

3-D Direction Aligned Wavelet Transform for Scalable Video Coding

Yu Liu[†], King Ngi Ngan[†], and Feng Wu[‡]

[†]Department of Electronic Engineering, The Chinese University of Hong Kong, Hong Kong

[‡]Internet Media Group, Microsoft Research Asia, Beijing, 100080, China

Abstract—This paper presents a novel 3-D direction aligned wavelet transform for scalable video coding. A new generalized 3-D directional threading technique is seamlessly incorporated into 3-D weighted adaptive lifting (WAL)-based wavelet transform to exploit the spatio-temporal correlation inside the video cube along the 3-D directional trajectory, leading to a new class of algorithm called 3-D direction aligned wavelet transform (DAWT). Experimental results show that the proposed 3-D WAL-based DAWT achieves better coding performance than the conventional 3-D discrete wavelet transform (3-D DWT).

I. INTRODUCTION

Wavelet video coding scheme can provide flexible spatial, temporal, quality and complexity scalability with fine granularity over a large range of bit rates, while maintaining high coding efficiency which is comparable with H.264-based JSVM scheme. The 3-D discrete wavelet transform (DWT) plays a important role in the wavelet video coding. In wavelet video coding, 3-D DWT is usually implemented by first applying 1-D lifting-based wavelet transform in the motion trajectory, which is generally called motion compensated temporal filtering (MCTF) or motion aligned temporal filtering (MATF), then applying 2-D lifting-based wavelet transform in the spatial domain. In the spatial domain, the conventional lifting scheme uses the elements in neighbor horizontal or vertical direction. Although very efficient in representing the horizontal and vertical edges, this kind of 2-D lifting structure does not work well when the edges are neither horizontal nor vertical. As a matter of fact, natural image/video often contain richly directional attributes, which can be commonly approximated as linear edges on a local level. These edges may be neither vertical nor horizontal. If not taken into account, such a fact would result in large magnitude in high frequency coefficients. This problem has been pointed out by many researchers in image processing field. Amongst them, directional wavelet transforms, such as adaptive directional lifting (ADL) [1], direction-adaptive DWT [2] and WAL [3], have demonstrated significantly subjective and objective quality improvement over conventional wavelet transform for image coding.

However, to the authors' best knowledge, there is no literature incorporating the directional wavelet transform into the framework of scalable video coding. In this paper, we will address scalable video coding with 3-D directional wavelet transform. The main contribution of this paper consists of two part: (1) a new generalized 3-D directional threading, which unifies the concept of temporal motion threading and

spatial directional threading, is proposed to exploit the spatio-temporal correlation inside the 3-D video cube; (2) the weighted adaptive lifting (WAL) scheme, which has achieved better coding performance than conventional lifting scheme and ADL scheme, is extended from 2-D image coding to 3-D video coding. Based on 3-D directional threading and 3-D weighted adaptive lifting, a new 3-D direction aligned wavelet transform (DAWT) is proposed for scalable video coding.

The remainder of this paper is organized as follows. Section II describes the proposed 3-D directional threading technique. The extension of weighted adaptive lifting to 3-D spatio-temporal transforms is presented in Section III. The experimental results are presented in Section IV, and Section V concludes the paper.

II. 3-D DIRECTIONAL THREADING

As inspired by temporal motion threading [4], [5], spatial directional threading can also be developed in the similar way. In temporal motion threading, pixels along the same motion trajectory are linked to form a thread according to the motion vectors of the blocks they belong to. For 2-D spatial signals, if the edges are not horizontal or vertical, we can get a better description by aligning the direction of the wavelet filtering to the direction of the edges, which leads to direction aligned spatial filtering (DASF). The principle of temporal motion threading for MATF can also be applied to the case of DASF, called 2-D spatial directional threading, which consists of two separable 1-D threading, horizontal directional threading and vertical directional threading. The displacements in horizontal and vertical axis form a spatial direction vector. Like temporal motion threading, pixels along the same directional trajectory are also linked to form a spatial directional thread according to spatial direction vectors of the blocks they belong to.

