

# 3D Object-based Scalable Wavelet Video Coding with Boundary Effect Suppression

Yu Liu<sup>1</sup>, Feng Wu<sup>2</sup>, and King Ngi Ngan<sup>1</sup>

<sup>1</sup>Department of Electronic Engineering, The Chinese University of Hong Kong, Hong Kong SAR

<sup>2</sup>Microsoft Research Asia, Beijing, 100080, China

**Abstract**—This paper extends the lifting-based motion threading technique from the frame-based coding to the object-based coding, attracted by the unique advantages of the object-based coding that do not exist in other coding schemes. The boundary effects, which exist in spatial and temporal wavelet transforms due to the manner of object-based coding, are well suppressed by 3D shape adaptive discrete wavelet transform (3D SA-DWT) via the lifting implementation in a unified framework. Based on the proposed object-based motion threading technique, a 3D object-based scalable wavelet video coder is developed. Experimental results show that the proposed coder achieves good performance compared with the existing object-based coders, and provides the scalability functionality at the same time.

## I. INTRODUCTION

The standard video coding paradigm is expected to migrate from the frame-based approach to the object-based approach with the emergence of the MPEG-4 standard [1] and the MPEG-7 standard [2]. Object-based coding enables accessibility and manipulability of object within a video sequence, and allows the structure of video content to survive the process of acquisition, editing and distribution, which is useful for content-based search and retrieval in MPEG-7. As an alternative to traditional video coding standard, 3D wavelet video coding has received much attention recently. A main advantage of 3D wavelet video coding is that it can provide full spatio-temporal-quality scalability with non-redundant 3D subband decomposition. In 3D wavelet video coding, motion compensation is usually incorporated into the temporal wavelet transform to achieve efficient coding performance, leading to a class of algorithms generally called as motion compensated temporal filtering (MCTF). However, to the authors' best knowledge, there is no literature incorporating the object-based motion compensation approach into lifting-based MCTF framework.

Xu et al. [3] propose a motion threading (MT) technique that employs longer wavelet filters to exploit the long-term correlation across frames along motion trajectory. The aim of MT is to form as many long threads as possible because

too many short threads will significantly increase the number of artificial boundaries. Luo et al. [4] propose an advanced MT technique to reduce the number of many-to-one mapping pixels and non-referred pixels in the original MT. Although the problem of boundary effects caused by the truncation of many-to-one mapping case in the original MT can be well solved by the advanced MT, the non-referred pixel case, which is assigned to use the motion vector from adjacent motion thread in the advanced MT, is not well solved because the assigned motion vector may be not accurate or even wrong for non-referred pixel and may cause some degradation on coding performance.

In object-based coding, the boundary effects also exist in 2D spatial wavelet transform for arbitrarily shaped video object plane. Due to the boundary effects in spatial and temporal wavelet reconstructions, these artificial boundaries will degrade greatly the coding performance. Therefore, it is better to solve the problem of boundary effects, which exist in spatial and temporal transforms of object-based coding simultaneously, in a unified framework. In this paper, we will address 3D object-based scalable wavelet video coding with boundary effect suppression.

## II. OBJECT-BASED MOTION THREADING USING LIFTING

To obtain more accurate motion trajectory of each pixel and the functionality of object-based coding with arbitrary regions of support, the object-based motion threading with the lifting structure is proposed, as shown in Fig.1. The object-based motion threading using lifting structure aims to reduce boundary effects of artificially terminating/emerging threads in the previous MT techniques.

In the previous MT techniques, when the pixel locates near the boundary of object, it is most likely that the best matching pixel in the reference frame lies outside of the object. That is, the previous motion thread reflects the fake motion trajectory of the pixel. Because of the wrong matching between the pixel inside the video object and the reference pixel outside the video object, this increases the probability of the many-to-one mapping case. In addition, due to the cover/uncover situations, some pixels may not appear in the previous or next VOP (video object plane). Thus a new motion thread has to be formed for each of non-referred pixel, which results in an artificially terminating/

