Redundant Wavelet Transform in Video Signal Processing

Suxia Cui (presenter) and Yonghui Wang
Department of Engineering Technology
Prairie View A&M University, Prairie View, TX 77446
email: sucui@pvamu.edu and yowang@pvamu.edu
tel: 936-857-4717, fax: 936-857-2097

name of conference: The 2006 International Conference on Image Processing,
Computer Vision, & Pattern Recognition (IPCV'06)

***Abstract*— In this paper, the authors conduct a review on redundant wavelet transform (RDWT) in video signal processing. As an over-complete version of discrete wavelet transform (DWT), RDWT provides several advantages over the traditional DWT. In video coding area, we utilize its shift-invariant property and its phase-diverse multihypothesis, which are not provided by DWT, to build high performance video coding system. While in video watermarking area, the RDWT provides a more accurate estimation on locations where watermarks can be embedded. Thus, using RDWT, we can embed more watermarks than DWT approach without harming the non-perceivable requirement. After exploring several successful applications in video coding and video watermarking areas, the authors give a discuss on some future issues.**

***Index Terms*— redundant wavelet transform, multihypothesis, video coding, video watermarking**

I. INTRODUCTION

Discrete Wavelet Transform (DWT) has been successfully used in still image processing, and proved to be superior than the widely used Discrete Cosine Transform (DCT). The wavelet-based still image compression standard MPEG2000 and the wavelet-based algorithm, such as set partitioning in hierarchical trees (SPIHT) [1], all set great light to the future applcation of DWT. For the past decade, there are a lot of discussions on how to apply DWT in the signal processing techniques related to video sequences. Among them are video compression and watermarking.

The modern video compression technique is based on decorrelation of video sequence. A video signal can be recognized as a sequence of consequent still images. So a set of video signal is highly corelated spatially as well as temporally. To decorrelate a video sequence in the temporal domain, motion estimation and motion compensation (ME/MC) is widely used in all the current video coding standards. To decorrelate the video sequence spatially, the residual image obtained from ME/MC is decomposed into transform domain. In the video coding standards, such as H.264[2] and MPEG-4 [3], discrete cosine transform (DCT) is used to decorrelate the residual image. The diagram of this traditional architecture is shown in Fig. 1.

To replace DCT with DWT in video coding is not just an extension of the compression of frame by frame still images. If we directly repalce the DCT with DWT from Fig. 1, the whole frame DWT will suffer from blocking artifacts caused by block-based ME/MC. Another approach is to move the ME/MC into the wavelet domain. But the ME/MC procedure requires the transform to be shift-invariant. The DWT has a downsampling process after each level of filtering which makes it to have a shift variant property. This prevent us to get the precise motion vectors if we drive the ME/MC in DWT domain. To solve this problem, several wavelet-based video coders are proposed to use redundant wavelet transform (RDWT)[4, 5] instead of DWT. RDWT removes the downsampling procedure and is shift-invariant. Also, RDWT retains all the phase information of wavelet transform and provides mutiple prediction possibilities in ME/MC in transform domain. To consider all the different predictions, phase-diversity multihypothesis was proposed [6, 7]. So RDWT not only provides wavelet-based video coding with shift-invariant property, but also increases the precision of the motion vectors by increasing the number of hypothesis of the prediction.

RDWT is also very promising in watermarking of both image and video. Because the redundancy in the transform domain is more robust in carrying watermarking information. Also remove the downsampling procedure facilitates a better detection of the video texture characteristics in a video sequence.

This paper is organized as follows. In Sec. II, we provide an overview of the theory behind the overcompleted transform. In Sec. III, we introduce research achievements both in video coding and video watermarking areas followed by Sec. IV. At last in Sec. V, we conclude with the future goals that we are targeting at.