Combining the concept of direction vector in spatial domain with the concept of motion vector in temporal domain, a new 3-D direction coordinate system in a unified framework can be devised. Fig. 1 shows the coordinate system of 3-D direction, where x , y , and z denote the horizontal, vertical, and temporal direction, respectively. The motion vector, $mv = \{dx, dy\}$, in temporal domain can be extended to the 3-D direction domain, which leads to the 3-D direction vector, $dv = \{dx, dy, dz\}$. Each component in the direction vector dv represents the corresponding displacement in each direction. To reduce the overhead bits representing the directional information, the tree-structured macroblock partitions in H.264/AVC is adopted and

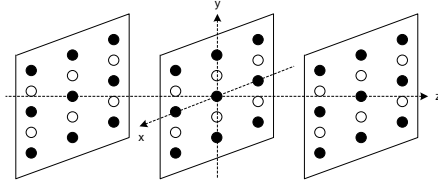


Fig. 1. 3-D Direction Coordinate System

a direction vector is assigned to each block. The dz component is used to indicate whether the current block goes through the temporal direction compensation process. In other words, the cases $dz = -1$, $dz = 1$, and $dz = 0$ indicate that the current block is forward, backward and not temporal direction compensated, respectively. For block that is not temporal direction compensated, it should go through the spatial direction prediction process. The dx and dy components represent the displacements in the horizontal and vertical axis no matter what the value of dz is. When bidirectional temporal direction compensated, the current block should be represented by two direction vectors. Our study shows that the temporal direction compensation process is so efficient that the energy of the prediction errors approximates to zero and the further spatial direction prediction can be saved, and vice versa. Therefore, if a block is temporal direction compensated ($dz \neq 0$), no further spatial direction prediction is needed and we only use default horizontal/vertical direction to predict it in the spatial transform, just like the conventional wavelet transform in the spatial domain. Otherwise, for a block that is not temporal direction compensated ($dz = 0$), it should go through further spatial direction prediction process after default temporal direction is applied to predict it in the temporal transform.

As seen from another view point, motion threading technique can also be considered as a special case of directional threading in temporal domain since the motion vector indicates the directional information of the pixel's movement in the temporal domain. Therefore, to unify the concepts of temporal motion threading and 2-D spatial directional threading, a new generalized 3-D directional threading is proposed to exploit the spatio-temporal correlation inside the video cube along the 3-D directional trajectory. In the 3-D directional threading technique, temporal directional threading and 2-D spatial directional threading are carried out separately; and 2-D spatial directional threading also involves two 1-D threading: horizontal directional threading and vertical directional threading, as shown in Fig. 2.

III. 3-D EXTENSION OF WEIGHTED ADAPTIVE LIFTING

A. Improved Weighted Lifting for 3-D Spatio-Temporal Transforms

In original WAL scheme [3], the weighted function f_i is used in the lifting stage, which can be any linear function that takes any pixels in the even subset as variables:

$$f_i(x_e) = \sum_k w_{i,k} x_e[m + \Delta m_{i,k}, n + \Delta n_{i,k}], \quad (1)$$

$$\text{under the constraint: } \sum_k w_{i,k} = 1; \quad (2)$$

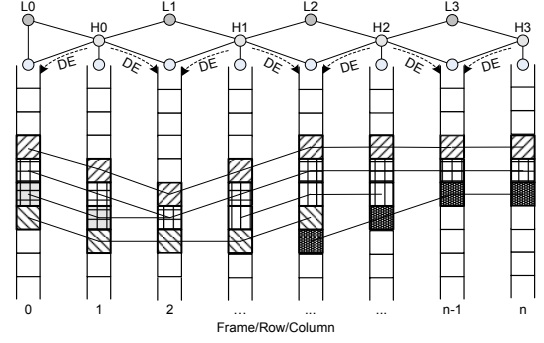


Fig. 2. The generalized separable 3-D directional threading

To solve the mismatch problem in the ADL scheme [1], the weighted lifting operators are proposed to make sure that the *predict* and *update* steps are consistent, as follows:

