A NEW 3-D SUBBAND VIDEO CODING TECHNIQUE *

Hong Man[1] *Ricardo de Queiroz*[2] *Mark J. T. Smith*[3]

[1]Department of ECE, Stevens Institute of Technology, Hoboken, NJ 07030
[2]Xerox Corporation, Webster, NY 14580
[3]School of ECE, Georgia Institute of Technology, Atlanta, GA 30332-0250

ABSTRACT

We present in this paper a new 3-D subband coding method with low complexity, high performance, and features that facilitate channel error resilience. Four variations of the coder are discussed: one employing an octave band decomposition: one using the discrete cosine transforms (DCT) and two using lapped transforms. The base coder uses generalized adaptive quantization to code the subbands. Good performance at high compression ratios is obtained without traditional motion compensated prediction. Furthermore, these performance improvements can be realized without the use of entropy coding, which can make the system attractive for operation in error prone environments. Experimental results have shown that the subband video coder and its variations are able to achieve significant performance improvements over MPEG-2.

1. INTRODUCTION

Compression of video has been under study for many years now. Most techniques addressing this problem have been based on motion compensated prediction. More recently, approaches based on 3-D subband and transform coding have been appearing in the literature, some claiming to achieve performance results better than their motion compensation prediction counterparts. In this paper, we introduce a set of 3-D subband coders with a number of attractive features and compare and contrast their performance parameters. The set of coders are distinguished by the type of front-end decomposition employed. Specifically, four variations of the coder are studied: one using the DCT, one using an octave band filter bank (or discrete wavelet transform [DWT]), one using a lapped orthogonal transform (LOT)[1], and one using a lapped bi-orthogonal transform (LBT) [2].

The proposed 3-D subband video coding algorithm consists of two sequential stages: the 3-D subband transform for data decorrelation, and the quantization/coding of the transformed coefficients. In the coding stage, we have applied an adaptive quantization technique, which was introduced in [3]. The overall system is described in detail in the next two sections, followed by a synopsis of the experimental performance outcomes. Interestingly, all versions of the new coder outperform MPEG-2 by a significant margin.

2. 3-D DATA TRANSFORMS

The DCT has been widely adopted in the image and video coding standards because it provides excellent energy compaction for images and can be implemented with fast algorithms. However, the DCT is usually implemented on a block-by-block basis, which can lead to "blocking artifacts" at high compression ratios.

The LOT [1] was introduced to overcome this problem by performing the transform using a sliding window. The adjacent input data blocks overlap with each other during the transform. With overlapping bases, the amplitude transitions across block edges become smoother than in the DCT case. At the same time, because the forward transform takes more input samples to produce the same amount of output coefficients, it can improve the data decorrelation capability as compared to the DCT.

Although the LOT can significantly reduce the "blocking artifacts," it may still produce some visible block edges. A small modification of the LOT yields another transform, the LBT [2], which attempts to minimize edge discontinuities.

Tree structured filter banks (particularly octave band decompositions (or DWTs)) are now perhaps the most popular decompositions for 2-D data, as evidenced by their adoption into the new JPEG2000 image coding standard. In contrast to the DWT, the M-point DCT, LOT and LBT can be viewed as M-band uniform subband decompositions. That is, they can be interpreted as filter banks with downsampling factors. All are 1-D structures that are applied separably to each

*THIS WORK WAS SUPPORTED IN PART BY NJCWT, AND BY U.S. ARMY RESEARCH OFFICE UNDER GRANT # DAAH04-96-1-0161.

of the rows and then to each of the resulting columns to achieve the 2-D decomposition. For 3-D video, an extra 1-D transform can be applied in the temporal direction.

It has been reported [4] that the LBT usually outperforms the DCT significantly, and may also outperform the DWT slightly in image compression. In terms of computation, the LBT and LOT are more complex than the DCT but simpler than the DWT. If we use well known conservative implementations, one can implement an 8-point DCT using **40** operations (additions or multiplications). An 8-point LOT or LBT can be implemented with **82** or **83** operations. The generalized version of the LBT can be implemented with slightly more than **100** operations. As a reference, the DWT using the biorthogonal 9/7 filters [5] can be implemented using **182** operations for a 3-stage pyramid decomposition. There are many implementations and techniques for reducing computation of various transforms, which we will not discuss in further detail. However, we point out that the techniques that are commonly used to achieve efficient DWT implementations, such as ladder filter structures (lifting schemes) can also be applied to the LOT/LBT.