II. REDUNDANT WAVELET TRANSFORM

The Redundant wavelet transform (RDWT) can be considered to be an approximation to the continuous wavelet transform that removes the downsampling operation from the traditional critically sampled DWT to produce an overcomplete representation. The shift-variance characteristic of the DWT arises from its use of downsampling, while the RDWT is shift invariant since the spatial sampling rate is fixed across scale. The RDWT has been given several appellations over the years,

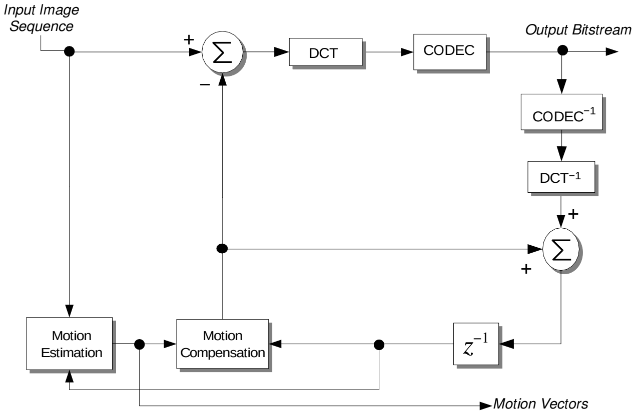

Fig. 1. The traditional hybrid coder with motion estimation and motion compensation (ME/MC) followed by a discrete cosine transform (DCT). z^{-1} = frame delay, *CODEC* is any still-image coder.

including the "undecimated DWT," the "overcomplete DWT," and the *algorithme à trous*. To describe the implementation of the RDWT in terms of filter-banks, let us first illustrate the same for the DWT.

A. Discrete Wavelet Transform

A 1D DWT and its inverse are illustrated in Fig. 2.

Here, $f[n]$ is the 1D input signal and $f'[n]$ is the reconstructed signal. $h[-k]$ and $g[-k]$ are the lowpass and highpass analysis filters, while the corresponding lowpass and highpass synthesis filters are $h[k]$ and $g[k]$. c_j and d_j are the low-band and high-band output coefficients at level j. DWT analysis, or decomposition, is, mathematically,

$$c_j[k] = (c_{j+1}[k] * h[-k]) \downarrow 2, \qquad (1)$$

and

$$d_j[k] = (c_{j+1}[k] * g[-k]) \downarrow 2, \qquad (2)$$

where $*$ denotes convolution, and $\downarrow 2$ denotes downsampling by a factor of two. That is, if $y[n] = x[n] \downarrow 2$, then

$$y[n] = x[2n]. \qquad (3)$$

The corresponding operation of DWT synthesis, or reconstruction, is

$$c_{j+1}[k] = (c_j[k] \uparrow 2) * h[k] + (d_j[k] \uparrow 2) * g[k], \qquad (4)$$

where $\uparrow 2$ denotes upsampling by a factor of two. That is, if $y[n] = x[n] \uparrow 2$, then

$$y[n] = \begin{cases} x[n/2], & n \text{ even}, \\ 0, & n \text{ odd}. \end{cases} \qquad (5)$$

In contrast, a 1D RDWT and its inverse are illustrated in Fig. 3. The RDWT eliminates downsampling and upsampling of coefficients, and at each scale, the number of output coefficients doubles that of the input. The filters themselves are upsampled to fit the growing data length. Specifically, the filters for scale j are

$$h_j[k] = h_{j+1}[k] \uparrow 2, \qquad (6)$$

and

$$g_j[k] = g_{j+1}[k] \uparrow 2. \qquad (7)$$

RDWT analysis is then

$$c_j[k] = (c_{j+1}[k] * h_j[-k]), \qquad (8)$$

and

$$d_j[k] = (c_{j+1}[k] * g_j[-k]), \qquad (9)$$

while RDWT synthesis is

$$c_{j+1}[k] = \frac{1}{2}(c_j[k] * h_j[k] + d_j[k] * g_j[k]). \qquad (10)$$

(6) through (10) are known as the *algorithme à trous* [8], since the filter-upsampling procedure inserts "holes" ("trous" in French) between the filter taps.