---

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

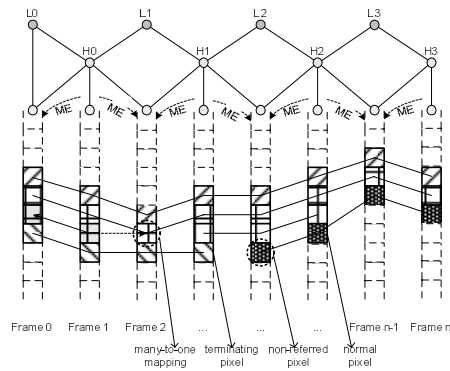


Fig. 1. The object-based motion threading technique with lifting structure

emerging thread. These cases increase the number of artificial boundaries and degrade the coding performance potentially. In the proposed object-based motion threading technique, the motion estimation is performed only on the pixels inside the video object at the macroblock level. In order to perform motion prediction on the VOP, the motion estimation of the blocks on the VOP boundaries uses polygon matching, instead of block matching, and the macroblock-based padding technique is adopted for the reference VOP. Finally, the pixels inside the video object along the same motion trajectory are linked to form an object-based motion thread. For those pixels outside the video object, no motion threading is needed.

For the non-referred pixel in the motion thread, unlike the advanced motion threading [4], which assigns a motion vector from adjacent motion thread to that of the non-referred pixel, the proposed object-based motion threading adopts a different method. Since the non-referred pixel originally indicates the boundary of the motion thread, the case of temporal filtering over the boundary of the terminating/emerging motion thread is similar to that of shape adaptive 2D spatial wavelet transform. We call this case as shape adaptive temporal wavelet transform for the motion threading signal segment. There are many approaches [5]-[7] in the literature for transforming arbitrary shaped signal segment to wavelet domain (we will introduce briefly these approaches in the next section). Among them, the shape adaptive transform method [7] of using bi-orthogonal symmetric filters with symmetric extensions over the boundaries is the most popular method. The temporal filtering over the boundary of the motion thread can also be treated in the same way by using the symmetric extension over the motion thread boundary.

In additional, due to the zoom out/zoom in of camera motion, this will also produce the many-to-one mapping pixels in a video sequence. For the pixel which is originally terminated in many-to-one mapping, it has been well solved by the advanced motion threading technique, which is proposed by Luo et al. [4]. Though the motion threads may be overlapped due to the many-to-one mapping generated

from motion estimation, those threads need not be truncated. And the lifting structure with the symmetric extension over the motion thread boundary ensures the temporal wavelet transform to be perfectly invertible. Therefore, the boundary effects of artificially terminating/emerging motion thread in the previous motion threading techniques can be well suppressed using lifting structure with symmetric extension.

### III. 3D SA-DWT VIA LIFTING

As mentioned in the previous section, many approaches [5]-[7] have been proposed for transforming arbitrary shaped signals to wavelet domain. Depending on the treatment of the boundary in the transform of finite length arbitrary shaped signals, these approaches can be divided into three categories: projections onto convex sets [5], boundary filters [6] and extension over the boundaries [7]. Among them, the most popular method is that of bi-orthogonal symmetric filters with symmetric extensions over the boundaries proposed by Li et al. in [7]. A video sequence consists of one or more video objects, and a video object appears in the form of video object plane at an instant time. That is, a 3D video object consists of a series of video object plane in temporal axis. The paper [7] focuses on the texture coding of 2D still objects or video object planes, and not 3D video objects. Therefore, there need efforts in developing coding techniques for arbitrarily shaped 3D video objects. The object-based motion threading technique provides us an opportunity to align a series of video object planes to form a 3D video object. By using the similar principle, 2D-SA-DWT is easily extended to 3D-SA-DWT by applying 1D-SA-DWT in the temporal motion threading segment and then applying 2D-SA-DWT in the 2D spatial signal segment (T+2D), or vice versa (2D+T).