$$d[m, n] = x_o[m, n] + \quad (3)$$

$$\sum_{i=0}^1 p_i \sum_k w_{i,k} x_e[m + \Delta m_{i,k}, n + \Delta n_{i,k}],$$

$$c[m, n] = x_e[m, n] + \quad (4)$$

$$\sum_{j=-1}^0 u_j \sum_l w_{j,l} d[m + \Delta m_{j,l}, n + \Delta n_{j,l}].$$

The above weighted lifting operators work well for the 5/3-tap wavelet filter, not only in spatial transform [3] but also in temporal transform [6]. However, it still has some problems, such as over/under-weighted update problems, which may result in coding inefficiency and even annoying boundary artifacts when the 9/7-tap wavelet filter is applied. Since pixels with different directions are continuously processed, it may cause the boundary artifacts, especially when block-based direction prediction scheme is employed. In addition, multiple lifting stages, such as the 9/7-tap wavelet filter, further enlarge the annoying boundary artifacts because of the over/under-weighted update in the intermediate lifting stages.

The main reason of the over/under-weighted update problems lies in that the update operator Eq. (4) does not fulfill the constraint condition $\sum_l w_{j,l} = 1$. When the sum of weighted parameters $\sum_l w_{j,l}$ is greater than 1, we call it over-weighted update problem. On the other hand, when the sum of weighted parameters $\sum_l w_{j,l}$ is less than 1, we call it under-weighted update problem. In order to solve these problems, we propose the improved weighted update operators to amend Eq. (4) to keep the update step under the weighted balance and maintain simultaneously the consistency between the predict and update steps as far as possible.

$$c[m, n] = x_e[m, n] + \quad (5)$$

$$\sum_{j=-1}^0 u_j \{ \beta_j \sum_l w_{j,l} d[m + \Delta m_{j,l}, n + \Delta n_{j,l}] + \gamma_j \}$$

where $\beta_j, \gamma_j, (j = -1, 0)$ are the amendment parameters which vary according to different cases as follows:

Case (1) For the over-weighted update problem, we amend Eq. (4) with the *normalized* parameters to update the even pixels:

$$\beta_{-1} = \frac{1}{\overline{W}_{-1}}, \beta_0 = \frac{1}{\overline{W}_0}, \gamma_{-1} = \gamma_0 = 0, \quad (6)$$

$$\text{if } \overline{W}_{-1} \geq 1 \text{ and } \overline{W}_0 \geq 1;$$

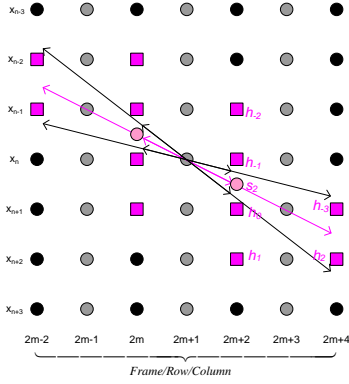


Fig. 3. Directional adaptive interpolation for 3-D spatio-temporal transforms

Case (2) For the over-and-under-weighted update problem, we amend Eq. (4) with the *symmetric-extension compensated* parameters to update the even pixels:

$$\begin{cases} \beta_{-1} = \frac{2-\overline{W}_0}{\overline{W}_{-1}}, \beta_0 = 1, \gamma_{-1} = \gamma_0 = 0, \\ \quad \text{if } \overline{W}_{-1} \geq 1 \text{ and } \overline{W}_0 < 1; \\ \beta_{-1} = 1, \beta_0 = \frac{2-\overline{W}_{-1}}{\overline{W}_0}, \gamma_{-1} = \gamma_0 = 0, \\ \quad \text{if } \overline{W}_{-1} < 1 \text{ and } \overline{W}_0 \geq 1; \end{cases} \quad (7)$$