B. Redundant Wavelet Transform

The RDWT [9] removes the downsampling operation from the traditional critically sampled DWT to produce an overcomplete representation. The shift-variance characteristic of the DWT arises from its use of downsampling, while the RDWT is shift invariant since the spatial sampling rate is fixed across scale. As a result, the size of each subband in an RDWT is the exactly the same as that of the input signal.

By appropriately subsampling each subband of an RDWT, one can produce exactly the same coefficients as does a critically sampled DWT applied to the same input signal. In fact, in a J-scale 1D RDWT, there exist 2^J distinct critically sampled DWTs corresponding to the choice between even- and odd-phase subsampling at each scale of decomposition. In the *algorithme à trous* [9] implementation, each block of 2^J coefficients at scale J of a J-scale 1D RDWT contains exactly one coefficient from each of these 2^J critically sampled DWTs. Since this block gathers all possible subsampling phases at this scale, we call this structure a "phase group." We can similarly define phase groups for scales $j < J$. These phase groups contain 2^j distinct phases, and there is a tree-like relationship among the phase groups at different scales. Specifically, each of the 2^j phases at scale j begets two phases at scale 2^{j+1}.

The situation is similar in 2D. A J-scale 2D RDWT consists of 4^J distinct critically sampled DWTs, while a phase group at scale j is a block of $2^j \times 2^j$ coefficients defined so as to contain all 4^j possible phases at that scale. As in the 1D transform, there is a tree-like relationship between the phases at different scales. Specifically, in the 2D RDWT, each phase at scale j begets four phases at scale $j+1$.

To invert a J-scale 2D RDWT, one can independently invert each of the 4^J critically sampled DWTs constituting the RDWT and average the resulting reconstructions together. However, this implementation of the inverse RDWT incurs unnecessary duplicate synthesis filterings of the highpass bands; thus, one usually alternates between synthesis filtering and reconstruction averaging on a scale-by-scale basis in practical implementations. The final reconstruction of this practical

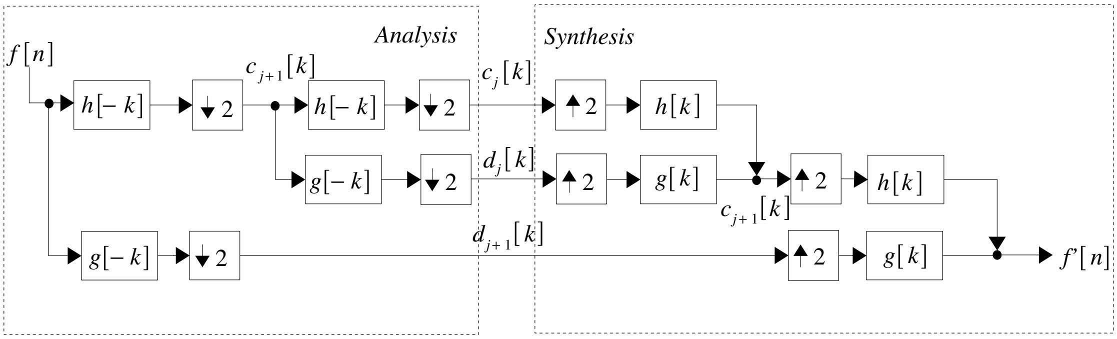

Fig. 2. Two level 1-D DWT analysis and synthesis filter banks.