3. ADAPTIVE QUANTIZATION

The proposed video coder is an extension of the subband domain adaptive quantization technique introduced in [3, 6]. Adaptive quantization is accomplished by identifying **significant** (or **important**) coefficients inside each subband and performing quantization at bit rates according to their importance. The importance of a certain coefficient is measured by its contribution to the overall quantization fidelity. The quantization rate distribution is then encoded explicitly and transmitted to the decoder as a part of the compressed data stream.

In our video coding structure, all the subband coefficients are grouped into pixel blocks, which we call quantization *units*. The magnitude of each *unit* is represented by its L_2 norm. Quantization/coding operation is applied to each of these quantization *units*. In order to achieve an embedded bit stream, the quantization/coding stage is carried out in bit planes of the magnitude. The significance of a *unit* is determined by comparing its magnitude with a threshold at each coding layer (bit-plane). This threshold is halved at each successive layer. For quantization/coding of the coefficients, two operations are performed during each layer pass: the **MAP** operation and the **QUAN** operation.

The **MAP** operation identifies the significant *units* at each layer and codes their locations through a 3-D **quadtree** representation [6]. It produces a sequence of bits in the output bit stream that we call the **map** bits. If a subband is found to contain one or more significant *units*, the symbol "**1**" is produced, and this subband will be evenly split into $(2 \times 2 \times 2)$ regions. Each such region corresponds to a 3-D quadtree branch. If any of these regions contains one or more significant *units*, the symbol "**1**" is appended, and the region will be further split into $(2 \times 2 \times 2)$ sub-regions (or sub-branches). Otherwise, the symbol "**0**" is appended, and no further tests will be performed on the subject branch. This process continues in a recursive way, until each of the leaf branches can not be split any further. An optional arithmetic coding can be used for further compression of the decision symbols from this quadtree coding procedure.

Once a significant *unit* is identified, quantization is performed using a multistage residual Lattice Vector Quantization (LVQ). We denote this stage as the **QUAN** operation. All the quantization indices are stored in the **quan** sections of the output data stream. An LVQ codebook contains highly structured lattice points that effectively span the signal space. It does not require any training and can be implemented efficiently without codeword storage. Two different LVQs have been designed for different quantization stages. Both are derived from the root lattice Z_4, which is the union of all integer points in the 4-dimensional space. A 6 bits/vector sphere truncated LVQ is used for the first layer quantization, which has the ability to achieve sufficient shape gain. A 4 bits/vector cubic LVQ is applied to all the successive refinement layers, which guarantees the convergence of all the quantized approximations.

For progressive coding, each quantization stage is consistent with each layer pass in the **MAP** operation, and the residual errors of all the quantization *units* are bounded by the threshold of the last bit-plane. Entropy coding is not applied to the LVQ quantization indices (i.e. the **quan** bits), which essentially enacts fixed length coding (FLC) for this part of the data stream.

A unique feature of our subband coding framework is that it classifies the compressed data streams into the **map** sections and the **quan** sections which have significantly different error sensitivity levels. This enables a good adaptation to different channel models and error protection protocols [6].

4. EXPERIMENTAL RESULTS

The 3-D subband implementations are achieved through spatial and temporal extensions of four 1-D transform methods including DCT, LOT, LBT and DWT. The extensions are symmetric in all three dimensions, i.e. the number of decomposition levels are the same. Eight-point DCT, LOT and LBT are used to generate $8 \times 8 \times 8$ uniform subbands, which corresponds to a 3-level $2 \times 2 \times 2$-band uniform decomposition. The LOT and LBT matrices are based on [1, 2]. The popular 9/7-tap biorthogonal filter bank [5] is used in the DWT. However, when a data block is shorter than 8, the 2-tap Haar filter bank is used instead.

Every 16 picture frames are grouped together to form the input data cubes. Using the DCT, LOT or LBT, an 8-band uniform decomposition is obtained in a single stage. In order to improve the energy compaction, an additional one level $2 \times 2 \times 2$ DWT decomposition is performed in the lowest frequency (or DC) subband. The resulting video coders are denoted as `DCT`, `LOT` and `LBT`. We also tested two DWT implementations. First, we replaced the uniform decomposition of the DCT/LOT/LBT by a 3-level $2 \times 2 \times 2$ uniform decomposition using the 9/7 filters. As in the DCT/LOT/LBT cases, the DC subband is further decomposed once using the DWT filter banks. We denote the coder with this transform as `DWT-I`. The other DWT coder using a 4-level octave-band decomposition is denoted `DWT-II`. It is clear that the lowest eight subbands in all these implementations have the same size.

