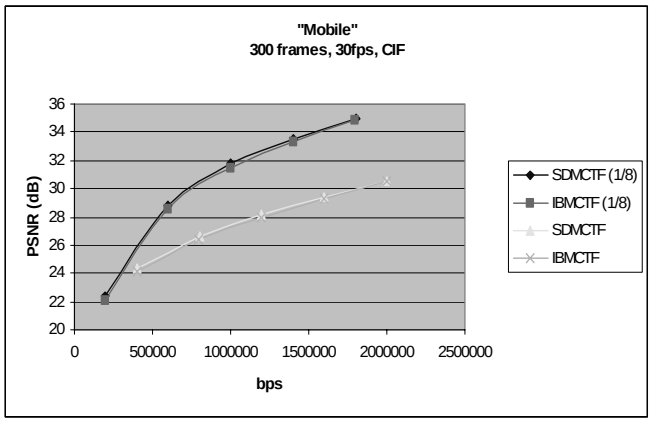

Figure 13: R-D performance of "Mobile" sequence for IBMCTF and SDMCTF.

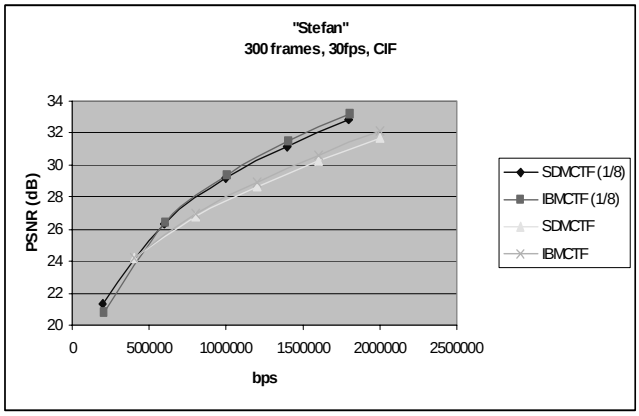

Figure 14: R-D performance of "Stefan" sequence for IBMCTF and SDMCTF.

5. CONCLUSION AND FUTURE RESEARCH DIRECTION

In this paper, we presented a novel 3-D lifting scheme that enables sub-pixel accuracy motion compensated temporal filtering in the overcomplete wavelet domain. In this scheme, we applied the MCTF for each band after the spatial domain wavelet transform. To overcome the inefficiency of motion estimation in the wavelet domain, the low band shifting method (LBS) was used to generate an overcomplete representation of the temporal reference frames both at the encoder and the decoder. This structure provides advantages over the SDMCTF in spatial scalability, blocking artifact reduction, intra mode coding, and etc. A novel interleaving algorithm for the overcomplete wavelet coefficient and the 3-D lifting algorithm for overcomplete wavelet domain temporal filtering were proposed to provide optimal implementation of sub-pixel accuracy motion estimation that was not possible in previously proposed IBMCTF schemes. Moreover, from an implementation perspective, the interleaving method enables usage of existing advanced motion estimation techniques (e.g. HVSBM etc.) used for SDMCTF. Simulation results demonstrated good coding performance similar or higher than conventional SDMCTF interframe wavelet coder.

Furthermore, the proposed adaptive IBMCTF framework can adapt the temporal filtering process in the various bands independently based on the spatial resolution, existing temporal correlations, content characteristics. Subsequently, we list briefly the various options which the IBMCTF framework allows and which will be future research topics.

- **Different accuracy of motion estimation.**

In interframe wavelet schemes the accuracy of the motion estimation and filtering is fixed. This is unfortunate, since different spatial resolutions require different accuracy. Alternatively, in the IBMCTF framework, we can vary the accuracy per band (each band corresponds to a specific resolution). Hence, for instance, we can employ coarse motion accuracy for lowest resolution subband while a finer motion accuracy is employed for finer resolution subbands.

- **Different prediction structures.**

Different temporal filters can be used at the various resolutions. For instance, bi-directional temporal filtering can be used for the low bands, while only forward temporal filtering can be used for the higher bands. Choosing a different filter can be done based on minimizing a distortion or a complexity measure (e.g. the low bands have less pixels and hence bi-directional and multiple reference temporal filtering can be employed, while for the high-pass bands that have a larger number of pixels, only forward estimation is performed).