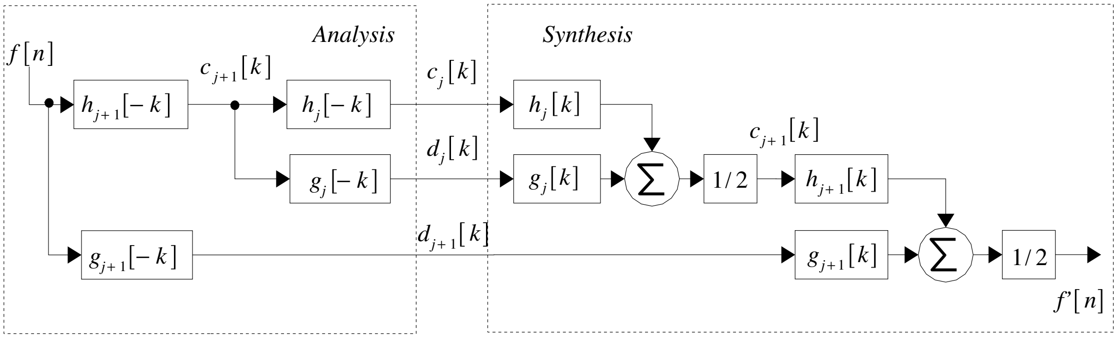

Fig. 3. Two level 1-D RDWT analysis and synthesis filter banks.

implementation, however, is identical to that produced by the conceptually simpler DWT-based approach.

C. Multihypothesis in Redundant Wavelet Transform Domain

Multihypothesis MC (MHMC) [10] forms a prediction of pixel $s(x, y)$ in the current frame as a combination of multiple predictions in an effort to combat the uncertainty inherent in the ME process. Assuming that the combination of these hypothesis predictions is linear, we have that the prediction of $s(x, y)$ is

$$\tilde{s}(x, y) = \sum_i w_i(x, y)\tilde{s}_i(x, y), \tag{11}$$

where the multiple predictions $\tilde{s}_i(x, y)$ are combined according to some weights $w_i(x, y)$. A number of MHMC techniques have been proposed over the last decade. Among them are fractional-pixel MC [11] and overlapped block motion compensation (OBMC) [12, 13], which we can category them into the spatial domain diversity multihypothesis. Bidirectional prediction (B-frames) as used in MPEG-2 and H.263 and long-term-memory motion compensation (LTMMC) [14] use the temporal domain diversity multihypothesis. RDWT provide us with another option in transform domain, because RDWT retains all the phase information by get rid of the downsampling procedure. Each of the critically sampled DWTs within a RDWT will "view" motion from a different perspective. Consequently, if motion is predicted in the RDWT domain, the inverse RDWT forms a multihypothesis prediction in the form of (11). Specifically, for a J-scale RDWT, the reconstruction from DWT i of the RDWT is $\tilde{s}_i(x, y)$, $0 \leq i < 4^J$, while $w_i(x, y) = 4^{-J}$, $\forall i$. So redundant wavelet multihypothesis (RWMH)[6] was proposed based on the phase-diversity multihypothesis in RDWT domain.

III. RWMH IN VIDEO CODING

As we have shown the algorithm behind the RDWT and RWMH, now we'll show some applications in video signal processing to illustrate the achievements that we have made by adopting multihypothesis in the transform domain. The experiments that we have done can be categoried into two major areas, one is video coding, the other is video watermarking. In this section, we'll introduce two approaches in video coding. In next section, we'll show the experiments in watermarking. Nowadays, the current video coding standard are based on the traditional close loop ME/MC system, we call that 2D system. The system diagram can be illustrated in Fig. 1, each video frame is processed individually, the next frame to be processed is based on the previous decoded information. Another scheme is the open loop 3D system which is becoming more and more important with the fast growing internet requirement.

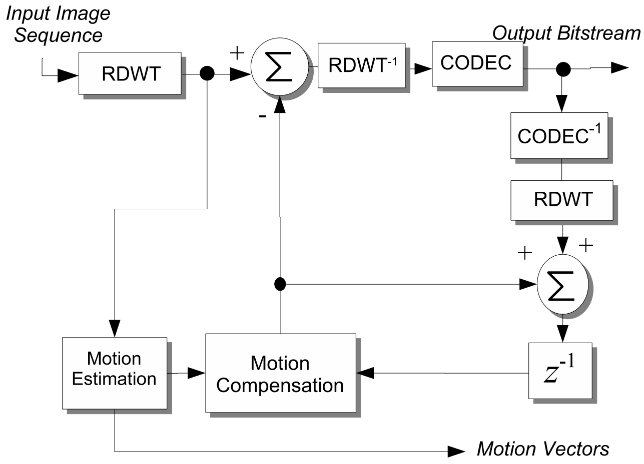

Fig. 4. The RWMH_OBMC coder. z^{-1} = frame delay, *CODEC* is any still-image coder.