The size of the quantization unit in the 3-D subband domain is set to $2 \times 2 \times 1$, in which the 1 is for the temporal dimension. The arithmetic coder based on [7] is applied only to the **map** section of the bit streams.

We compared our subband video coders with the MPEG-2 coder. In the MPEG-2 setting, each group of pictures (GOP) contains 15 frames, which is close to the 16-frame setting in our coders. The I/P frame distance is set to 3. The testing video sequences were the luminance components of the "Akiyo" and "Hall Monitor" sequences, both in CIF (352×288) resolution and at 30 frames-per-second (fps). The testing bit rates were $1,520,640$ bps (or 0.5 bpp) and $760,320$ bps (or 0.25 bpp).

Figure 1 shows the frame-by-frame peak-signal-to-noise ratio (PSNR) results of all the coders on the "Akiyo" sequence. From these plots we can see that all our video coders outperform the MPEG-2 coder, sometimes by as much as **5** dB. A summarized comparison is provided in Table 1, in which the PSNR results are averaged over the 120 picture frames.

All these results reveal the fact that high performance video coding can be achieved by a simple 3-D transform coding approach without the complicated motion compensated prediction. Some observations worthy of note include:

- Uniform subband representations under these conditions slightly outperform octave-band decompositions.

PSNR	mean	max	min	mean	max	min
Akiyo	0.25 bpp			0.50 bpp		
LBT	**46.56**	48.75	44.06	**50.79**	52.28	49.17
LOT	46.16	48.24	44.10	50.48	52.02	48.85
DCT	46.22	48.34	44.09	50.53	52.07	49.01
DWT-I	45.65	48.11	41.31	49.90	51.83	46.76
DWT-II	44.82	46.07	42.14	48.33	49.14	46.97
MPEG-2	42.77	43.81	37.95	45.74	46.65	42.66
Hall	0.25 bpp			0.50 bpp		
LBT	**40.14**	41.67	38.94	42.37	43.63	41.47
LOT	40.03	41.50	39.30	**42.56**	43.69	42.15
DCT	**40.14**	41.61	39.43	**42.56**	43.61	42.18
DWT-I	39.74	41.23	38.52	42.24	43.44	41.07
DWT-II	37.74	39.59	36.32	41.16	42.41	39.71
MPEG-2	35.71	36.91	32.67	38.66	40.17	36.34

Table 1. Performance comparison of all tested video coders in PSNR (dB).

- The LOT/LBT outperforms the DWT and DCT when used in the new coder.
- Entropy coding is not particularly important in the new coder because only a small fraction of the compression is obtained through arithmetic coding on the **map** section of the bit stream.
- The potential saving in computation by using LOT/LBT is significant, relative to the DWT.

5. CONCLUSION

In this paper we have presented a new 3-D subband video coder. It consists of a simple 3-D data transform and an adaptive quantization procedure. We have implemented several video coders based on the DCT, LOT, LBT and DWT methods. All these video coders are able to provide performance superior to MEG-2. Simulation results have shown the viability of this approach in applications such as wireless video communications. In particular, we believe that a video coder with a simple 3-D LBT/LOT transform and a few levels of quadtree coding followed by a look-up table based LVQ can be an attractive choice for portable computing devices.

REFERENCES

[1] H. S. Malvar and D. H. Staelin, "The lot: transform coding without blocking effects," *IEEE Trans. on Acoustics, Speech, and Signal Processing*, vol. ASSP-37, pp. 553–559, Apr. 1989.

[2] H. S. Malvar, "Biorthogonal and nonuniform lapped transforms for transform coding with reduced blocking and ringing artifacts," *IEEE Trans. on Signal Processing*, vol. 46, pp. 1043–1053, Apr. 1998.