Case (3) For the under-weighted update problem, we amend Eq. (4) with the *inverse-direction compensated* parameters to update the even pixels:

$$\begin{cases} \beta_{-1} = \beta_0 = 1, \\ \gamma_j = (1 - \overline{W}_j) DA_{2m-2(m+j)+1}(d)[m+j, n], \\ \quad \text{if } \overline{W}_{-1} < 1 \text{ and } \overline{W}_0 < 1; \end{cases} \quad (8)$$

B. Directional Adaptive Interpolation for 3-D Spatio-Temporal Transforms

To improve the orientation property of the interpolated image and adapt to statistical property of each image, a directional adaptive interpolation is proposed in the original WAL scheme. The new interpolation method consists of two part: (1) directional interpolation, which indicates the adaptation of the interpolation to directional information used for the predict step; and (2) adaptive interpolation filter, which is designed to minimize the energy of the direction matching error.

The directional interpolation not only can be applied to spatial domain, but also can be extended to temporal domain, as shown in Fig. 3. In temporal domain, in order to interpolate the sub-pixel s_2 in frame $2m+2$, not only the integer pixels $\{h_{-2}, h_{-1}, h_0, h_1\}$ in frame $2m+2$, but also the integer pixels $\{h_{-3}, h_2\}$ in frame $2m+4$ along the predicted direction vector are used. For this purpose, the direction vector between frame $2m+1$ and frame $2m+2$ will be spread to reach the frame $2m+4$.

Nevertheless, the predicted direction vector between frame $2m+2$ and frame $2m+4$ may not be accurate or even wrong. In order to reduce the direction prediction error and improve further the coding efficiency for each individual frame, the adaptive interpolation filter can also be used in the case of temporal domain. In order to estimate the filter coefficients of the

adaptive directional interpolation filter for temporal domain, the minimization problem can also be solved by the Wiener-Hopf equation. Finally, the optimal filter and the default filter will be decided by the rate-constrained filter selection process, just like the WAL scheme for spatial domain.

C. 3-D WAL-based Direction Aligned Wavelet Transform

The 3-D directional threading technique has the property of separable processes and is perfectly matched with 3-D WAL-based wavelet transform, which has the similar three separable processes. Therefore, the 3-D direction prediction process can be seamlessly incorporated into the 3-D WAL-based wavelet transform, leading to a new class of algorithm called direction aligned wavelet transform (DAWT), which consists of direction aligned temporal filtering (DATF) and 2-D direction aligned spatial filtering (DASF). Although wavelet transform in temporal domain is usually called motion aligned temporal filtering (MATF), the MATF can also be considered as the DAWT in temporal axis since the motion vector indicates the directional information of the pixel's movement along the temporal axis. The 3-D directional threading technique provides us an opportunity to align a series of video frames to form a 3-D video cube along its 3-D directional trajectory, while the 3-D WAL-based wavelet transform is performed to decompose the 3-D video cube into a 3-D multi-spatio-temporal resolution video pyramid, which provides full spatio-temporal-quality scalability for scalable video coding.

IV. EXPERIMENTAL RESULTS

In the experiment, MSRA 3-D wavelet video coder VIDWAV 2.0 [7] is used as the reference software. We report the experimental results of two MPEG standard test sequences: *Carphone* and *Foreman*. In the tests, each sequence is temporally decomposed by a four-level lifting-based wavelet transform into five temporal subbands. Each temporal subband is further spatially decomposed by a three-level lifting-based wavelet transform. The resulted wavelet coefficients are entropy coded by 3-D EBCOT. In order to evaluate the proposed 3-D WAL performance as objectively as possible, we replace the 3-D DWT module of VIDWAV reference software with the proposed 3-D WAL module and use the same bit-plane coding and 3-D EBCOT technique as VIDWAV reference software.