A. RWMH in 2D Video Coding Application

As has been stated in Sec. I, in the traditional diagram of a video coder, to replace the DCT with DWT is not promising because of the "shift-variant" property caused by downsampling. Now use RDWT without downsampling is one solution. The new system diagran is shown in Fig. 4, we can see the major change is that firstly the RDWT is performed outside of the ME/MC loop so both motion estimation and motion compensation are dealing with the RDWT coefficients instead with the spatial domain values. Secondly, after motion compensation, we need an inverse of the RDWT coefficients to finalize the mutihypothesis prediction. Since the weighting of the individual predictions is carried out implicitly in the form of an inverse transform, no additional side information need be sent to the decoder[6]. Another aspect of RWMH approach is that the spatial coherence of RDWT coefficients reside at the same location as the spatial domain coefficients which makes the combination with other MHMC techniques, such as OBMC, fractional-pixel accuracy, and LTMMC possible and we can further improve the performance. Experiment results show at least 1dB gain over both low motion sequences and high motion sequences[6].

B. RWMH in 3D Video Coding Application

The diagram of 3D video coding system [15] is shown in Fig. 5. In this approach, we do not have the closed loop as in the traditional 2D video coding system. This change makes the 3D coder do not depend on the previous decoded frame which leads to the facts that the fully scalability to be a possible. The encoder of our 3D-RWMH video-coding system, depicted in Fig. 5, first performs a spatial RDWT on each frame and then performs MCTF in the redundant-wavelet domain. This is in contrast to many prior MCTF techniques [16–19] in which MCTF takes place in the spatial domain. Since MCTF is performed in the RDWT subbands, it is overcomplete spatially; consequently, before coding the temporal subbands, we remove this spatial redundancy by performing an inverse spatial RDWT on each frame. In essence, each RDWT phase in each frame can be considered to have viewed the MCTF from a different perspective and thus forms an independent hypothesis about the temporal filtering taking place. The inverse spatial RDWT implicitly combines these hypotheses into a multihypothesis estimate of what the true temporal filtering should be. After the inverse spatial transform, the temporally transformed frames are coded by a suitable 3D coder. In this system, 3D-SPIHT [20] is used, but other coders are possible. In the 3D-RWMH system, motion is tracked using a triangular mesh deployed in each of the subbands of the RDWT decomposition of each frame. Since all RDWT subbands are the same size, the same triangle mesh is used for all subbands of a frame. Experimental results show that for high-motion video sequences, such as "Football," a significant gain on the order of 0.5 dB over the spatial-domain system is seen.

IV. RDWT IN VIDEO WATERMARKING

In [21], we proposed a novel video watermarking system by using redundant wavelet transform, which is PWM-RDWT-3D. This algorithm applies the pixel-wise masking technique (PWM [22]) used in image watermarking into video sequence. In this watermarking system, the core technology is to find the most correct locations to embed the watermarks. We use a weighing function to direct the strength of watermarks to be embedded to different pixels. This weighing function is defined as

$$w_0^\theta(i,j,k) = \Theta(0,\theta)\Lambda(0,i,j,k)\Xi(0,i,j,k)^{0.2} \tag{12}$$

where 0 means level 0. We decide to embed watermarks only in the three highest subbands, i.e., $l = 0$. The three terms of this weighing function are based on the facts of:

- The eye is less sensitive to noise in high resolution subbands.
- The eye is less sensitive to noise in those volume of the video sequence where brightness is high or low.
- The eye is less sensitive to noise in highly textured volumes, but more sensitive near the edges.