In a typical 3D wavelet video coding, temporal transform and 2D spatial transform are done separately, and 2D spatial transform also involves the two separable 1D transforms, as shown in Fig. 2(b). As discussed before, the temporal filtering over the boundary of the motion thread can also be treated in the same way as the 2D shape adaptive spatial transform by using the symmetric extension over the boundary. Therefore, this allows us to focus only on 1D shape adaptive wavelet transform, either for temporal motion threading segment or for spatial arbitrary shaped signal segment, with the lifting structure for the analysis in this section. The lifting structure [8] is an efficient implementation of the wavelet transform with low memory and computational complexity. Consider a 1D signal segment with finite length  $n>1$  (No transform is needed when  $n=1$ ). A lifting wavelet with a symmetrical boundary extension is applied on each segment independently, i.e., as shown in Fig. 2 (a). Each lifting steps only updates half of the nodes, and the original value of the updated nodes will not be needed in subsequent steps. For the wavelet

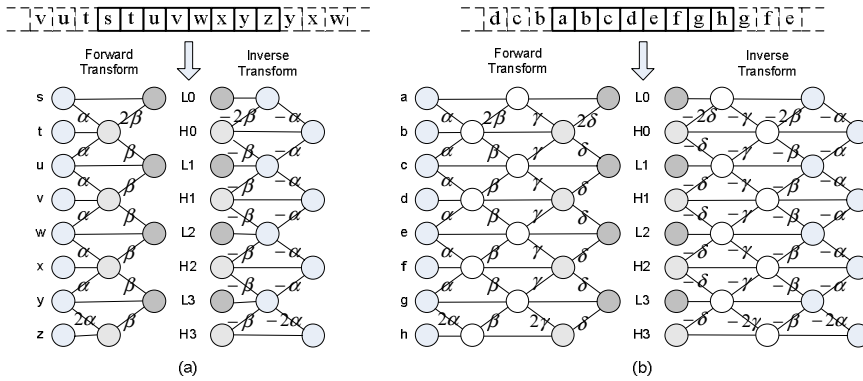


Fig. 2. Separable 3D shape adaptive wavelet transform via lifting implementation. (a) Shape adaptive lifting wavelet for temporal motion threading segment, with biorthogonal 5/3-tap filters; (b) Shape adaptive lifting wavelet for 2D spatial signal segment, with biorthogonal 9/7-tap filters

transform, low-pass and high-pass coefficients are interleaved. Each lifting step can be straightforwardly inverted with an inverse lifting unit.

The biorthogonal 5/3 lifting wavelet for temporal motion threading segment can be depicted by a 2-stage lifting with coefficients,  $\alpha = -0.5$  and  $\beta = 0.25$ . And the biorthogonal 9/7 lifting wavelet for 2D spatial signal segment can be depicted by a 4-stage lifting with coefficients,  $\alpha = -1.586$ ,  $\beta = -0.052$ ,  $\gamma = 0.883$ , and  $\delta = 0.444$ .

Based on the shape-adaptive wavelet transform discussed above, the 3D SA-DWT via the lifting implementation (T+2D) for an arbitrarily shaped 3D video object can be described as follows:

- 1) Apply the object-based motion threading to align a series of video object planes to form a 3D video object.
- 2) Within the bounding box of the 3D video object, using shape information identifies the pixels belonging to the video object to be transformed.
- 3) Within each group of plane (GOP), apply 1D shape adaptive lifting wavelet with biorthogonal 5/3-tap filters to each temporal motion threading segment.
- 4) Perform the above operation to the low-pass band video object until the desired level of temporal wavelet decomposition is reached.
- 5) Within each video object plane (VOP), apply 2D separable shape adaptive lifting wavelet with biorthogonal 9/7-tap filters to 2D spatial signal segment,
  - a) Apply the 1D shape adaptive lifting wavelet to each row signal segment;
  - b) Apply the 1D shape adaptive lifting wavelet to each column signal segment;
  - c) Perform the above operation to the low-low-pass band video object until the desired level of spatial wavelet decomposition is reached.

The above 3D SA-DWT via the lifting algorithm provides a way to efficiently decompose an arbitrarily shaped 3D video object into a 3D multi-spatio-temporal resolution video object pyramid. This non-redundant multi-spatio-temporal resolution pyramid decomposition structure

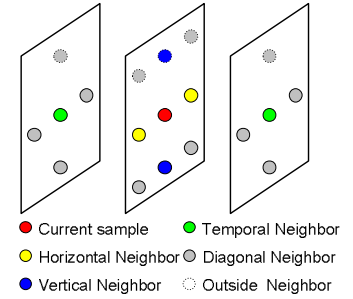


Fig. 3. Coding neighbors in context Information

can provide full spatio-temporal-quality scalability for object-based coding. The 3D wavelet coefficients are coded with a bitplane coding scheme named 3D Embedded Block Coding with Optimal Truncation (3D EBCOT). The extension from 3D EBCOT to object-based coding is straightforward and efficient. Only those wavelet coefficients that belong to the video object are coded in 3D object-based EBCOT. If any neighboring pixels fall outside the video object, we assume that their value is zero in context modeling, as shown in Fig.3.