Fig. 4 presents the experimental results for video sequences *Carphone* and *Foreman* at different resolutions and frame rates with 5/3 or 9/7-tap spatial wavelet filters. In the tests, temporal filtering with 5/3-tap filter is used in each video sequence and only the Y component of each video sequence is presented in the PSNR bit-rate curves. As seen from the rate distortion curves at different test points, the proposed 3-D WAL method outperforms consistently the 3-D DWT method. For 5/3-tap spatial filter, the average PSNR gains of 3-D WAL method over 3-D DWT method are 0.89 dB for *Carphone* and 1.12 dB for *Foreman*, respectively; and for 9/7-tap spatial filter, the average PSNR gains are 0.69 dB for *Carphone* and 1.00 dB for *Foreman*, respectively. The highest coding gain of the 3-

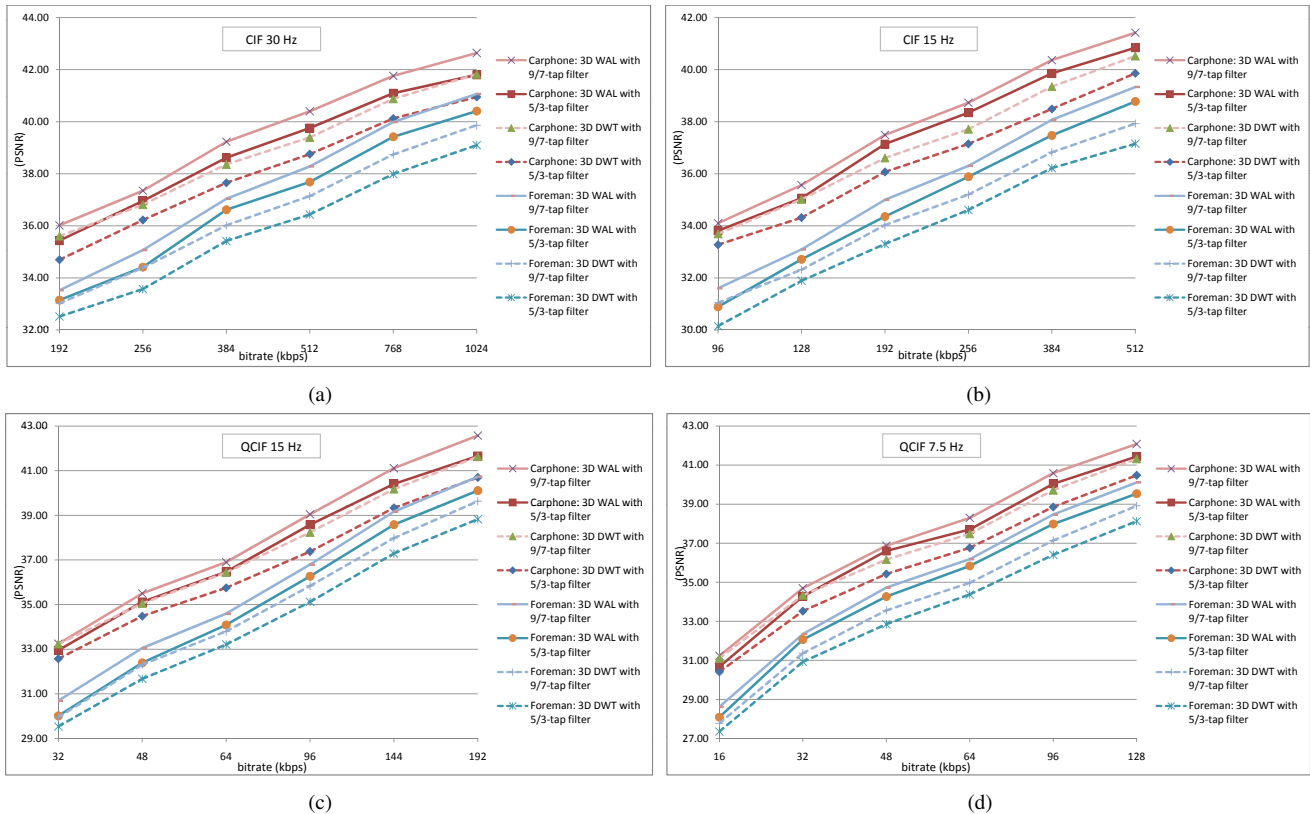


Fig. 4. The performance comparison of 3-D WAL-based video coder and 3-D DWT-based video coder for the Y component of *Carphone* and *Foreman* video sequences with 5/3 and 9/7-tap spatial filter. (a) CIF 30 Hz, (b) CIF 15 Hz, (c) QCIF 15 Hz, (d) QCIF 7.5 Hz