This algorithm deploys the video watermarking in the RDWT domain. The advantage of using an over-completed wavelet transform instead of the traditional critically sub-sampled discrete wavelet transform is that the redundancy in the transform domain facilitates a better detection of the image texture in a video sequence. Thus leads to an efficient watermark-casting scheme. This scheme first applies a temporal 1D DWT transform to the original video sequence. Then a 2D RDWT transform is performed on each individual video frame. The PWM method proposed in [22] is used to embed watermark in those high texture volumes where the HVS is less sensitive to the added watermark. This makes the PWM-RDWT-3D enable to capture more accurately the video textures than the PWM-DWT-3D. Experimental results show that PWM-RDWT-3D is more robust than the PWM-DWT-3D algorithm.

V. CONCLUSIONS

In this paper, we review the RDWT application in video signal processing. As an over-complete version of wavelet

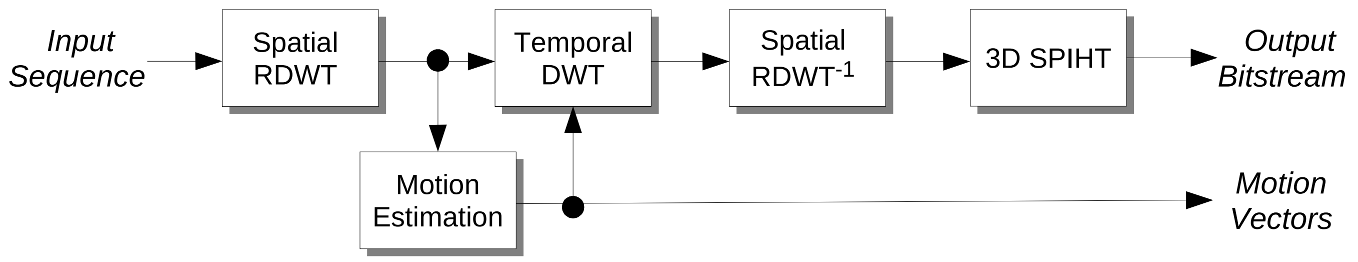

Fig. 5. The RWMH-3D video coding system.

transform, RDWT demostrates more advantages over DWT. In video coding, we utilize its shift-invariant characteristic. This characteristic makes sure motion compensation/estimation can be done in wavelet domain. Also, we build a new multihypothesis scheme in wavelet domain through RDWT, which we call it as RWMH. During motion compensation, We build a multihypothesis by implicitly averaging all the phases in wavelet domain. In video watermarking, RDWT can retain the spatial location of all the salient features across all the subbands thanks to the nonsubsampling operation. This makes the RDWT watermarking enable to capture more accurately the video textures than the DWT watermarking.

While RDWT provides so many advantages over DWT, RDWT does have several drawbacks. The first one is that RDWT will take up more memory space due to its redundancy. Large memory requirement hinders its implementation in hardware. The second one is that the computational complexity of RDWT is higher than that of DWT. This makes RDWT based video coding system slower than DWT based system. Our future work will be focus on solving these problems. Also, based on the success of RDWT in video coding and video watermarking areas, the authors expect that RDWT can be applied into other video processing areas.

VI. REFERENCES

[1] A. Said and W. A. Pearlman, "A new, fast, and efficient image codec based on set partitioning in hierarchical trees," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 6, no. 3, pp. 243–250, June 1996.

[2] ITU-T, *Advanced Video Coding for Generic Audiovisual Services*, May 2003, ITU-T Recommendation H.264.

[3] ISO/IEC 14496-2, *Information Technology—Coding of Audio-Visual Objects—Part 2: Visual*, 1999, MPEG-4 Coding Standard.

[4] H.-W. Park and H.-S. Kim, "Motion estimation using low-band-shift method for wavelet-based moving-picture coding," *IEEE Transactions on Image Processing*, vol. 9, no. 4, pp. 577–587, April 2000.

[5] S. Cui, Y. Wang, and J. E. Fowler, "Mesh-based motion estimation and compensation in the wavelet domain using a redundant transform," in *Proceedings of the International Conference on Image Processing*, Rochester, NY, September 2002, vol. 1, pp. 693–696.