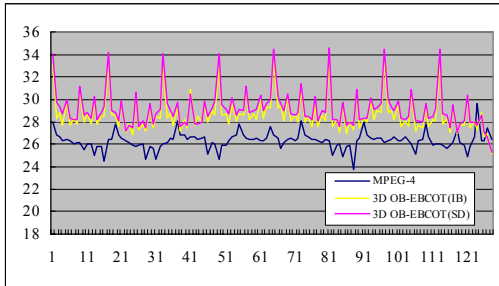
#### IV. EXPERIMENTAL RESULTS

The MPEG-4 standard test sequences are employed in the simulation and the corresponding shape masks of these video objects are supplied by MPEG. The PSNR is computed with respect to the foreground object pixels, and the bit costs are only for texture coding, where the bit costs of shape mask information are not accounted in the object-based coding. The proposed object-based scalable wavelet video coder (3D OB-EBCOT) includes two versions: spatial-domain MCTF -based coder, named 3D OB-EBCOT (SD), and in-band MCTF -based coder, named 3D OB-EBCOT(IB). We compared the proposed object-based scalable wavelet video coder with MPEG-4 object coding. We first run the MPEG-4 on a sequence to obtain an average bit rate and then code the sequence using 3D OB-EBCOT with the same average bit rate again. Table I lists the average PSNR (in dB) results given by different coders on the four foreground object sequences, Akiyo, Bream, Children and Foreman. Fig. 4 shows the comparisons of 3D OB-EBCOT with MPEG-4 in terms of PSNR plots (in dB) for the foreground objects, Children and Foreman.

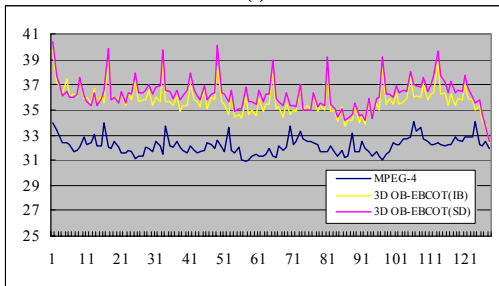
We also compared the performance of 3D OB-EBCOT with those of different 3D SPIHT coding schemes. Table II lists the average PSNR (in dB) results from coding the QCIF foreground video objects, Akiyo and Bream, with a frame rate 30 fps at different bitrates. The experimental results of different 3D SPIHT coding schemes comes from the literature [9]. AROS means arbitrary regions of support.

TABLE I: COMPARISON OF 3D OB-EBCOT WITH MPEG-4 IN TERMS OF AVERAGE PSNR (IN DB) FOR FOREGROUND OBJECTS

Sequence (CIF)	Akiyo			Breem			Children			Foreman		
Bitrate (kbps)	112			160			219			210		
Component	Y	U	V	Y	U	V	Y	U	V	Y	U	V
MPEG-4	34.95	40.32	41.11	28.19	33.55	39.53	26.30	31.96	29.18	32.16	38.50	37.77
3D OB-EBCOT(IB)	38.12	42.45	43.25	30.12	35.11	40.06	28.70	33.98	30.95	35.75	41.30	41.00
3D OB-EBCOT(SD)	38.43	42.81	43.68	31.46	36.22	40.24	29.16	34.08	31.14	36.27	41.06	41.33



(a)



(b)

Fig. 4. Comparison of 3D OB-EBCOT with MPEG-4 in terms of PSNR plots (in dB) for foreground objects, (a) Children, (b) Foreman

The 3D OB-EBCOT coder outperforms the AROS T with modified 3D SPIHT by 4.12-5.80 dB at the same bitrate. It is worth pointing out that the video objects need not be compressed multiple times by 3D OB-EBCOT to achieve the different target bitrates. The 3D OB-EBCOT coder is an object-based scalable video coder and is able to achieve all of these different target bitrates in a single bitstream.