D WAL method can be up to 1.62 dB for *Foreman* sequence compared with the 3-D DWT method.

For the computational complexity comparison of the two 3-D wavelet video coders, it is obvious that the 3-D DWT-based video encoder has lower computational complexity. Although the 3-D DWT-based video encoder has time-consuming motion estimation process in temporal domain, the 3-D WAL-based video encoder increase considerably the computational complexity due to 3-D direction estimation process, which is not only performed in temporal domain but also in 2-D spatial domain to obtain the 3-D directional information. In addition, the proposed 3-D WAL method needs to solve the Wiener-Hopf equations to determine the adaptive interpolation filter coefficients, which slightly increases the computational complexity. On the other hand, the computational complexity of 3-D WAL-based video decoder is comparable with that of 3-D DWT-based video decoder. The reason is that both of the two coders are asymmetric, and the inverse transform of 3-D WAL-based video decoder is only slightly more computationally expensive than that of 3-D DWT-based video decoder, once the additional side information is given.

V. CONCLUSION

In this paper, we present a 3-D directional wavelet coding technique for scalable video coding. The proposed 3-D directional threading is seamlessly incorporated into 3-D weighted adaptive lifting-based wavelet transform to exploit the spatio-temporal correlation inside the video cube along

the 3-D directional trajectory. Experimental results show that the proposed 3-D WAL-based DWT consistently outperforms the conventional 3-D DWT for scalable video coding, and up to 1.62 dB improvement in PSNR is observed.

ACKNOWLEDGMENT

This work was supported in part by a grant from CUHK Direct Grant for Research Scheme under Project 2050383 and is affiliated with the Microsoft-CUHK Joint Laboratory for Human-centric Computing and Interface Technologies.

REFERENCES

- [1] W. Ding, F. Wu, X. Wu, S. Li, and H. Li, "Adaptive directional lifting-based wavelet transform for image coding," *IEEE Trans. Image Process.*, vol.16, no.2, pp.416-427, Feb. 2007
- [2] C.-L. Chang and B. Girod, "Direction-adaptive discrete wavelet transform for image compress," *IEEE Trans. Image Process.*, vol.16, no.5, pp.1289-1302, May 2007
- [3] Y. Liu and K.N. Ngan, "Weighted adaptive lifting-based wavelet transform," *2007 IEEE Int. Conf. Image Process. (ICIP2007)*, San Antonio, USA, Sept. 2007
- [4] J. Xu, Z. Xiong, S. Li and Y.-Q. Zhang, "Three-dimensional embedded subband coding with optimized truncation (3D ESCOT)," *Applied and Computational Harmonic Analysis*, pp.290-315, 2001
- [5] Y. Liu, F. Wu, and K.N. Ngan, "3-D object-based scalable wavelet video coding with boundary effect suppression", *IEEE Trans. Circuits Syst. Video Technol.*, vol.17, no. 5, pp.639-644, May 2007
- [6] R. Xiong, F. Wu, J.Xu, S. Li and Y.-Q. Zhang, "Barbell lifting wavelet transform for highly scalable video coding," *Picture Coding Symposium 2004*, USA, Dec 2004
- [7] R. Xiong, X. Ji, D. Zhang, J. Xu, G. Pau, M. Trocan, and V. Botreau, "Vidwav Wavelet Video Coding Specifications," *ISO/IEC JTC1/SC29/WG11/M12339*, Poznan, July 2005