- **Different GOF structures.**

The group of frames (GOF) to be filtered together by MCTF can also be adaptively determined per band. For instance, the LL-bands might have a very large GOF, while the H-bands can use limited GOFs. The GOF sizes can be varied based on the sequence characteristics, complexity or resiliency requirements.

Besides the adaptive temporal filtering, adaptive texture coding of the various spatial bands can also be employed (see Fig. 3). Wavelet or DCT-based schemes can be used. If DCT-based coding is used, we have the advantage that intra-coded blocks can be inserted anywhere within the GOF to deal efficiently with covering and uncovering situations. Also, 'adaptive intra refresh' concepts from MPEG-4/H.26L can be easily employed to provide improved resiliency and different refresh rates can be used for the various bands to obtain different resiliencies. This is especially beneficial since the lower resolution bands can be used for concealing the higher resolution bands and hence, their resiliency is more important.

Another advantage of the proposed framework is the complexity scalability of the decoder. If we have many devices with different computation power and displays, the same scalable bitstream can be used to support all those devices through SNR/spatial/temporal scalability. For example, the decoder with low complexity can decode only low resolution spatial and temporal decomposition level, which incurs only small computational burden, while a decoder with sophisticated decoding power can decode the whole bit stream to achieve the full spatial and temporal resolution. Also, the complexity of the temporal filtering or texture coding can increase for higher spatial subbands.

REFERENCES

[1] H. Gharavi, "Subband Coding of Video Signals" in *Subband Image Coding, Kluwer Academic Publishers*, 1991

[2] Y.-Q. Zhang and S. Zafar, "Motion-compensated wavelet transform coding for color video compression", *IEEE Trans. CSVT*, vol. 2, no.3, pp. 285-96, Sept. 1992.

[3] D. Taubman and A. Zakhor, "Multirate 3-D subband coding of video", *IEEE Trans. Image Proc.*, vol. 3, pp. 572–588, Sept. 1994.

[4] J. R. Ohm, "Three-dimensional subband coding with motion compensation", *IEEE Trans. Image Proc.*, vol. 3, no. 5, Sept. 1994.

[5] S.-T. Hsiang and J. W. Woods, "Embedded video coding using invertible motion compensated 3-D subband/wavelet filter bank", *Signal Processing: Image Commun.* vol. 16, pp. 705-724, 2001.

[6] S.-J. Choi and J. W. Woods, "Motion compensated 3-D subband coding of video," *IEEE Trans. Image Proc.*, vol. 8, no. 2, February 1999.

[7] 3D MC-EZBC Software package available at ftp://ftp.cipr.rpi.edu/personal/chen, see also S.T. Hsiang, J.W. Woods : "Invertible three-dimensional analysis/synthesis system for video coding with half-pixel accurate motion compensation", VCIP 1999, SPIE Vol. 3653, pp. 537-546.

[8] Y. Andreopoulos, M. van der Schaar, A. Munteanu, J. Barbarien, P. Schelkens, and J. Cornelis, "Open-loop, in-band, motion-compensated temporal filtering for objective full-scalability in wavelet video coding," *ISO/IEC JTC1/SC29/WG11*, *m9026*, MPEG meeting, Shanghai, China, October 2002.

[9] M. v d Schaar, J. Ye, Y. Andreopoulos, "Fully scalable 3-D overcomplete wavelet video coding using adaptive motion compensated temporal filtering," *ISO/IEC JTC1/SC29/WG11*, *m9037*, MPEG 62nd meeting, Shanghai, China, October 2002.

[10] H.-W. Park and H.-S. Kim, "Motion Estimation Using Low-Band-Shift Method for Wavelet-Based Moving-Picture Coding," *IEEE Trans. Image Processing*, vol. 9, no. 4, pp. 577-587, Apr. 2000.

[11] Joeri Barbarien, Yiannis Andreopoulos, Adrian Munteanu, Peter Schelkens, Jan Cornelis, "Coding of motion vectors produced by wavelet-domain motion estimation", *ISO/IEC JTC1/SC29/WG11*, *m9249*, MPEG 63nd meeting, Awaji Island, Japan, December, 2002.