WAVELET CODING OF VIDEO OBJECT BY OBJECT-BASED SPECK ALGORITHM

Zhitao Lu and William A. Pearlman

Center for Next Generation Video
Electrical, Computer and Systems Engineering Department
Rensselaer Polytechnic Institute Troy NY 12180-3590
luz2, pearlw@rpi.edu

ABSTRACT

In this paper, we present the OB-SPECK algorithm (object-based set partitioned embedded block coder) for wavelet coding of arbitrary shaped video objects. Based on the SPECK algorithm introduced in [1], the shape information of the video object in the wavelet domain is integrated into the coding process for efficient coding of a video object of arbitrary shape. The whole coding system is a motion prediction coding system. First, the object in inter frame is predicted by the block-based motion estimation and compensation; then the video object (in intra frame) and motion residue (in inter frame) are transformed by the object-based wavelet transform, which applies a proper sampling method so that the number of wavelet coefficients is the same as that of the original object and the spatial correlation within the object is preserved. Finally, the wavelet coefficients are encoded by the OB-SPECK algorithm. The proposed scheme achieves high coding efficiency and preserves the features of an embedded bitstream, low computational complexity and exact bitrate control.

1. INTRODUCTION

With the emergence of multimedia applications, functions such as access, searching, indexing and manipulation of visual information at the semantically meaningful object level, are becoming very important in research and standardization efforts. For example, one of the outstanding features of the MPEG-4 standard is the possibility of object-based image access to coded video data. In order to support these functions, object-based representation and compression of video signals are required.

There have been continuous efforts in developing coding techniques for arbitrarily shaped video objects. Among them, the shape-adaptive DCT (SA-DCT) [2] is the most popular one and it has been applied to the MPEG-4 verification model. There are several disadvantages of the SA-DCT technique. First of all, it inherits the block effect from the DCT. Second, in the implementation of SA-DCT, the alignment of the coefficients destroys the spatial correlation to some extent; therefore, the coding efficiency is degraded. Since wavelet transform scheme can avoid the block effect introduced in the DCT scheme, many different wavelet schemes have been proposed [3, 4, 5, 6]. In [3, 4], padding techniques are used to extend the video object to rectangular shape or even length in the horizontal and vertical directions in order to apply the conventional wavelet transform. The number of wavelet coefficients is larger than that of the original video object. In [5], the same alignment process as in the SA-DCT is applied, which will pejoratively affect the coding efficiency. In [6], the ZTE (zerotree entropy coding) algorithm [7] is extended to arbitrarily shaped video object coding. As in the ZTE, it improves the coding efficiency, but the feature of scalable bitstream is destroyed.

In this paper, we propose the object-based SPECK algorithm. Based on the object-based wavelet transform [8], shape information of video objects in wavelet subbands is integrated into the coding process. Besides the high coding efficiency, the proposed approach preserves the features of embedded bitstream, exact bitrate control and low computation complexity. The paper is organized as following: In Section 2, we introduce the object-based wavelet transform. Section 3 present the proposed OB-SPECK coding algorithm. Simulation results on MPEG-4 foreground sequences are given in Section 4. , Section 5 concludes the paper.

2. OBJECT-BASED WAVELET TRANSFORM

The object-based wavelet transform we used is proposed by Barnard [8]. Separable filters are used to decompose the given arbitrarily shaped object into four subbands (LL, LH, HL and HH), then the LL band is further decomposed. At each decomposition level, the signal is properly extended and down-sampled so that the original signal with length N is represented by the same number of wavelet coefficients with perfect reconstruction property.