[3] H. Man, F. Kossentini, and M. Smith, "A family of efficient and channel error resilient

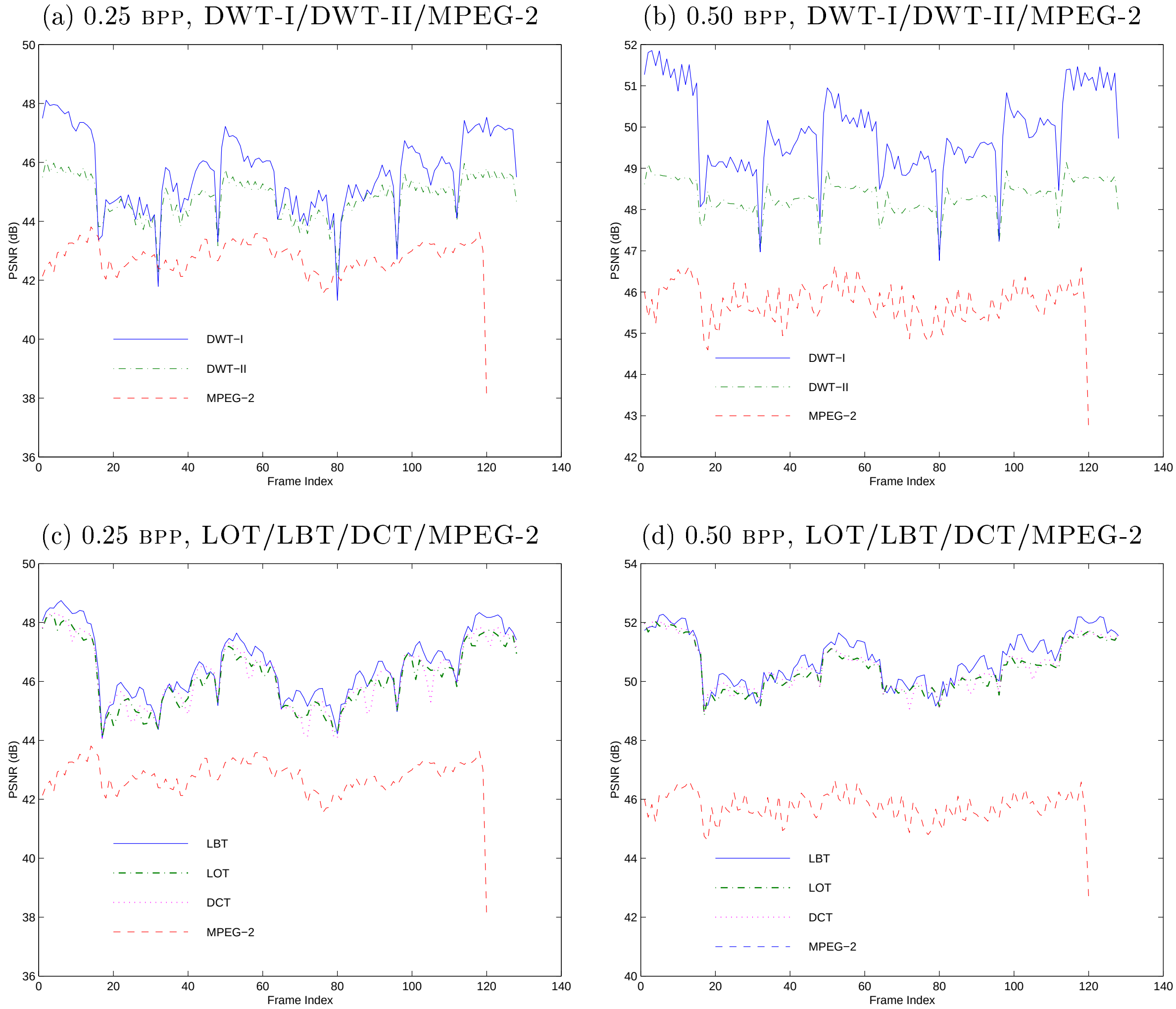

Figure 1. PSNR performance of the 3-D subband video coders DWT-I, DWT-II, LBT, LOT, DCT and the MPEG-2 coder for "Akiyo" sequence.

wavelet/subband image coders," *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 9, pp. 95–108, Feb. 1999.

[4] R. L. de Queiroz and T. D. Tran, "Lapped transforms for image compression," in *The Handbook on Transforms and Data Compression* (K. R. Rao and P. Yip, eds.), ch. 5, CRC Press, 2001.

[5] M. Antonini, M. Barlaud, P. Mathieu, and I. Daubechies, "Image coding using wavelet transform," *IEEE Trans. on Image Processing*, vol. 1, pp. 205–220, Feb. 1992.

[6] H. Man and M. Smith, "A new three-dimensional subband coding technique for digital video broadcasting," in *Second Annual UCSD/IEEE Conference on Wireless Communications*, (San Diego, CA), Feb.-Mar. 1999.

[7] I. H. Witten, R. M. Neal, and J. G. Cleary, "Arithmetic coding for data compression," *Communications of the ACM*, vol. 30, pp. 520–540, Jun. 1987.