[6] S. Cui, Y. Wang, and J. E. Fowler, "Multihypothesis motion compensation in the redundant wavelet domain," in *Proceedings of the International Conference on Image Processing*, Barcelona, Spain, September 2003, vol. 2, pp. 53–56.

[7] Y. Wang, S. Cui, and J. E. Fowler, "3D video coding using redundant-wavelet multihypothesis and motion-compensated temporal filtering," in *Proceedings of the International Conference on Image Processing*, Barcelona, Spain, September 2003, vol. 2, pp. 755–758.

[8] P. Dutilleux, "An implementation of the "algorithme à trous" to compute the wavelet transform," in *Wavelets: Time-Frequency Methods and Phase Space*, J.-M. Combes, A. Grossman, and P. Tchamichian, Eds., pp. 298–304. Springer-Verlag, Berlin, Germany, 1989, Proceedings of the International Conference, Marseille, France, December 14–18, 1987.

[9] M. J. Shensa, "The discrete wavelet transform: Wedding the à trous and Mallat algorithms," *IEEE Transactions on Signal Processing*, vol. 40, no. 10, pp. 2464–2482, October 1992.

[10] G. J. Sullivan, "Multi-hypothesis motion compensation for low bit-rate video coding," in *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing*, Minneapolis, MN, April 1993, vol. 5, pp. 437–440.

[11] B. Girod, "Motion-compensating prediction with fractional-pel accuracy," *IEEE Transactions on Communications*, vol. 41, no. 4, pp. 604–612, April 1993.

[12] S. Nogaki and M. Ohta, "An overlapped block motion compensation for high quality motion picture coding," in *Proceedings of the IEEE International Symposium on Circuits and Systems*, San Diego, CA, May 1992, vol. 1, pp. 184–187.

[13] M. T. Orchard and G. J. Sullivan, "Overlapped block motion compensation: An estimation-theoretic approach," *IEEE Transactions on Image Processing*, vol. 3, no. 5, pp. 693–699, September 1994.

[14] T. Wiegand, X. Zhang, and B. Girod, "Long-term memory motion-compensated prediction," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 9, no. 1, pp. 70–84, February 1999.

[15] Y. Wang, S. Cui, and J. E. Fowler, "3D video coding with redundant-wavelet multihypothesis," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 16, no. 2, pp. 166–177, February 2006.

[16] A. Secker and D. Taubman, “Lifting-based invertible motion adaptive transform (LIMAT) framework for highly scalable video compression,” *IEEE Transactions on Image Processing*, vol. 12, no. 12, pp. 1530–1542, December 2003.
[17] J.-R. Ohm, “Three-dimensional subband coding with motion compensation,” *IEEE Transactions on Image Processing*, vol. 3, no. 5, pp. 559–571, September 1994.
[18] P. Chen and J. W. Woods, “Comparison of MC-EZBC and H.26L TML 8 on digital cinema test sequences,” ISO/IEC JTC1/SC29/WG11, MPEG2002/M8130, Jeju Island, South Korea, March 2002.
[19] B. Pesquet-Popescu and V. Bottreau, “Three-dimensional lifting schemes for motion compensated video compression,” in *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing*, Salt Lake City, UT, May 2001, vol. 3, pp. 1793–1796.
[20] B.-J. Kim, Z. Xiong, and W. A. Pearlman, “Low bitrate scalable video coding with 3-D set partitioning in hierarchical trees (3-D SPIHT),” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 10, no. 8, pp. 1374–1387, December 2000.
[21] Y. Wang and S. Cui, “Using redundant wavelet transform in video watermarking,” in *Proceedings of SPIE*, Boston, MA, October 2005, vol. 6015.
[22] M. Barni, F. Bartolini, and A. Piva, “Improved wavelet-based watermarking through pixel-wise masking,” *IEEE Transactions on Image Processing*, vol. 10, no. 5, pp. 783–791, May 2001.