When separable filters are used, the 2D arbitrarily shaped

[0]This work was supported in part by the National Science Foundationunder grant number EEC-9812706. The government has certain rights in this material.

object based transform is converted to two 1D arbitrary length signal transforms (in the horizontal and vertical directions). In order to preserve the spatial correlation within an arbitrarily shaped video object, two down-sampling methods are used for the segment lines which start at an even-numbered row (or column) position and at an odd-numbered row (or column) position. When the segment line starts at an even-numbered row (or column) position, the down-sampling keeps the even-numbered samples; when the segment line starts at an odd-numbered row (or column) position, the down-sampling keeps the odd-numbered samples which are also located on the even-numbered row (or column). For each above case, when the length N of a segment line is odd, one more pixel is added to extend the length of the segment line to even. In order to apply the same down-sampling methods as that of the even-length segment line, when the segment line starts at an even-numbered row (or column) position, the added pixel is located at the end of the segment line; when the segment line starts at an odd-numbered row (or column) position, the added pixel is located at the beginning of the segment line. Therefore, there are four cases of signal extension and down-sampling according to the parity of the position which the segment line starts and the parity of the segment line length. These four cases are summarized as below:

1. The segment line starts at an even-numbered column (or row) and its length is even. In this case, the filtering and down-sampling process take place as in the conventional wavelet transform. The even numbered coefficients are kept and odd numbered coefficients are discarded.

2. The segment line starts at an odd-numbered column (or row) and its length is even. Now the down-sampling process keeps the odd-numbered coefficients and discards the even-numbered coefficients.

3. The segment line starts at even-numbered column (or row) and its length is odd. One pixel is added at the right-hand side of the segment such that the last high-pass coefficient is zero.

4. The segment line starts at odd-numbered column (or row) and its length is odd. One pixel a is added at the left-hand side of the original signal instead of at the right-hand side, the value of a makes the first two coefficients equal to each other so that the energy added to the high frequency subband is small.

The transform we used here keeps the number of wavelet coefficients the same as the number of pixels in the original object. It also keeps the alignment and shape of the original object so that there is low increase of unnecessary energy in high frequency subbands.

3. OBJECT-BASED SPECK ALGORITHM

3.1. SPECK algorithm

The SPECK coding algorithm belongs to a class of embedded tree structured significance mapping schemes [9, 10, 1]. They are based on a hierarchical structure of pyramid subband transformation, such as wavelet transform. The transform coefficients are grouped into subband subsets related through a quadtree structure. In the SPECK algorithm, the quadtree is formed by successive recursive splitting of a subband block (parent) into four quadrants (children). The coding process consists of a *sorting pass* and a *refinement pass*. In the *sorting pass*, as shown in Fig. 1, when data set S is significant, it is partitioned into four small child sets, $O(S)$; each of these four child sets is further tested and partitioned until all the significant coefficients are found. In the *refinement pass*, the significant coefficients found in the *sorting pass* are transmitted to decoder according to the bit-plane transmission. The idea behind this is to exploit the clustering of energy found in transformed images and concentrate on those areas of the image which have high energy. For detailed information about the SPECK algorithm, please refer to the original papers [1].

3.2. Object-based SPECK algorithm

A straightforward extension of the SPECK algorithm to coding of video objects of arbitrary shape is that we set all the coefficients outside the object in each subband to zero. Then the original SPECK algorithm can be applied just as if the support of the object were rectangular. No modification of the algorithm iwould be required. This method is in efficient, since one bit must be transmitted to tell the decoder that the node or branch outside the object is insignificant under each threshold.

In our scheme, the shape information of video objects in the wavelet domain is integrated into the coding process. Similar to the object-based wavelet decomposition, the shape image is also decomposed into a pyramid of subbands, called the *shape mask pyramid*. In this way, the regions which belong to the object in each subband are known by both the encoder and the decoder. Each pixel of the shape mask has a 2-bit mask value: 1 bit is used to distinguish if the current wavelet coefficient is within the object; and the other bit is used to tell if its child branch is within the object. When the spatial orientation tree is constructed, which node and/or child branch are inside/outside the video object is known. Before the coding process, we prune the node and branch which are outside the video object. During the *sorting pass* in the SPECK algorithm, those nodes and branches are not added into any list of LSP, LIP and LIS. Therefore, no information about these nodes and branches are transmitted. When the encoder and decoder scan these nodes and branches, they will be informed by the *shape mask pyramid* and skip over them. The parent-child relation in the object-based SPECK algorithm is illustrated in Fig. 2.

The whole coding system is a motion prediction coding system. First, the object in inter frame is predicted by the block-based motion estimation and compensation; then the video object (in intra frame) and motion residue (in inter frame) are transformed by the object-based wavelet transform. Finally, the wavelet coefficients are encoded by the OB-SPECK algorithm.