In addition, to demonstrate that object-based motion threading with lifting structure further improves the performance of the original motion threading, we compared the performance of 3D OB-EBCOT with those of MPEG-4, AROS T with modified 3D SPIHT, and object-based 3D ESCOT[3] on coding the QCIF foreground video objects, Akiyo and Coastguard (Object one). Table III lists the average PSNR (in dB) results given by different coders on these two video objects. Observed from the results of the video object Coastguard, the improvement of object-based motion threading with lifting structure is considered significant, about 1.27dB gain over object-based 3D ESCOT with object-based motion threading.

## V. CONCLUSION

In this paper, an object-based motion threading with the lifting structure is proposed to improve the previous motion threading techniques for object-based video coding. The

TABLE II: COMPARISON OF 3D OB-EBCOT WITH THREE 3D SPIHT CODING SCHEMES IN TERMS OF AVERAGE PSNR (IN DB) FOR FOREGROUND OBJECTS

Sequence (QCIF)	Akiyo			Breem		
Bitrate (kbps)	20	40	60	20	40	60
3D SPIHT Method 1	26.78	31.08	33.84	19.66	21.28	22.33
3D SPIHT Method 2	28.99	33.24	36.07	22.13	23.73	25.00
3D SPIHT Method 3	29.04	33.29	36.10	22.17	23.75	25.02
3D OB-EBCOT(SD)	33.72	37.81	40.22	26.29	28.20	30.82

TABLE III: COMPARISON OF 3D OB-EBCOT WITH OTHER CODING SCHEMES IN TERMS OF AVERAGE PSNR (IN DB) FOR FOREGROUND OBJECTS

Sequence (QCIF)	Akiyo	Breem	
Bitrate (kbps)	24.57	45.91	35.45
MPEG-4	31.63	34.87	28.13
ARPS T with modified 3D SPIHT	-	34.38	-
3D OB-ESCOT without MT	32.10	-	28.48
3D OB-ESCOT with block-based MT	-	-	29.91
3D OB-ESCOT with object-based MT	-	-	30.57
3D OB-EBCOT(SD)	34.99	38.48	31.84

boundary effects, which exist in spatial and temporal wavelet transforms due to the manner of object-based coding, are well suppressed by 3D SA-DWT via lifting in a unified framework. Based on the object-based motion threading technique, a 3D object-based scalable wavelet video coder is developed. Experimental results show that the proposed coder outperforms the existing object-based coders, MPEG-4, 3D SPIHT with AROS, and object-based 3D ESCOT, on average, and provides the scalability functionality at the same time.

## REFERENCES

- [1] "MPEG-4 Information Technology: Coding of Audio-Visual Objects, Part 2: Visual", ISO/IEC 14496-2, 2000
- [2] "MPEG-7 Overview (version 10)", ISO/IEC JTC1/SC29/WG11 N6828, Palma de Mallorca, October 2004
- [3] 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
- [4] L. Luo, F. Wu, S. Li, Z. Xiong, Z. Zhuang, "Advanced motion threading for 3D wavelet video coding", *Signal Process.: Image Comm.*, vol. 19, no 7, pp 601-616, 2004.
- [5] D.C. You, "Mathematical theory of image restoration by the method of convex projections", *Image Recovery: Theory and Applications*, H. Stark, Ed. New York: Academic, 1987
- [6] C. Herley, "Boundary filters for finite-length signals and time-varying filter banks", *IEEE Trans. Circuits Syst. II*, Vol.42, pp.102-144, 1995
- [7] S. Li, W. Li, "Shape-adaptive discrete wavelet transforms for arbitrarily shaped visual object coding", *IEEE Trans. Circuits and Systems for Video Technology*, Vol.10 (5), pp.725-743, Aug 2000
- [8] W. Sweldens and P. Schroder, "Building your own wavelets at home", *ACM SIGGRAPH Course Notes*, pp.15-87, 1996
- [9] G. Minami, Z. Xiong, A. Wang, and S. Mehrotra, "3-D wavelet coding of video with arbitrary regions of support", *IEEE Trans. CSVT*, Vol.11 (9), pp.1063-1068, Sept. 2001