4. EXPERIMENTAL RESULT

We use two MPEG-4 foreground object sequences, the Akiyo and Weather Sequences in both QCIF and CIF format, to test the performance of the OB-SPECK algorithm. The objects in these sequences are the persons. We test the OB-SPECK algorithm in two modes: the intra mode and the inter mode. In the intra mode, the video object is encoded as a series of still images without motion compensation. In the inter mode, the object is first predicted by the block-based motion estimation and compensation, then the motion residues are encoded. In intra mode tests, the video objects are encoded at 0.2 bpp, 0.5 bpp and 1.0 bpp. The bitrate (bpp) is calculated based on the number of pixels within the objects. We compare the coding results with the shape-adaptive DCT (SA-DCT) [2] which is adopted in MPEG-4 verification model and other wavelet-based schemes, such as Egger's EZW schemes [5] , OWT [3] and shape-adaptive ZTE (SA-ZTE) [6]. The PSNR results for frame 0 of the Akiyo sequence (CIF) are listed in Table 1. Figure 3 shows frame 0 of the Akiyo sequence decoded at bitrate of 0.2 bpp, 0.5 bpp and 1.0 bpp. In inter mode trials, the test sequences are: Akiyo and Weather sequences in QCIF format with 300 frames. The sequences are encoded/decoded at the frame rate of 10 fps, with all frames, except the first frame, encoded in the inter mode. The bitrates are at 20 kbps, 30 kbps, 40 kbps, 50 kbps, 60 kbps and 90 kbps. We compare the performance of our algorithm with that of the SA-DCT technique [11] which is adopted by the MPEG-4 verification model. The rate distortion performance for both the SA-DCT and the OB-SPECK for the Weather sequence are shown in Fig. 4. The improvement of average PSNR is 0.05 dB to 1.7 dB.

5. CONCLUSION

In this paper, an efficient coding technique, object-based SPECK algorithm, for arbitrary shape video objects is addressed. The experiment results on MPEG-4 test sequences show that the proposed algorithm achieves very high coding efficiency. As no arithmetic coding is used at present, the coding efficiency can be further improved by applying arithmetic coding especially at very low bitrate. The algorithm also preserves the features of embedded bitstream, exact bitrate control and low computation complexity. It can encode and decode the video objects independently so that each object can be accessed and manipulated in the compressed domain.

Table 1. Coding results of the OB-SPECK and other reported algorithms for frame 0 of the sequence Akiyo (CIF) in intra mode

Algorithm	bitrate	PSNR-Y	PSNR-U	PSNR-V
OB-SPECK	1.	37.55	42.55	42.25
SA-DCT [2]	1.	37.09	42.14	42.36
SA-ZTE [6]	1.	38.06	43.43	43.25
Egger's [5]	1.	36.40	42.53	42.40
OWT [3]	0.875	34.13	N/A	N/A

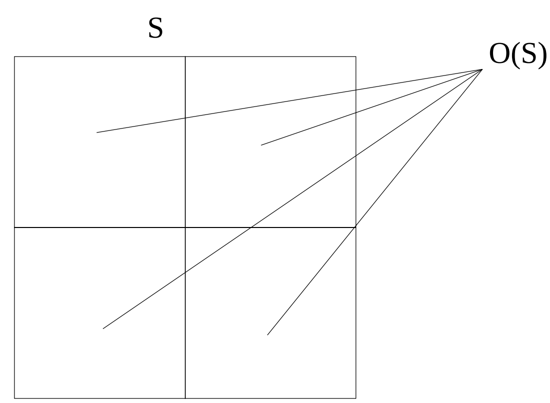

Fig. 1. Partitioning of set S

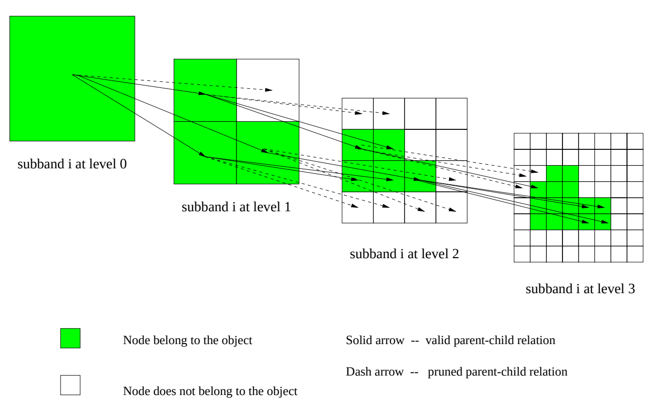

Fig. 2. In the parent-child relation in the OB-SPECK algorithm, the branches , which correspond to the nodes outside the object (represented by the dash arrows), are pruned before the coding process begins.

Fig. 3. Original and decoded frame 0 of the Akiyo sequence (QCIF): original (top left), 0.2 bpp (top right), 0.5 bpp (bottom left) and 1.0 bpp (bottom right)

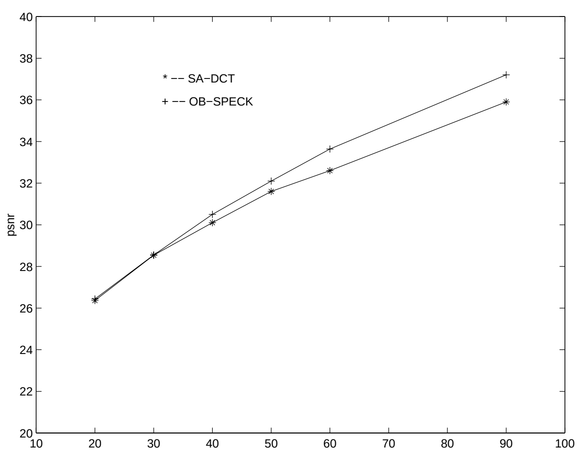

Fig. 4. Rate distortion performance of the SA-DCT and the OB-SPECK on the Weather sequence (QCIF) (inter mode)

6. REFERENCES

[1] A. Islam and W. A. Pearlman, "An embedded and efficient low-complexity hierarchical image coder," in *Proc. of SPIE (3653) on Visual Communications and Image Processing*, vol. 2, 1999.

[2] T. Sikora and B. Makai, "Low complexity shape-adaptive DCT for coding of arbitrarily shaped image segments," *Signal Process: Image Communications*, vol. 7, pp. 381–395, Nov. 1995.

[3] H. Katata, N. Ito, T. Aono, and H. Kusao, "Object wavelet transform for coding of arbitrarily shaped image segments," *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 7, Feb. 1997.

[4] J. H. Kim, J. Lee, E. Kang, and S. Ko, "Region-based wavelet transform for image compression," *IEEE Transactions on Circuits and Systems -II: Analog and Digital Signal Processing*, vol. 45, pp. 1137–1140, Aug. 1998.

[5] O. Egger, P. Fleury, and T. Ebrahimi, "Shape adaptive wavelet transform for zerotree coding," in *Proc. European Workshop Image Anal. Coding, Rennes*, vol. 1, 1996.

[6] S. Li and W. Li, "Shape-adaptive discrete wavelet transforms for arbitrarily shaped visual object coding," *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 10, pp. 725–743, Aug. 2000.

[7] S. A. Martucci, I. Sodagar, T. Chiang, and Y.-Q. Zhang, "A zerotree wavelet video coder," *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 7, pp. 109–118, Feb. 1997.

[8] H. J. Barnard, *Image and Video Coding Using a Wavelet Decomposition*. PhD thesis, Delft University of Technology, Dec. 1996.

[9] J. M. Shapiro, "Embedded image coding using zerotrees of wavelet coefficients," *IEEE Trans. on Signal Processing*, vol. 41, pp. 3445–3462, Dec. 1993.

[10] A. Said and W. A. Pearlman, "A new, fast and efficient image codec based on set partitioning in hierarchical trees," *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 6, June 1996.

[11] A. Kaup, "Object-based texture coding of moving video in mpeg-4," *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 9, pp. 5–15, Feb. 1999.